Theme V – Models and Techniques for Analyzing Seismicity

Basic models of seismicity: Temporal models

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

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

Contents

Abstract In this and subsequent articles, we present an overview of some models of seismicity that have been developed to describe, analyze and forecast the probabilities of earthquake occurrences. The models that we focus on are not only instrumental in the understanding of seismicity patterns, but also important tools for time-independent and time-dependent seismic hazard analysis. We intend to provide a general and probabilistic framework for the occurrence of earthquakes. In this article, we begin with a survey of simple, one-dimensional temporal models such as the Poisson and renewal models. Despite their simplicity, they remain highly relevant to studies of the recurrence of large earthquakes on individual faults, to the debate about the existence of seismic gaps, and also to probabilistic seismic hazard analysis. We then continue with more general temporal occurrence models such as the stress-release model, the Omori-Utsu formula, and the ETAS (Epidemic Type Aftershock Sequence) model.

1 Introduction

In this article, we present an overview of several temporal point-process models that were well studied during the 1990's and 2000's. We cannot go through all the existing interesting models, and we only focus on some generic models, since most of the complicated models are constructed on the basis of these elegant generic models. These models include the Poisson model, the renewal models, the Omori-Utsu formula, the ETAS (Epidemic Type Aftershock Sequence) model, and the stress release model. To address them systematically, we classify these models into the following four classes: (1) 1D/temporal models, which include the stationary and non-stationary Poisson model, the renewal model, the Omori-Utsu (Reseanberg-Jones) model, temporal ETAS models and single-region stress release models; (2) discrete multi-region models, including linked stress release models and self- and mutually exciting models; (3) spatiotemporal models, including homogeneous and non-homogeneous Poisson models, space-time ETAS models; and, (4) models with auxiliary information, which incorporate information from observation or/and calculation results of other physical variables. In this article, we only discuss models from Class 1, i.e., temporal models that are formulated only based on the information of the catalogs. Please see other papers in this series for basic models from the 2nd and 3rd classes, and the CORSSA articles by Hainzl et al and Iwata (this theme) for the 4th class of models.

To illustrate how to utilize these models for seismicity analysis, we give examples that are extracted from various published papers. We encourage the reader to repeat these examples and to apply similar analysis to their own datasets.

2 Basic mathematical concepts

Earthquakes are caused by the rapid energy release accompanying the movement or fracture of faults in the crust or upper mantle of the earth, causing ground shaking of a duration from seconds to several minutes. However, generally only point-information is available for each earthquake, namely the origin time, hypocenter and magnitude. If we compress the time axis of the seismograph to view all the records of earthquakes in a large time scale, the waveforms of each earthquake become a pulse. We treat earthquakes as points in time and space. Such idealization enables us to study the earthquake process by using point process models.

Mathematically, a point process is a random object that takes value of a countable subset in a given spatiotemporal space. Here we can simply regard a point process as a type of stochastic model that defines probabilistic rules for the occurrence of points (i.e., earthquakes) in time and/or space. As well as purely temporal point process models, we consider in this article marked point processes that include earthquake magnitudes. In a marked point process, a mark or size (e.g., magnitude) is assigned to each point.

Denote a point process in time by N and a certain temporal location by t. We assume in the following discussions that we have known observations up to time t. The most important characteristic is the waiting time, say u, to the next event from time t. We consider the following cumulative probability distribution function of waiting time

$$F_t(u) = \Pr\{\text{from } t \text{ onwards, waiting time to next event} \leq u \,\}, \qquad (1)$$

with corresponding probability density function (p.d.f.)

$$f_t(u)\,\mathrm{d}u = \Pr\{\text{from } t \text{ onward, waiting time is between } u \text{ and } u + \mathrm{d}u \,\}, \qquad (2)$$

the survival function

$$S_t(u) = \Pr\{\text{from } t \text{ on, waiting time} > u \,\}$$
$$= \Pr\{\text{no event occurs between } t \text{ and } t + u\}, \qquad (3)$$

and the hazard function

$$h_t(u)\,\mathrm{d}u = \Pr\{\text{next event occurs between } t + u \text{ and } t + u + \mathrm{d}u$$
$$\mid \text{it does not occur between } t \text{ and } t + u.\}. \qquad (4)$$

One can easily prove that these functions are related in the following ways:

$$S_t(u) = 1 - F_t(u) = \int_u^\infty f_t(s)\,\mathrm{d}s = \exp\left[-\int_0^u h_t(s)\,\mathrm{d}s\right], \tag{5}$$

$$f_t(u) = \frac{\mathrm{d}F_t}{\mathrm{d}u} = -\frac{\mathrm{d}S_t}{\mathrm{d}u} = h_t(u)\,\exp\left[-\int_0^u h_t(s)\,\mathrm{d}s\right], \tag{6}$$

$$h_t(u) = \frac{f_t(u)}{S_t(u)} = -\frac{\mathrm{d}}{\mathrm{d}u}[\log S_t(u)] = -\frac{\mathrm{d}}{\mathrm{d}u}[\log(1 - F_t(u))]. \tag{7}$$

In the above equations, the first and second equality signs in each are obvious. Here we only outline how to derive the third equality signs in each. Equation (7) obtained from (4), i.e.,

$$\begin{aligned} h_t(u)\,\mathrm{d}u &= \frac{\Pr\{\text{next event occurs between } t+u \text{ and } t+u+\mathrm{d}u\ \}}{\Pr\{\text{the waiting time is greater than } u.\}} \\ &= \frac{f(u)\,\mathrm{d}u}{S(u)} = -\mathrm{d}\log S(u). \end{aligned}$$

To obtain the exponential term in (5), replace u by x in the above equality and integrate both sides with respect to x from 0 to u to get

$$\int_0^u h_t(x)\,\mathrm{d}x = -\int_0^u \mathrm{d}\,\log S(x) = \log S(0) - \log S(u) = -\log S(u), \tag{8}$$

i.e.,

$$S(u) = \exp\left[-\int_0^u h_t(x)\,\mathrm{d}x\right]. \tag{9}$$

The last equality sign in (6) can be proved by taking the derivative of the above equation with respect to u.

In the above concepts, the hazard function h_t is the key to understand our theory because it has the nice property of additivity. Suppose that a point process N consists of two independent sub-process, N_1 and N_2. Using the notation in (1) to (7) for N and denoting the waiting time distribution function, the survival function and the hazard function of the sub-processes by $F_t^{(i)}$, $S_t^{(i)}$ and $h_t^{(i)}$, respectively, with $i = 1$ or 2, respectively, for N_1 or N_2. Then, we can prove

$$F_t(u) = F_t^{(1)}(u) + F_t^{(2)}(u) - F_t^{(1)}(u)\,F_t^{(2)}(u), \tag{10}$$

$$f_t(u) = f_t^{(1)}(u) + f_t^{(2)}(u) - f_t^{(1)}(u)\,F_t^{(2)}(u) - F_t^{(1)}(u)\,f_t^{(2)}(u), \tag{11}$$

$$h_t(u) = \frac{f_t(u)}{1 - F_t(u)} = \frac{f_t^{(1)}(u)}{1 - F_t^{(1)}(u)} + \frac{f_t^{(2)}(u)}{1 - F_t^{(2)}(u)} = h_t^{(1)}(u) + h_t^{(2)}(u). \tag{12}$$

Another property of the hazard function is the following: if there is no event occurring between t and $t + u$ and $v \geq u \geq 0$, then $h_t(v) = h_{t+u}(v - u)$. This is followed by conditioning on no event occuring in $(t, t + u]$, then, when $x \geq 0$.

$$
\begin{aligned}
S_{t+u}(x) &= \Pr\{\text{no event occurs in } (t + u, t + u + x) \mid \text{no event occurs in } (t, t + u]\} \\
&= \frac{\Pr\{\text{no event occurs in either } (t + u, t + u + x) \text{ or } (t, t + u]\}}{\Pr\{\text{no event occurs in } (t, t + u]\}} \\
&= \frac{\Pr\{\text{no event occurs in } (t, t + u + x)\}}{\Pr\{\text{no event occurs in } (t, t + u]\}} \\
&= \frac{S_t(u + x)}{S_t(u)}
\end{aligned}
\tag{13}
$$

with $x \geq 0$ and so

$$
h_{t+u}(x) = \frac{\mathrm{d}}{\mathrm{d}x} \log S_{t+u}(x) = \frac{\mathrm{d}}{\mathrm{d}x} \log S_t(u + x) = h_t(u + x).
\tag{14}
$$

If we set $x = 0$ in the above equation, we have $h_t(u) = h_{t+u}(0)$ when no events occur in $(t, t + u]$, i.e., h collapses to a function of one variable. Thus, as a function of t, $h_t(0)$ is identical to $h_{t_0}(t - t_0)$ from a time t_0, until the occurrence of another event, say t_1, where $h_t(0)$ is identical to $h_{t_1}(t - t_1)$.

If an event occurs at t, then $h_t(0)$ has two possible explanations: one is the hazard function at $t - t_{prev}$ from the occurrence time t_{prev} of the last previous event, the other is the hazard function at 0 from the event occurring at t. To distinguish them, we use $h_{t_-}(0)$ and $h_{t_+}(0)$ for the first and the second explanations, respectively. Conventionally, the first explanation is used for $h_t(0)$, i.e., $h_t(0) = h_{t_-}(0)$.

Example 1 If the distribution time from t is an exponential distribution, i.e.,

$$
F_t(u) = 1 - \exp[-\lambda u], \text{ and } f_t = \lambda \exp[-\lambda u],
\tag{15}
$$

where λ is a positive constant, then

$$
h_t(u) = \lambda.
\tag{16}
$$

This is the case of the Poisson process, which will be discussed in detail in next section.

Since the function $h_t(0)$ plays a key role in our understanding of the theory of point processes, it is worthwhile to give it a specific name. Nowadays, it is usually named as the conditional intensity and denoted by $\lambda(t)$. Researchers also frequently write it as $\lambda(t|\mathcal{H}_t)$, where $\mathcal{H}_t$ represents the observation history up to time t but not including t, since the value of $\lambda(t)$ also can depend on what is in the observation history before t, but not on what happens at t.

From the definition of the hazard function, we have

$$\lambda(t)\, dt = \Pr\{\text{one or more events occur in } [t, t + dt) \mid \mathcal{H}_t\}. \tag{17}$$

This is usually used as the definition of the conditional intensity. When the point process N is simple, i.e., there is at most 1 event occurring at the same time location, then $\Pr\{N[t, t + dt) > 1\} = o(dt)$. Using this relationship, $\lambda(t)$ can be also defined as follows:

$$
\begin{aligned}
\mathbf{E}\left[N[t, t + dt) \mid \mathcal{H}_t\right] &= \sum_{n=0}^{\infty} n \Pr\{N[t, t + dt) = n \mid \mathcal{H}_t\} \\
&\approx \Pr\{N[t, t + dt) = 1 \mid \mathcal{H}_t\} + o(dt) \\
&\approx \lambda(t)\, dt + o(dt).
\end{aligned}
\tag{18}
$$

The conditional intensity has a natural form for the purpose of forecasting.

Given an observed dataset of a point process N, say $\{t_1, t_2, \cdots, t_n\}$, in a given time interval $[S, T]$. The likelihood function L is the joint probability density of waiting times to each of these events, i.e., the joint probability that the waiting time from S is in the range $(t_1 - S, t_1 - S + dt_1)$ and the waiting time from t_1 is in the range $(t_2 - t_1, t_2 - t_1 + dt_2)$ and so on, with the final term being that the waiting time from t_n is greater than $(T - t_n)$. The probability each of these terms is conditional on what has happened previously, so we can write the likelihood function as

$$
\begin{aligned}
&L(N; S, T)\, dt_1\, dt_2 \cdots dt_n \\
={}& \Pr\{\text{The waiting time from } S \text{ is in } (t_1 - S, t_1 - S + dt_1)\} \\
&\times \Pr\{\text{The waiting time from } t_1 \text{ is in } (t_2 - t_1, t_2 - t_1 + dt_2) \mid \text{what happens before } t_1\} \\
&\times \cdots \\
&\times \Pr\{\text{The waiting time from } t_n \text{ is greater than } T - t_n \mid \text{what happens before } t_n\} \\
={}& f_S(t_1)\, dt_1 \times f_{t_1}(t_2 - t_1)\, dt_2 \times \cdots \times f_{t_{n-1}}(t_n - t_{n-1})\, dt_n \times S_{t_n}(T - t_n) \\
={}& h_S(t_1 - S) \exp\left[-\int_S^{t_1} h_S(u - S)\, du\right] dt_1 \\
&\times \prod_{i=1}^{n-1} \left\{ h_{t_i}(t_{i+1} - t_i) \exp\left[-\int_{t_i}^{t_{i+1}} h_{t_i}(u - t_i)\, du\right] dt_i \right\} \exp\left[-\int_{t_n}^{T} h_{t_n}(u - t_n)\, du\right] \quad (\because (6)) \\
={}& \left[\prod_{i=1}^{n} h_{t_i}(0)\, dt_i\right] \exp\left[-\int_S^{T} h_u(0)\, du\right] = \left[\prod_{i=1}^{n} \lambda(t_i)\, dt_i\right] \exp\left[-\int_S^{T} \lambda(u)\, du\right].
\end{aligned}
\tag{19}
$$

We usually write the above formula as its logarithm, i.e.,

$$\log L(N; S, T) = \sum_{i=1}^{n} \log \lambda(t_i) - \int_S^{T} \lambda(u)\, du. \tag{20}$$

A direct use of the likelihood function is for parameter estimation: when the model involves some unknown regular parameters, say θ, we can estimate θ through maximizing the likelihood, i.e., the MLE (maximum likelihood estimate) is

$$\hat{\theta} = \arg_\theta \max L(N; S, T; \theta). \tag{21}$$

Given several models fit to the same dataset, the optimal one can be selected by using the Akaike Information Criterion (AIC, see *Akaike* 1974). The statistic

$$AIC = -2 \max_{\boldsymbol{\theta}} \log L(\boldsymbol{\theta}) + 2k_p \tag{22}$$

is computed for each of the models fit to the data, where k_p is the total number of estimated parameters for a given model. Comparing models with different numbers of parameters, the addition of the quantity $2k_p$ roughly compensates for the additional flexibility that the extra parameters provide. The model with the lowest AIC value is selected as the optimal choice for forward prediction purposes.

3 Long-term temporal models

In this section, we present the one-dimensional temporal models for long-term earthquake forecasting, including Poisson models, renewal models and stress release models. Roughly speaking, these models are intended to be used at time scales of 10 to 100 years or even longer.

3.1 Stationary Poisson models

We suppose that the readers of this section already have some related background knowledge of the Poisson process; if not, the readers are referred to Introduction to Probability Models (*Ross* 2003).

The Poisson process is named after the French mathematician Siméon-Denis Poisson (1871-1840). The process is also known as the model of complete randomness. Thus, it naturally serves as the null model in many hypothesis tests to clarify whether systematic structure is contained in the observations. For example, *Gardner and Knopoff* (1974) discussed whether the seismicity in California is Poissonian or not after the aftershock clusters are removed. Furthermore, the stationary Poisson process forms the basis of most of today's probabilistic seismic hazard analysis (*Cornell* 1968).

Let $N(a, b)$ be the number of events in a point process N that fall in the time interval (a, b). A point process is called a stationary Poisson process if the system has the following characteristics:

a) Independent increments: The numbers of events occurring in two disjoint time intervals are independent of each other. That is,

$$\Pr\{N(a,b) = m, N(c,d) = n\} = \Pr\{N(a,b) = m\}\Pr\{N(c,d) = n\} \qquad (23)$$

for all non-negative integers m, n, and any interval pairs $(a,b) \cap (c,d) = \varnothing$.

b) Stationarity: The probability distribution of the number of events falling in a time interval only depends on the length of the time interval.

c) Simplicity: There is never an occasion that two or more events occur simultaneously.

It can be shown that a stationary Poisson process has the following properties:

a) The number of events occurring in an interval of length S has a Poisson distribution, i.e.,

$$\Pr\{N(a, a+S) = n\} = \frac{\lambda^n S^n}{n!}\mathrm{e}^{-\lambda S}, \qquad (24)$$

where λ is the long term average of the number of events in an interval of unit length.

b) Given a fixed time, the forward occurrence time t_f (time to next event) and the backward occurrence time t_b (time back to the previous event) have the same exponential distribution as the waiting time τ between two consecutive events, i.e.,

$$\Pr\{t_f < x\} = \Pr\{t_b < x\} = \Pr\{\tau < x\} = 1 - \mathrm{e}^{-\lambda x}. \qquad (25)$$

c) The mean and the variance of the number of observations are the same:

$$\mathbf{E}\left[N(a,b)\right] = \lambda(b-a), \qquad \mathbf{Var}\left[N(a,b)\right] = \lambda(b-a).$$

d) Binomial distribution. Given that n is the total number of events occurring in the interval $[0, T]$, then the number of events in $[0, S]$, where $S < T$, is a non-negative integer-valued random variable from a binomial distribution with parameters $(n, S/T)$, i.e.,

$$\Pr\{N[0,S] = k \mid N[0,T] = n\} = \binom{n}{k}\left(\frac{S}{T}\right)^k\left(1 - \frac{S}{T}\right)^{n-k} \qquad (26)$$

e) Uniform distribution: Given that at least one event occurs in $[0, T]$, then the location of an arbitrary event is uniformly distributed in $[0, T]$.

f) Beta distribution: Given that there are a total of n events occurring in $[0, 1]$, $n \geq 1$, the location $t_{(k)}$ of the kth event $(k < n)$, from the left to the right, has a beta distribution, i.e.,

$$\Pr\{t_{(k)} \in (x, x+\mathrm{d}x) \mid N[0,1] = n\} = \frac{\mathrm{d}x}{\mathrm{B}(k, n-k)}\, x^{k-1}(1-x)^{n-k-1}, \qquad (27)$$

where $B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,\mathrm{d}t$ is the beta function.

Conditional intensity Recall the definition of the conditional intensity function of a point process, which is defined by

$$\lambda(t)\mathrm{d}t = \Pr\left\{N[t, t+dt) = 1 \mid \mathcal{H}_t\right\}. \tag{28}$$

For the stationary Poisson processes with rate λ, the right-hand side is $\lambda\, \mathrm{d}t\, \exp[-dt] + o(\mathrm{d}t)$, i.e., $\lambda(t) = \lambda$. This is why we use λ to denote both the rate of the Poisson process and the conditional intensity in the case of more general point processes.

Likelihood functions and maximum likelihood estimates of the Poisson process Suppose that all observations of a realization of a Poisson process with a rate of λ on a time interval $[S, T]$ is $\{t_1, t_2, \cdots, t_n\}$, where $S \le t_1 < t_2 < \cdots < t_n < T$. The likelihood of the observation is

$$L[S, T; \lambda] = \lambda^n \mathrm{e}^{-\lambda(T-S)}, \tag{29}$$

and the logarithm of the likelihood function is

$$\log L(S, T; \lambda) = n \log \lambda - \lambda(T - S). \tag{30}$$

Usually, the rate λ is unknown. We can estimate it through maximizing the likelihood function, i.e., the maximum likelihood estimate (MLE) of λ is

$$\hat{\lambda} = \arg_\lambda \max L(S, T; \lambda) \tag{31}$$

Taking the derivative of (30) and setting it to zero to obtain a maximum, we have

$$\left.\frac{\partial}{\partial\lambda} \log L(S, T; \lambda)\right|_{\lambda=\hat{\lambda}} = 0, \tag{32}$$

which yields

$$\hat{\lambda} = n/(T - S), \tag{33}$$

which is simply the average rate of events in the interval $[S, T]$.

3.2 Non-stationary Poisson models

A point process is a non-stationary Poisson process if it satisfies the independent increments and simplicity conditions, but violates the stationarity condition. Consequently, the Poisson process has a rate that is a function of time, denoted by $\lambda(t)$.

With the definition $\Lambda(t) = \int_0^t \lambda(x)\, dx$ and $\Lambda(t+s) = \int_0^{t+s} \lambda(x)\, dx$, the probability of observing n events in the interval $[t, t+s)$ is

$$\Pr\{N[t, t+s) = n\} = \frac{e^{-(\Lambda(t+s)-\Lambda(t))}[\Lambda(t+s) - \Lambda(t)]^n}{n!}.$$

That is, $N[t, t + s)$ is Poisson distributed with expectation value $\int_t^{t+s} \lambda(u)\, du$ where $\lambda(t)$ is the time dependent intensity. Also, given an event that falls in $[S, T]$, the probability density of its location is proportional to $\lambda(\cdot)$ on $[S, T]$, i.e., $\lambda(t)/\int_S^T \lambda(u)\, du$.

Likelihoods and maximum likelihood estimates of non-stationary Poisson processes
Suppose that the observation of a realization of a non-stationary Poisson process with a rate of $\lambda(t)$ on a time interval $[S, T]$ is $\{t_1, t_2, \cdots, t_n\}$, where $S \le t_1 < t_2 < \cdots < t_n < T$. In a similar way as for the stationary Poisson model, the likelihood of the observations is

$$L(S, T; \lambda) = e^{-\int_S^T \lambda(u)\, du} \prod_i \lambda(t_i). \tag{34}$$

The corresponding logarithm of the likelihood function is

$$\log L(S, T; \lambda) = \sum_i \log \lambda(t_i) - \int_S^T \lambda(u)\, du. \tag{35}$$

There are relative few applications of non-stationary Poisson models in seismology. This type of model is usually used as a natural alternative to the null stationary Poisson model when an apparent trend or seasonality is visible in the data. Some examples of non-stationary Poisson models are the exponential trend model in *Zheng and Vere-Jones* (1994) and the polynomial trend models and the seasonality models in *Ogata and Katsura* (2004) and *Ma and Vere-Jones* (1997).

3.3 Renewal/recurrence models

Renewal models are simple extensions of the Poisson model. One of their uses is to model the characteristic recurrence of earthquakes on a particular fault or in a particular region. This class of models is widely used in seismicity and seismic hazard analysis. For example, *Field* (2007) summarized how the Working Group on California Earthquake Probabilities (WGCEP) estimates the recurrence probabilities of large earthquakes on major fault segments using various recurrence models to produce the official California seismic hazard map. These models are sometimes justified by the elastic rebound theory proposed by *Reid* (1910). According to his theory, large earthquakes release the elastic strain that has built up since the last large earthquake. Some seismologists deduce that the longer it has been since the last earthquake, the more probable is an imminent event (e.g. *Nishenko and Buland* 1987; *Nishenko* 1991; *Sykes and Menke* 2006), while others contend that the data contradict this view (e.g. *Davis et al.* 1989; *Kagan and Jackson* 1995). Renewal models are often used to quantitatively demonstrate whether earthquakes occur temporally in clusters or quasi-periodically.

A renewal process is a generalization of the Poisson process. In essence, the Poisson process has independent identically distributed waiting times that are exponentially distributed before the occurrence of the next event. A renewal process is defined as a point process with the waiting times having a more general distribution, which is not necessarily exponential distribution. In this section, we denote the density function of the waiting times by $f(\cdot)$, which is also usually called the renewal density.

The conditional intensity of the renewal process can be derived as follows. Recall that the time interval between any two adjacent events is independent from other intervals. This means that the occurrence of the next event depends only on the time of the last event but not on the full history. Thus, from Section 1, the conditional intensity function is the same as the hazard function, i.e.,

$$\lambda(t) = \frac{f(t - t_{N(t_-)})}{1 - F(t - t_{N(t_-)})}, \tag{36}$$

where $t_{N(t_-)}$ is the occurrence time of the last event before t and F is the cumulative probability function of f.

The following probability functions are often chosen as the the renewal densities.

1. The gamma density

$$f(u; k, \theta) = u^{k-1} \frac{e^{-u/\theta}}{\theta^k \, \Gamma(k)}, \quad \text{for } u \geq 0 \text{ and } k, \theta > 0 \,, \tag{37}$$

with a hazard function

$$h(u; k, \theta) = \frac{u^{k-1} \, \mathrm{e}^{-\frac{u}{\theta}}}{\theta^k \Gamma_{\frac{u}{\theta}}(k, \frac{u}{\theta})} \,, \tag{38}$$

where θ is called the scale parameter, k the shape parameter, and Γ and Γ_α are the gamma and the incomplete gamma functions, respectively, defined by $\Gamma(x) = \int_0^\infty t^{x-1} \mathrm{e}^{-t} \, \mathrm{d}t$ and $\Gamma_\alpha(x) = \int_\alpha^\infty t^{x-1} \mathrm{e}^{-t} \, \mathrm{d}t$.
2. The log-normal density function is

$$f(u; \mu, \sigma) = \frac{1}{u\sigma\sqrt{2\pi}} e^{-\frac{(\ln(u)-\mu)^2}{2\sigma^2}}, \quad \text{for } u \geq 0, \tag{39}$$

with a hazard function

$$h(u; \mu, \sigma) = \frac{2}{u\sigma\sqrt{2\pi}} \frac{e^{-\frac{(\ln(u)-\mu)^2}{2\sigma^2}}}{1 - \mathrm{erf}\left[\frac{\ln(u)-\mu}{\sigma\sqrt{2}}\right]} \,, \tag{40}$$

where μ and σ are the mean and standard deviation of the variable's natural logarithm[1]. and erf is the error function.

3. The Weibull distribution is one of the most widely used lifetime distributions in reliability engineering. It is named after Waloddi Weibull who described it in detail in 1951 (*Weibull* 1951). The probability density function of a Weibull random variable X is

$$f(u;\mu,k) = \begin{cases} \frac{k}{\mu}\left(\frac{u}{\mu}\right)^{k-1} e^{-(u/\mu)^k} & u \geq 0, \\ 0 & u < 0, \end{cases} \tag{41}$$

where $k > 0$ is the shape parameter and $\mu > 0$ is the scale parameter of the distribution. Its cumulative distribution function is a stretched exponential function

$$F(u;\mu,k) = 1 - e^{-(u/\mu)^k}$$

and its hazard function is

$$h(u;k,\mu,k) = \frac{k}{\mu}\left(\frac{u}{\mu}\right)^{k-1}.$$

4. *Kagan and Knopoff* (1987) used the inverse Gaussian distribution to model the evolution of stress as a random walk with a tectonic drift. Later on, a particularly interesting renewal model, called the *Brownian passage-time* model, was introduced by *Matthews et al.* (2002) and *Ellsworth et al.* (1999) based on the properties of the Brownian relaxation oscillator (BRO). In this conceptual model of *Matthews et al.* (2002), the loading of the system has two components: (1) a constant-rate loading component, λt, and (2) a random component, $\epsilon(t) = \sigma W(t)$, that is defined as a Brownian motion (where W is a standard Brownian motion and σ is a nonnegative scale parameter). The Brownian perturbation process for the state variable $X(t)$ (see Figure 2 in *Matthews et al.* (2002)) is defined as:

$$X(t) = \lambda t + \sigma W(t).$$

An event will occur when $X(t) \geq X_f$; event times are seen as "first passage" or "hitting" times of Brownian motion with drift. The BRO are a family of stochastic renewal processes defined by four parameters: the drift or mean loading (λ), the perturbation rate (σ^2), the ground state (X_0), and the failure state (X_f). On the other hand, the recurrence properties of the BRO (response times) are described by a Brownian passage-time distribution which is characterized by two parameters: (1) the mean time or period between events, (μ), and (2)

---

[1] If X is a normal random variable with mean μ and variance σ^2, then $Y = \exp X$ has a log-normal distribution with a density of (39).

the aperiodicity of the mean time, α, which is equivalent to the coefficient of variation (defined in as the ratio of the variance to the mean occurrence time). The probability density for the Brownian passage-time (BPT) model is given by:

$$f(t; \mu, \alpha) = \left(\frac{\mu}{2\pi\alpha^2 t^3}\right)^{\frac{1}{2}} \exp\left\{-\frac{(t-\mu)^2}{2\alpha^2\mu t}\right\}, \quad t \geq 0 \tag{42}$$

with a cumulative probability function

$$F(t; \mu, \alpha) = \Phi\left(\frac{t-\mu}{\alpha\sqrt{\mu x}}\right) + \exp\left(\frac{2}{\alpha^2}\right)\Phi\left(-\frac{t+\mu}{\alpha\sqrt{\mu x}}\right), \tag{43}$$

where Φ is the cumulative probability distribution function of a standard normal random variable, i.e., F is the inverse Gaussian or Wald distribution.

Forward recurrence time In practice, we usually do not start our observation at the occurrence of an event. Then we are interested in the distribution of waiting times until the next event starting from this starting time, i.e., the forward recurrence time from an arbitrary time. This problem is equivalent to finding the distribution of the distance between an arbitrary point to the next event in the time line, and can be solved in three steps.

1. Assume that the length of each inter-event interval Δ has a density f with finite mean μ_0, i.e., $\int_0^\infty u\, f(u)\, \mathrm{d}u = \mu_0$.
2. The arbitrary time point falls in each inter-event interval Δ with a probability proportional to its length. That is to say, this arbitrary time is uniformly distributed on the time axis and has more chances to fall in larger intervals; or, $\Pr\{$Arbitrary point falls in $\Delta \mid \Delta \in [z, z + \mathrm{d}z]\} \propto z$.
3. From 2, given the length $\Delta = z$ of the inter-event interval that the arbitrary time falls in, the probability that the forward recurrence time X is less than t is 1 if $z < t$, or t/z if $z \geq t$. In other words, the probability that the forward recurrence time X is greater than t is 0 if $z < t$, or $(z - t)/z$ if $z \geq t$. In summary,

$$\begin{aligned}
\Pr\{X \geq t\} &= \int_t^\infty \Big[\Pr\{\text{Arbitrary point falls in } \Delta \mid \Delta \in [z, z + \mathrm{d}z]\} \\
&\qquad \times \Pr\{\Delta \in [z, z + \mathrm{d}z]\} \times \Pr\{X > t \mid \Delta \in [z, z + \mathrm{d}z)\}\Big] \\
&\propto \int_t^\infty z\, f(z)\, \frac{(z-t)}{z}\, \mathrm{d}z \\
&= \int_t^\infty (z - t) f(z)\, \mathrm{d}z \\
&= \mu^c(t) - t[1 - F(t)]
\end{aligned} \tag{44}$$

where

$$\mu^c(t) = \int_t^\infty z\, f(z)\, \mathrm{d}z.$$

Similarly, it can be shown that

$$\Pr\{X < t\} \propto t[1 - F(t)] + \mu(t) \tag{45}$$

where

$$\mu(t) = \int_0^t z\, f(z)\, \mathrm{d}z.$$

We find the normalizing factor in (44) by using

$$1 = \Pr\{X < \infty\} = \Pr\{X \geq 0\}, \tag{46}$$

which gives

$$\Pr\{X < t\} = \frac{t[1 - F(t)]}{\mu_0} + \frac{\mu(t)}{\mu_0} \tag{47}$$

and

$$\Pr\{X \geq t\} = \frac{\mu^c(t)}{\mu_0} - \frac{t\,[1 - F(t)]}{\mu_0}, \tag{48}$$

where $\mu_0 = \mathbf{E}\,\Delta = \int_0^\infty z\, f(z)\, \mathrm{d}z$.

It is easy to see that the phrase "the forward recurrence time" can be replaced by "the backward recurrence time" in the above discussion. That is to say, the backward recurrence time Y has the same distribution as the forward distribution function. The density function w_f of the forward recurrence times is obtained by taking the derivative of (47), i.e.,

$$w_f(t) = \frac{1 - F(t)}{\mu_0}, \tag{49}$$

with corresponding hazard function

$$h_f(t) = \frac{w_f(t)}{\Pr\{X > t\}} = \frac{1 - F(t)}{\mu^c(t) - t[1 - F(t)]}. \tag{50}$$

Table 1 Historical Earthquakes at Nankai Trough (DATA FILE: *example2.data*)

Occurrence times	Years
t1	684
t2	887
t3	1099
t4	1233
t5	1361
t6	1498
t7	1605
t8	1707
t9	1854
t10	1946

Table 2 Results from fitting renewal models to the Nankai Trough data

Model	MLE	$\log L$	AIC
Poisson	$\lambda = 7.092 \times 10^{-3}$	-59.488	120.975
Gamma	$k = 7.768, \theta = 16.865$	-54.156	112.312
log-normal	$\mu = 4.813, \sigma = 0.365$	-54.580	113.160
Weibull	$k = 2.886, \mu = 144.849$	-55.176	114.351
BPT	$\mu = 123.881, \alpha = 0.419$	-55.978	113.954

Likelihood function Suppose that the realization of a renewal process is $\{t_1, t_2, \cdots, t_n\}$ in $[0, T]$, and we do not know what happens before time 0. The likelihood function can be written as

$$L(0, T; \cdot) = w_f(t_1) \times \prod_{i=2}^{n} f(t_i - t_{i-1}) \times F(T - t_n). \tag{51}$$

Another form of the likelihood is through the conditional intensity function,

$$\log L(0, T; \cdot) = \sum_{i:\, t_i \in [0,T]} \log \lambda(t_i) - \int_0^T \lambda(u)\, \mathrm{d}u. \tag{52}$$

It can be shown that both forms are equivalent.

Example 2 Table 1 [Data file: *example2.data*] lists 10 great earthquakes of $M8+$ that have occurred in the Nankai Trough south of Japan since 600 A.D., taken from Table 10.5 in *Utsu* (1999). *Ogata* (2002) analyzed this data set by using recurrence models and Bayesian methods to evaluate the uncertainties in the estimated occurrence times. Here, as a simple illustration, we fit the Poisson and the four renewal models discussed in this section. The estimated parameters and the corresponding likelihood and AIC scores are listed in Table 2, indicating that, in this simple example, the Gamma model fits the data the best.

Among the models mentioned in this section, the BPT model has become the most popular one and has been used in obtaining some interesting results (see, e.g., *Ogata* 2002 and *Nomura et al.* 2011). However, there is no clear evidence showing that one model is superior to the others. This might be caused by the relatively large standard errors of the parameter estimates that are due to very few data points in the historical catalogs. Another reason may due to the nature of renewal processes (e.g., *MacFadden and Weissblum* 1965; *Daley and Vere-Jones* 2003, Pages 77–79): The superposition of independent nontrivial renewal processes does not result in a renewal process unless all of the superposed processes are simply Poisson. This property requires that the renewal process is applied to a fault system that is relatively disjoint in behavior from the surrounding environment, or is loaded by a constant tectonic stressing rate. However, the identification of such fault systems seems technically difficult in practice.

3.4 Stress release models

Based on the elastic rebound theory (*Reid* 1910), the stress release model was introduced in a series of papers by Vere-Jones and others (see, e.g., *Zheng and Vere-Jones* 1991, 1994). It was extended to the linked stress release model by *Shi et al.* (1998), *Liu et al.* (1998), *Lu et al.* (1999), *Lu and Vere-Jones* (2000), and *Bebbington and Harte* (2001, 2003). This model assumes that the stress level in a certain region is gradually built up, linearly with time by tectonic movements, and drops down suddenly coinciding with earthquakes, i.e., the stress level at t, $X(t)$, can be written as

$$X(t) = X(0) + \rho t - S(t), \tag{53}$$

where $X(0)$ is the initial stress level, ρ is the constant loading rate from the tectonic movements, and $S(t)$ is the accumulated stress release from earthquakes in the period $[0, t)$, namely

$$S(t) = \sum_{i:\, t_i < t} s_i. \tag{54}$$

It is assumed that the stress release related to individual earthquakes scales with the magnitude m_i according to

$$s_i = 10^{0.75 m_i}, \tag{55}$$

which is hinted from the Benioff strain. The conditional intensity is a monotonically increasing function of the stress level, usually chosen as

$$\lambda(t) = \Psi(X) = \mathrm{e}^{\gamma X} = \exp[\gamma(X(0) + \rho t - S(t))] = \exp[a + bt - c\, S(t)], \tag{56}$$

where γ is a single parameter representing the strength and heterogeneity of the regional crust, and $a = \gamma X(0)$, $b = \rho\gamma$, and $c = \gamma$ are the final model parameters. The assumption of an exponential distribution is in agreement with laboratory experiments of stress corrosion where the mean waiting time until fracture can be approximated by an exponential function of negative applied stress (*Scholz* 2002).

The amount of stress s_i released by an earthquake has a distribution with a density function of $\xi(s)$ and cumulative probability function $\Xi(s)$, which is, for simplification, assumed to be independent from the stress level. That is to say, the full model has a conditional intensity function in the form of

$$\lambda(t, s) = \Psi[X(t)]\,\xi(s). \tag{57}$$

Given the observations $\{(t_i, s_i) : i = 1, 2, \cdots, n\}$ in a time period $[T_1, T_2]$, the log-likelihood function for the full model is

$$\log L(T_1, T_2; \cdot) = \sum_{i=1}^{n} \log \lambda_i(t_i, s_i) - \int_{T_1}^{T_2} \int_{0}^{\infty} \lambda(u, v)\,\mathrm{d}v\,\mathrm{d}u$$
$$= \sum_{i=1}^{n} \log \Psi(X(t_i)) - \int_{T_1}^{T_2} \Psi(X(u))\,\mathrm{d}u + \sum_{i=1}^{n} \log \xi(s_i). \tag{58}$$

In the above equation, the first two terms are from the usual likelihood for pure temporal processes, and the third term represents the joint distribution of stress reductions caused by the earthquakes.

Usually, we consider the following three types of distributions for S: the Pareto, the truncated Pareto and the tapered Pareto distributions:

1. The Pareto distribution of S corresponds to the case where the magnitude distribution follows the Gutenberg-Richter magnitude-frequency relation, i.e., the exponential distribution. It has a probability density function

   $$\xi(s) = \frac{k-1}{S_0}\left(\frac{s}{S_0}\right)^{-k}; \quad s \geq S_0, \;\; k > 1, \tag{59}$$

   where k is linked with the Gutenberg-Richter b-value by $k = \frac{4}{3}b + 1$.

2. The truncated Pareto distribution of s corresponds to the case where the magnitude distribution follows the truncated Gutenberg-Richter magnitude-frequency relation. It has a probability density function

   $$\xi(s) = \frac{k-1}{S_0\left[1 - \left(\frac{S_u}{S_0}\right)^{1-k}\right]}\left(\frac{s}{S_0}\right)^{-k}, \quad S_0 \leq s \leq S_u, \tag{60}$$

   where S_u and S_0 are the upper and lower thresholds, respectively.

3. The tapered Pareto (or Kagan) distribution (see, *Vere-Jones et al.* 2001; *Kagan and Schoenberg* 2001) has a cumulative distribution function

$$\Xi(s) = 1 - \left(\frac{s}{S_0}\right)^{-k+1} \exp\left(\frac{S_0 - s}{S_c}\right), \quad S_0 \leq s < \infty, \tag{61}$$

and a density

$$\xi(s) = \left(\frac{k-1}{s} + \frac{1}{S_c}\right)\left(\frac{s}{S_0}\right)^{-k+1} \exp\left(\frac{S_0 - s}{S_c}\right), \quad S_0 \leq s < \infty, \tag{62}$$

where S_c is a parameter governing the strength of the exponential taper affecting frequency of large events.

All three distributions are illustrated in Fig. 1.

The magnitude-frequency distribution of earthquake is well described by the Gutenberg-Richter magnitude-frequency relation (*Gutenberg and Richter* 1944), which takes the form of an exponential distribution, i.e.,

$$\Pr\{\text{magnitude} > m\} = 10^{-b(m-m_c)}, \quad m \geq m_c, \tag{63}$$

where b is the so-called Gutenberg-Richter b-value. To reduce the possibility of extremely large earthquakes in simulation, the following truncated form is also often adopted:

$$\Pr\{\text{magnitude} > m\} = \begin{cases} 10^{-b(m-m_c)} = e^{-\beta(m-m_c)}, & \text{if } m_c \leq m \leq M_{max}, \\ 0 & \text{otherwise,} \end{cases}$$

where M_{max} is the upper bound of magnitudes.

Example 3 The following example is taken from *Zheng and Vere-Jones* (1994) where the stress release model was fit to the historical catalog from North China during the period from 1480 to 1992. Table 3 (Data file *example3.data*) lists this historical catalog from North China, Figure 2 shows the division of subregions, and Table 4 gives the corresponding results. Figure 3 shows the temporal variations of the conditional intensity functions in each subregion.

4 Temporal clustering models

On short time scales, seismicity has been shown to be clustered. The terms foreshock, main shock, aftershock, and earthquake swarm are concepts related the clustering of
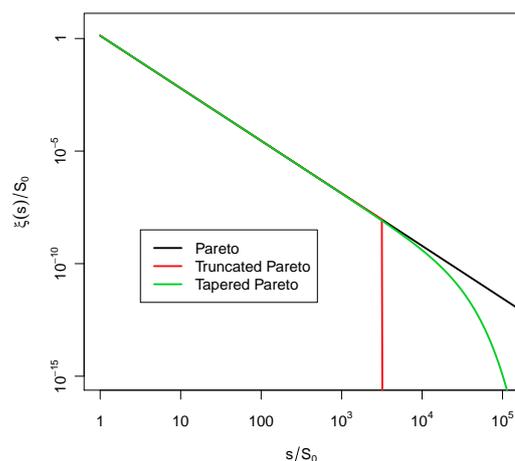
Fig. 1 Probability density functions $\xi(s)$ of the Pareto, the truncated Pareto, and the tapered Pareto distribution. Here we take $k = 7/3(b = 1)$, $S_c = 10^4 S_0$, and $S_u = 10^{3.5} S_0$.
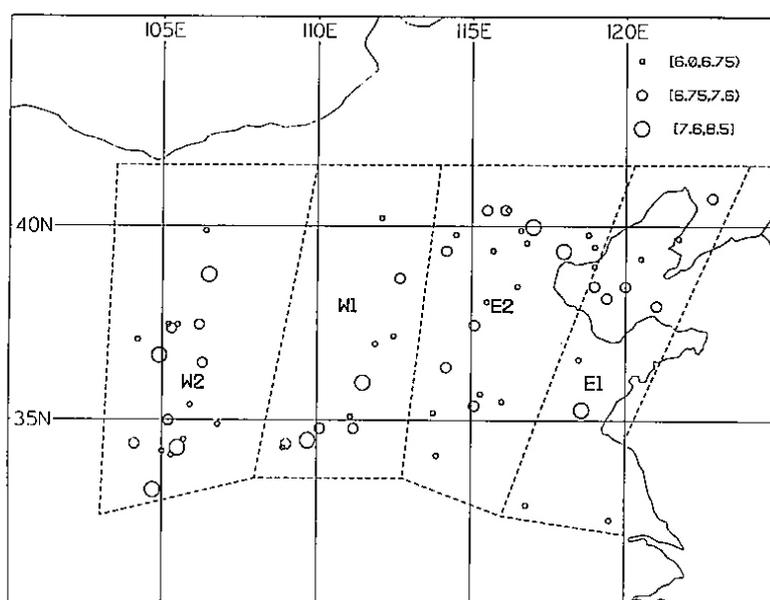
Fig. 2 Subregions and earthquake locations for the North China dataset [from *Zheng and Vere-Jones* (1994)].

the earthquake events. They are often used loosely in the many studies on seismicity. However, a precise definition of each is generally not agreed, and indeed, it is often difficult to correctly label observed events prospectively as they occur.

Short-term statistical seismicity models developed in the last century, including the Omori-Utsu formula, the multiple Omori-Utsu formula, and the epidemic-type

Table 3 DATA FILE: *example3.data* Historical large earthquake from north China, 1480-1996, reprinted from Zheng and Vere-Jones (1994).

Date	Lat.	Long.	Mag.	Reg.	Date	Lat.	Long.	Mag.	Reg.
1484.01.29	40.40	116.10	6.70	W1	1487.08.10	34.30	108.90	6.20	E2
1501.01.19	34.80	110.10	7.00	E2	1502.10.17	35.70	115.30	6.50	W1
1536.10.22	39.60	116.80	6.00	W1	1548.09.13	38.00	121.00	7.00	W2
1556.01.23	34.50	109.70	8.00	E2	1561.07.25	37.50	106.20	7.20	E1
1568.04.25	34.40	109.00	6.70	E2	1568.05.15	39.00	119.00	6.00	W2
1573.01.10	34.40	104.10	6.70	E1	1587.04.10	35.20	113.80	6.00	W1
1597.10.06	38.50	120.00	7.00	W2	1604.10.25	34.20	105.00	6.00	E1
1614.10.23	37.20	112.50	6.50	E2	1618.05.20	37.00	111.90	6.50	E2
1618.11.16	39.80	114.50	6.50	W1	1622.03.18	35.50	116.00	6.00	W1
1622.10.25	36.50	106.30	7.00	E1	1624.02.10	32.40	119.50	6.00	W2
1624.04.17	39.80	118.80	6.20	W1	1624.07.04	35.40	105.90	6.00	E1
1626.06.28	39.40	114.20	7.00	W1	1627.02.15	37.50	105.50	6.00	E1
1634.01.-	34.10	105.30	6.00	E1	1642.06.30	35.10	111.10	6.00	E2
1654.07.21	34.30	105.50	8.00	E1	1658.02.03	39.40	115.70	6.00	W1
1665.04.16	39.90	116.60	6.50	W1	1668.07.25	35.30	118.60	8.60	W2
1679.09.02	40.00	117.00	8.00	W1	1683.11.22	38.70	112.70	7.00	E2
1695.05.18	36.00	111.50	8.00	E2	1704.09.28	34.90	106.80	6.00	E1
1709.10.14	37.40	105.30	7.50	E1	1718.06.29	35.00	105.20	7.50	E1
1720.07.12	40.40	115.50	6.70	W1	1730.09.30	40.00	116.20	6.50	W1
1739.01.03	38.80	106.50	8.00	E1	1815.10.23	34.80	111.20	6.70	E2
1820.08.03	34.10	113.90	6.00	W1	1829.11.19	36.60	118.50	6.00	W2
1830.06.12	36.40	114.20	7.50	W1	1831.09.28	32.80	116.80	6.20	W2
1852.05.26	37.50	105.20	6.00	E1	1861.07.19	39.70	121.70	6.00	W2
1879.07.01	33.20	104.70	8.00	E1	1882.12.02	38.10	115.50	6.00	W1
1885.01.14	34.50	105.70	6.00	E1	1888.06.13	38.50	119.00	7.50	W2
1888.11.02	37.10	104.20	6.20	E1	1920.12.06	36.70	104.90	8.50	E1
1922.09.29	39.20	120.50	6.50	W2	1937.08.01	35.40	115.10	7.00	W1
1945.09.23	39.50	119.00	6.20	W1	1966.03.22	37.50	115.10	7.20	W1
1967.03.27	38.50	116.50	6.30	W1	1969.07.18	38.20	119.40	7.40	W2
1975.02.04	40.70	122.80	7.30	W2	1976.04.06	40.20	111.10	6.20	E2
1976.07.28	39.40	118.00	7.80	W1	1976.09.23	39.90	106.40	6.20	E1
1979.08.25	41.20	108.10	6.00	E1	1989.10.18	40.00	113.70	6.00	E2

Table 4 Results from fitting the stress release model to the historical earthquakes from each of the four subregions in North China. N represents the number of events in each subregion.

	a	b	c	$\log L$ (SRM)	$\log L$ (Poisson)	ΔAIC	N
E1	-5.397	0.0209	0.571	-52.04	-56.76	-5.44	11
E2	-3.900	0.0176	0.880	-83.34	-87.58	-4.48	22
W1	-2.764	0.0035	1.731	-50.36	-52.99	-1.24	11
W2	-5.152	0.0238	0.410	-91.97	-77.84	-7.75	18

aftershock sequence (ETAS) model, all emphasize the evolutionary nature of a developing earthquake cluster. In this section, we give an introduction to these models. The models described in this section do not have a spatial component. Hence, there is an implicit assumption that the spatial area is sufficiently small so that any given
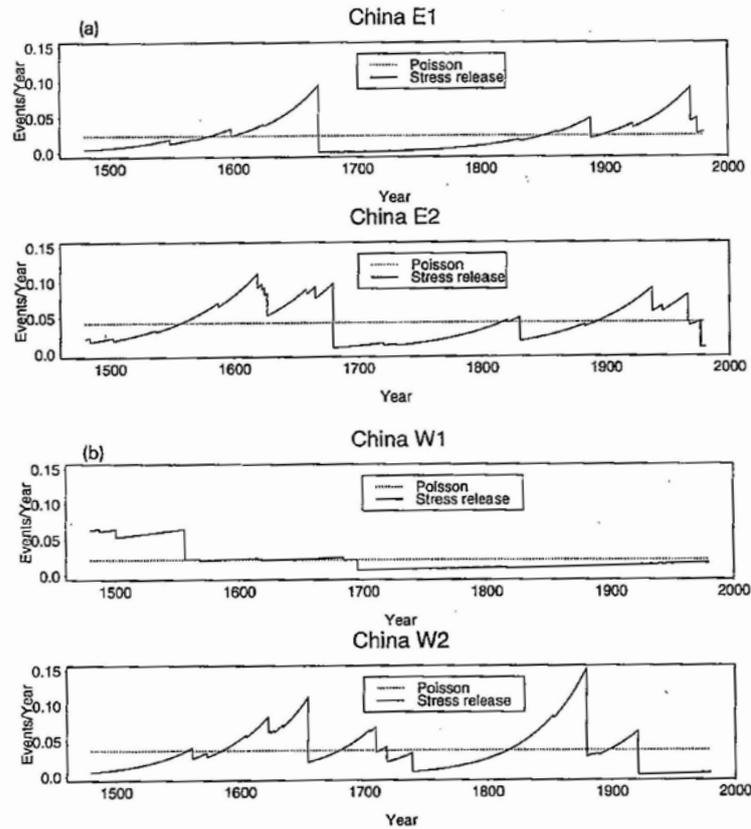
Fig. 3 Conditional intensity curves for the four subregions of the North China dataset [from *Zheng and Vere-Jones* (1994)].

event can conceivably interact with all following events, regardless of their spatial locations.

4.1 The Omori-Utsu formula

The Omori-Utsu formula describes the decay of the aftershock frequency with time after a mainshock as an inverse power law. When *Omori* (1894) was studying the aftershocks of the 1891 $M_s 8.0$ Nobi earthquake, he first tried to use an exponential decay function to fit the data, but unsatisfactory results were obtained. Then he found that the number of aftershocks occurring each day can be well described by the equation

$$n(t) = K(t + c)^{-1}, \tag{64}$$

where t is the time from the occurrence of the mainshock, K and c are constants.

Utsu (1957) postulated that the decay of the aftershock numbers could vary, and showed that

$$n(t) = K(t + c)^{-p} \tag{65}$$

yields better fitting results. Equation (65) is now called the modified Omori formula or Omori-Utsu formula.

The Omori-Utsu formula has been used extensively to analyze, model and forecast aftershock activity. *Utsu et al.* (1995) reviewed the values of p for more than 200 aftershock sequences and found that it ranges between 0.6 and 2.5 with a median of 1.1. He found no clear relationship between estimates of p-values and mainshock magnitudes.

In the context of statistical seismology, we usually make use of (65) in the form of a conditional intensity function, given by

$$\lambda(t) = \frac{K}{(t + c)^p}. \tag{66}$$

When the magnitude distribution of the aftershocks is also considered, the conditional intensity for the full model is written as

$$\lambda(t, m) = \frac{K\, s(m)}{(t + c)^p}, \tag{67}$$

where $s(m)$ is the magnitude probability density function and usually takes the form of the Gutenberg-Richter magnitude-frequency relation. This model has been used by Reasenberg and Jones for forecasting aftershock activity (see, e.g., *Reasenberg and Jones* 1989, 1994). Therefore, some researchers also call (67) the Reasenberg-Jones model.

Likelihood of the single and multiple Omori-Utsu formulas Suppose that the occurrence time of the mainshock is 0. As discussed in Section 2, given observations $N = \{(t_i, m_i) : i = 1, 2, \cdots, n\}$ in the the time interval $[S, T]$, $T > S \geq 0$, the log-likelihood function for the model of the Omori-Utsu formula can be written in the form

$$\log L(N; S, T) = \sum_{i:\, t_i \in [S,T]} \log \lambda(t_i, m_i) - \int_{\mathcal{M}} \int_S^T \lambda(u, m)\, \mathrm{d}u\, \mathrm{d}m$$

$$= \sum_{i:\, t_i \in [S,T]} \log \lambda(t_i) - \int_S^T \lambda(u)\, \mathrm{d}u + \sum_{i:\, t_i \in [S,T]} \log s(m_i), \tag{68}$$

where the first two terms on the righthand side represent the contribution to the likelihood from the occurrence times and the third term represents the contribution
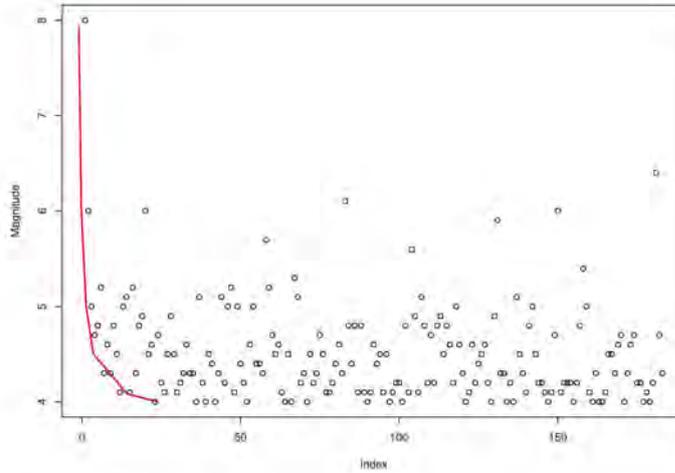
Fig. 4 A plot of magnitudes against indices of the aftershocks of the 2008 Wenchuan earthquake in China. The red line indicates the estimation of the completeness magnitude.

from the magnitudes. The parameters in the Omori-Utsu formula can be estimated from maximizing the temporal part of the likelihood (first two terms in the righthand side)

$$
\sum_{i:\,t_i \in [S,T]} \log \lambda(t_i) - \int_S^T \lambda(u)\,\mathrm{d}u
$$

$$
= N \log K - p \sum_{i:\,t_i \in [S,T]} \log(t_i + c) - \frac{K}{p-1}\left[(S+c)^{1-p} - (T+c)^{1-p}\right]. \quad (69)
$$

Example 4 The aftershock sequence of the 2008 $M_S 8.0$ Wenchuan earthquake, China: Table 5 [Data file *example4.data*] lists the mainshock and the aftershocks ($M_S \geq 4.0$) of the Wenchuan earthquake (2009-5-12, $M_s 8.0$), occurring within 25 days after the mainshock. Since aftershocks immediately after a large mainshock are usually missing due to detection and recording problems (see, e.g., *Kagan and Jackson* 1995), the Omori-Utsu formula should only be fit to a period where the records in the considered magnitude range are complete. A simple and direct way to detect this starting point is by plotting each event magnitude against its catalog sequential number of event. As shown in Figure 4, we arbitrarily take $T_0 = 0.3$ days, corresponding to the occurrence time of the 40th event. The estimated parameters from fitting the Omori-Utsu formula are $\hat{K} = 45.228$, $\hat{c} = 0.129$ days, and $\hat{p} = 1.107$; with $\log L = 270.575$.
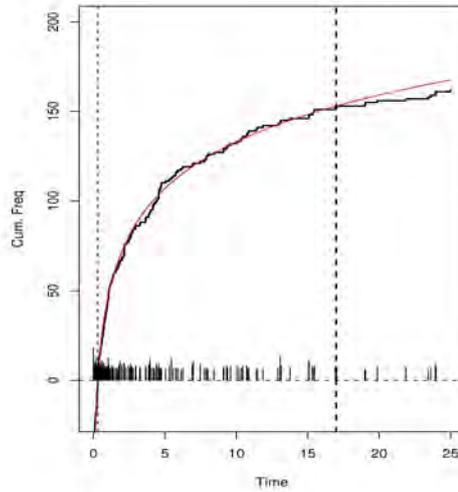
Fig. 5 Comparison of the observed and modeled cumulative number of $M \geq 4$-aftershocks of the Wenchuan earthquake: observation (black) and Omori-Utsu formula (red).

Multiple Omori-Utsu formula It is often observed that not only mainshocks trigger aftershocks, but also large aftershocks may trigger their own aftershocks. To model such phenomena, *Utsu* (1970) used the following model

$$\lambda(t) = K/(t - t_0 + c)^{-p} + \sum_{i=1}^{N_T} \frac{K_i H(t - t_i)}{(t - t_i + c_i)^{-p_i}}, \qquad (70)$$

where t_0 is the occurrence time of the mainshock, t_i, $i = 1, \ldots N_T$, define the occurrence times of the triggering aftershocks and H is the Heaviside function. The likelihood for the multiple Omori-Utsu formula is sightly more complicated than for the simple Omori-Ustu formula, but can be written in a similar way.

Example 5 This example is on the 1965 Rat Island earthquake of $M_W 8.7$ and its aftershocks, and is taken from *Ogata and Shimazaki* (1984). Earthquakes within the range $170° - 180°E$ and $48° - 55°N$ and with magnitudes $m_b \geq 4.7$ are selected. An ordinary log-log plot of the daily number of events against time is time shown in Figure 6. The results from fitting a simple Omori-Utsu formula are

$$\hat{K} = 85.088, \quad \hat{c} = 0.204, \quad \hat{p} = 1.055, \quad AIC = -811.4,$$

and from the multiple formula, by assuming that the largest aftershock of $M_W 7.6$ also has its own aftershocks, are

$$\hat{K} = 82.284, \quad \hat{K}_1 = 6.117, \quad \hat{c} = \hat{c}_1 = 0.204, \quad \hat{p} = \hat{p}_1 = 1.055, \quad AIC = -873.4.$$
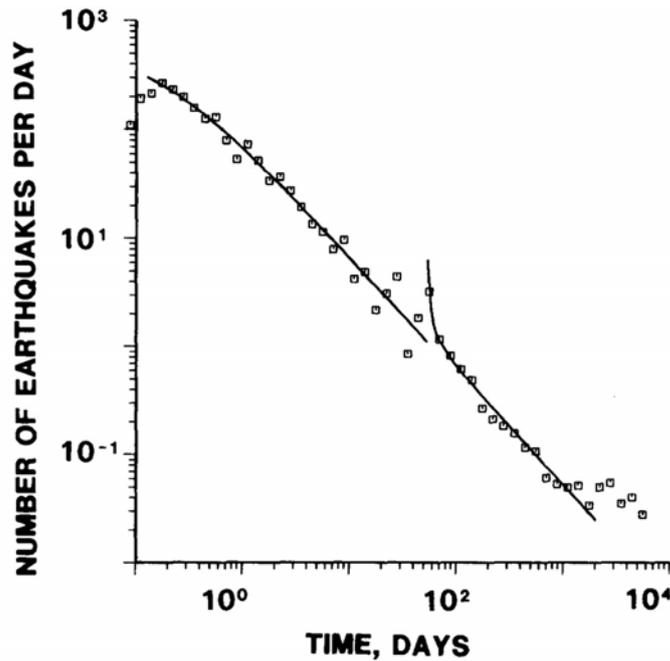
Fig. 6 An ordinary log-log plot of the number of aftershocks of the 1965 M_W 8.7 Rat Island event per day against time (*Ogata and Shimazaki* 1984). The curve represents the intensity function fitted to observed data points marked by the squares.

Here the AIC criteria selects the model where $c = c_1$ and $p = p_1$ as the best model among the class of models with $N_T = 1$.

One difficulty in applying the multiple Omori-Utsu formula is to determine which earthquakes are triggering events. The largest aftershocks often have secondary aftershocks, but not always. We can use the techniques of residual analysis in Section 5, as a diagnostic tool, to find out which events have secondary aftershocks.

4.2 The Epidemic-Type Aftershock Sequence (ETAS) model

In most cases, aftershock activity consists of secondary aftershock clustering as described by the multiple Omori-Utsu formula (70). However, even though possible triggering events could potentially be determined through a residual analysis, separating triggering earthquakes from the others is difficult. *Ogata* (1988) generalized this model by not needing to make a distinction between triggering events and the other events. He proposed that each event, irrespective of whether it is a small or a big event, can in principle trigger its own offspring. More precisely, the temporal

conditional intensity of this model is

$$\lambda(t) = \mu + \sum_{i:\,t_i < t} \kappa(m_i)\, g(t - t_i), \tag{71}$$

or, the full form of the condition intensity for this marked point process is

$$\lambda(t, m) = s(m) \left[\mu + \sum_{i:\,t_i < t} \kappa(m_i)\, g(t - t_i) \right], \tag{72}$$

where $s(m) = \beta\, e^{-\beta(m - m_0)}, m \geq m_0$, is the p.d.f. form of the Gutenberg-Richter relation, m_0 being the magnitude threshold, $\kappa(m) = A \exp[\alpha(m - m_0)]$ is the mean number of events directly   triggered by an event of magnitude m and $g(u) = (p - 1)(1 + t/c)^{-p}/c$ is the probability density function (p.d.f.) of the time difference between the parent event and its children, i.e., the p.d.f. form of the Omori-Utsu formula. *Ogata* (1988) named (71) the Epidemic-Type Aftershock Sequence (ETAS) model, based on an analogy with the spread of epidemics. The ETAS model also belongs to the more general class of self-exciting Hawkes processes (*Hawkes* 1971a,1971b; *Hawkes and Oakes* 1974).

Example 6 [Data file: *example6.data*] The aftershock sequence of the Darfield earthquake, New Zealand ($M_W 7.1$, 2010-9-4). The dataset is taken from the New Zealand catalog compiled by GNS, in the region ($43.2 - 44.0°S$, $171.0 - 173.5°E$) and in the time interval from 1985-1-1 to 2011-7-13. The magnitude threshold is $M3.8$. By using the ETAS fitting program in the SASeis software (*Ogata* 2006), the MLEs of the model parameters are $\mu = 0.001414$ (events/day), $K = 10.32$, $c = 0.006006$ (day), $\alpha = 1.832$, $p = 1.0167$. Figure 7 shows the conditional intensity for the fitted ETAS model.

Criticality and branching ratio A condition for stability of the ETAS model is that the process must be subcritical. In data analysis this important condition justifies whether estimated model parameters are reasonable or irrational. Also, when extending the ETAS model or specifying an explicit form of an ETAS-like branching model, the stability conditions must be satisfied. Forecasting with model parameters that specify a supercritical model definitely overestimate the earthquake risk in the medium- or long-term future.

The stability of the ETAS model or a more general branching process is closely related to the concepts of criticality and branching ratio. These two concepts are identical for the ETAS model, but not the same for general branching processes. To illustrate the differences between them, we start our discussions on criticality and the branching ratio with more general assumptions. Readers who would like to know more details are recommended to read through the following box.
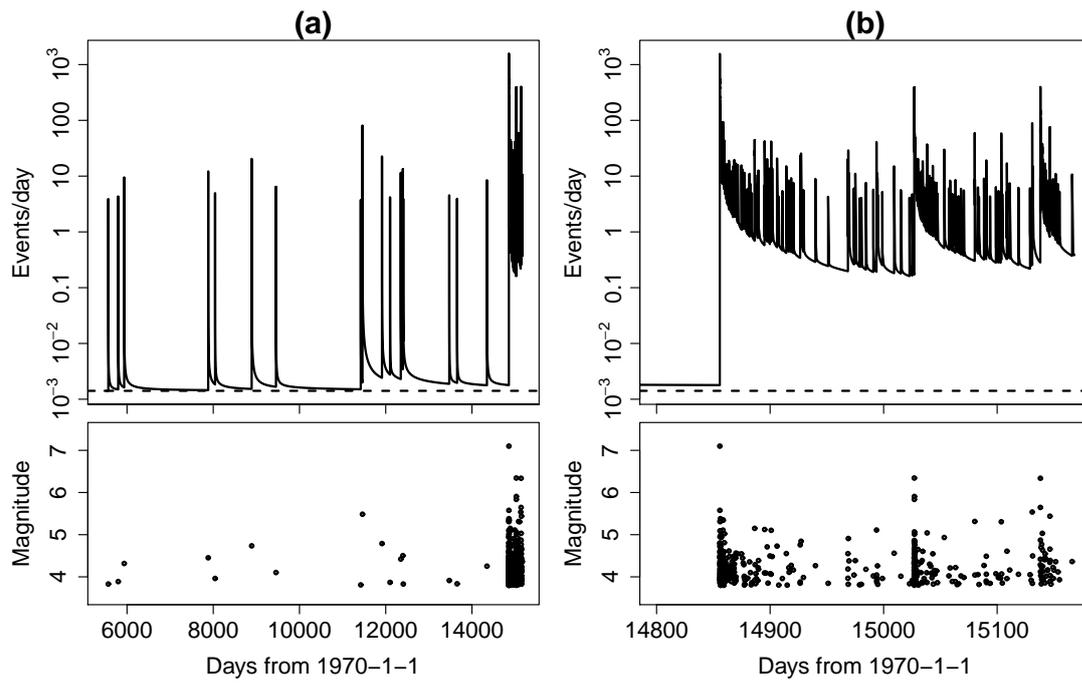
Fig. 7 Temporal variation of the conditional intensity from fitting the ETAS model to the Darfield earthquake sequence for (a) the whole time period, and for (b) the time period of the aftershock sequence only. The magnitudes against occurrence times are plotted in lower panels.

We start our discussions with more general branching models with the following assumptions:

(1) The mean number of direct children produced by an event of magnitude m' is a Poisson random variable with a mean of $\kappa(m')$;
(2) The magnitudes of the children from a parent of magnitude $m\prime$ are independent and identical distributed (i.i.d.) random variable samples (r.v.s) with a density $s(\cdot \mid m')$;
(3) The intensity function of background events is $G_0(m) = C\,s_0(m)$, where $s_0(m)$ is the p.d.f. of the magnitude density for background events and C is a constant.

It is easy to see that the intensity function for the first generation (events of direct children from background events) is

$$G_1(m) = \int_S \kappa(m')\,s(m \mid m')\,G_0(m')\,\mathrm{d}m',$$

where S represents the range of magnitudes, and for the second generation is

$$G_2(m) = \int_S \kappa(m')\,s(m \mid m')\,G_1(m')\,\mathrm{d}m'.$$

Similarly,

$$G_{n+1}(m) = \int_S \kappa(m')\,s(m \mid m')\,G_n(m')\,\mathrm{d}m' = \int_S K^{[n]}(m, m')\,G_0(m')\,\mathrm{d}m,$$

where $K^{[n]}$ ($n \geq 1$) is defined by the following recursion:

$$K^{[1]}(m, m') = \kappa(m')\,s(m \mid m'),$$
$$K^{[n]}(m, m') = \int \kappa(m^*)\,s(m \mid m^*)\,K^{[n-1]}(m^*, m')\,\mathrm{d}m^*, \quad \text{for } n = 2,\,3,\,\cdots.$$

Suppose that $a(m')$ and $b(m)$ are the left and right eigenfunctions of K corresponding to the maximum eigenvalue ϱ, i.e.,

$$\varrho\,a(m') = \int_S a(m)\,K(m; m')\,\mathrm{d}m, \tag{73}$$

and

$$\varrho\,b(m) = \int_S K(m; m')\,b(m')\,\mathrm{d}m', \tag{74}$$

respectively, satisfying

$$\int_S a(m)\, b(m)\, \mathrm{d}m = 1. \tag{75}$$

Let

$$\Omega(m; m') = a(m')\, b(m), \tag{76}$$

i.e., Ω is the projection operator of K corresponding to ϱ, or

$$\int_S \Omega(m; m^*)\, K(m^*; m')\, \mathrm{d}m^* = \int_S K(m; m^*)\, \Omega(m^*; m')\, \mathrm{d}m^*$$
$$= \varrho\, \Omega(m; m'). \tag{77}$$

When both κ and $s(m \mid m')$ are stepwise continuous functions, i.e., linear integral equations (73) and (74) can be viewed as the continuum limit of eigenvalue equations of the form

$$\sum_j M_{i,j} v_j = \varrho v_i,$$

the maximum eigenvalue is separated from the others, and when $n \to \infty$,

$$\frac{K^{[n]}}{\varrho^n} \to \Omega, \tag{78}$$

and thus

$$G_n(m) \to \varrho^n \int_S \Omega(m; m')\, G_0(m')\, \mathrm{d}m'. \tag{79}$$

From (79), we can see that if $\varrho < 1$, then $G_n \to 0$ when $n \to \infty$, and that if $\varrho > 1$, then $G_n \to \infty$ when $n \to \infty$. Here ϱ is called the criticality parameter because if $\varrho < 1$, the process is stable; otherwise the population might explode and become infinitely large.

Equation (79) can be rewritten as

$$G_n(m) \to \varrho^n \int_S b(m)\, a(m')\, G_0(m')\, \mathrm{d}m'$$
$$= \varrho^n\, b(m) \int_S a(m')\, G_0(m')\, \mathrm{d}m'$$
$$= \varrho^n\, b(m) \times \mathrm{const}, \tag{80}$$

implying that $b(m)$ is asymptotically proportional to the intensity of the population when $n \to \infty$.

The eigenfunction $a(m')$ can be interpreted as the asymptotic ability in producing offspring, directly and indirectly, from an ancestor $\{m'\}$ because

$$\lim_{n\to\infty} \sum_{i=n}^{\infty} \int_S K^{[i]}(m; m')\, \mathrm{d}m = \lim_{n\to\infty} \sum_{i=n}^{\infty} \varrho^i \int_S \Omega(m; m')\, \mathrm{d}m$$

$$= \lim_{n\to\infty} \sum_{i=n}^{\infty} \varrho^i\, a(m') \int_S b(m)\, \mathrm{d}m$$

$$= \lim_{n\to\infty} \frac{\varrho^n}{1 - \varrho}\, a(m') \times \text{const.} \qquad (81)$$

For the ETAS model, where the magnitude density is separable and the background rate is constant, the eigenvalue equations are

$$\varrho\, a(m') = \kappa(m') \int_S a(m)\, s(m)\, \mathrm{d}m, \quad \text{and} \qquad (82)$$

$$\varrho\, b(m) = s(m) \int_S \kappa(m')\, b(m')\, \mathrm{d}m', \qquad (83)$$

where $\mathcal{M}$ is the magnitude range. We can see that

$$a(m') = C_1\, \kappa(m'), \qquad (84)$$

and

$$b(m) = C_2\, s(m). \qquad (85)$$

Substituting $a(m')$ and $b(m)$ back into (82) and (83), we get

$$\varrho = \int_S \kappa(m)\, s(m)\, \mathrm{d}m. \qquad (86)$$

The <u>branching ratio</u> is defined as the proportion of triggered events in all the events. To obtain the branching ratio for general branching models, consider a model with the following intensity

$$\lambda(t, m) = \mu\, s_0(m) + \sum_{t_i : t_i < t} \kappa(m_i)\, g(t - t_i)\, s(m \mid m_i). \qquad (87)$$

Taking expectations on both sides with respect to time, we get,

$$\bar{\lambda}\, s_1(m) = \mathbf{E}\left[\lambda(t, m)\right] = \mu\, s_0(m) + \mathbf{E}\left[\sum_{t_i : t_i < t} \kappa(m_i)\, g(t - t_i)\, s(m \mid m_i)\right], \qquad (88)$$

where $\bar{\lambda}$ is the total average rate, and $s_1(m)$ is the magnitude density for overall events. The expectation of the summation in the right-hand side can be written as

$$\mathbf{E}\left[\sum_{t_i:\,t_i<t}\kappa(m_i)\,g(t-t_i)\,s(m\mid m_i)\right]$$

$$=\mathbf{E}\left[\int_S\int_{-\infty}^{t}\kappa(m^*)\,g(t-u)\,s(m\mid m^*)\times\bar{\lambda}\,s_1(m^*)\,\mathrm{d}u t\,\mathrm{d}m^*\right]$$

$$=\bar{\lambda}\int_S\kappa(m^*)\,s(m\mid m^*)\,s_1(m^*)\,\mathrm{d}m^*,$$

and hence

$$\bar{\lambda}\,s_1(m)=\mu\,s_0(m)+\bar{\lambda}\int_S\kappa(m^*)\,s(m\mid m^*)\,s_1(m^*)\,\mathrm{d}m^*. \tag{89}$$

Integrating on both sides with respect to m gives

$$\bar{\lambda}=\mu+\bar{\lambda}\int_S\kappa(m^*)\,s_1(m^*)\,\mathrm{d}m^*. \tag{90}$$

The branching ratio is obtained by

$$\omega=1-\frac{\mu}{\bar{\lambda}}=\int_S\kappa(m^*)\,s_1(m^*)\,\mathrm{d}m^* \tag{91}$$

which is also the average number of events that are triggered by an arbitrary event.

For the ETAS model, the magnitude is completely separable from the whole intensity, and, thus, ω is identical to the criticality parameter ϱ. But this does not hold for general cases.

Example 7 : Here we assume that $h(m)$ is a probability density function and $H(m)$ is the corresponding c.d.f. Set

$$s(m\mid m')=\frac{h(m)}{H(am')},\quad 0<m<am',\quad 0<a\leq 1.,$$

and set $m_c=0$ to abbreviate the notation. That is, the magnitudes of offspring are limited to not be greater than am', where m' is the magnitude of the ancestor of interest. By Equation (82),

$$\rho V(m')=\frac{\kappa(m')}{H(a\,m')}\int_0^{a\,m'}h(m)V(m)\,\mathrm{d}m.$$

To obtain the eigenvalue, we apply the limit operation to both side of the above equality. When $M_i \to 0$, notice that $H(0) = 0$ and $h(M) = H'(M)$, then

$$\rho V(0) = \kappa(0) \lim_{m' \to 0} \frac{\int_0^{a\,m'} h(m)V(m)\,\mathrm{d}m}{H(am')} = \frac{a\kappa(0)V(0)}{a}.$$

Thus the criticality parameter ρ is

$$\rho = \kappa(0).$$

In this example, if $\kappa(m)$ is a monotonically increasing function of m, then, by (91), the branching ratio is

$$\omega = \int_{\mathcal{M}} \kappa(m^*)\, s_1(m^*)\,\mathrm{d}m^* > \int_{\mathcal{M}} \kappa(0)\, s_1(m^*)\,\mathrm{d}m^* = \kappa(0) = \varrho.$$

For the ETAS model, substituting $\kappa(m) = A\,\mathrm{e}^{\alpha(m-m_0)}$ and $s(m) = \beta\,\mathrm{e}^{-\beta(m-m_0)}$ into (86), the criticality parameter is

$$\varrho = \int_{m_0}^\infty s(m)\,\kappa(m)\,\mathrm{d}m = \frac{A\beta}{\beta - \alpha}.$$

When $\varrho < 1$, the ETAS model is stable and stationary, which requires $\beta \geq \alpha$ and $A \leq 1 - \alpha/\beta$. When $\varrho \geq 1$, there is a finite probability that the number of events in a unit time interval becomes infinite as t increases to infinity. More details on the behavior of the ETAS model were discussed by *Helmstetter and Sornette* (2002), *Zhuang and Ogata* (2006) and *Saichev and Sornette* (2007).

Applications There have been many applications of the ETAS model. For example: (i) detection of anomalous seismicity patterns (e.g. *Ogata* 2005) (ii) characterization of clustering characteristics by (regional) variations of the ETAS parameters (e.g. *Enescu et al.* 2009); (iii) detection of fluid-related forcing signals (e.g. *Hainzl and Ogata* 2005; *Lombardi et al.* 2010); (iv) regional probabilistic earthquake forecasting (e.g., *Helmstetter et al.* 2006), and so forth. Here we refer to *Zhuang et al.* (2011) for further discussions.

5 Transformed time sequence and residual analysis

Residual analysis is a powerful tool for assessing the fit of a particular model to a set of occurrence times (e.g., *Papangelou* 1972; *Ogata* 1988; *Daley and Vere-Jones*

2003). Assume that we have a realization of a point process with event times denoted by $t_1, t_2, \cdots, t_n$. We then calculate transformed event times, denoted by $\tau_1, \tau_2, \cdots, \tau_n$, in such a way that the transformed times have the same distributional properties as the a homogeneous Poisson process with unit rate parameter. Suppose that the conditional intensity is $\lambda(t)$. The transformed time sequence, for $i = 1, 2, \cdots, n$, is calculated as

$$\tau_i = \int_0^{t_i} \lambda(u) \, du. \tag{92}$$

Then the sequence $\{\tau_i : i = 1, 2, \cdots, n\}$ forms a Poisson process with unit rate.

To prove the above statement, we need only to prove that, from any transformed time $\tau = \int_0^t \lambda(u) \, du$, the waiting transformed time Y to next event is an exponential distribution with the unit rate. If the true model for the point process has a conditional $\lambda(t)$, by (5) and the equivalence between the conditional intensity and the hazard function, the survival function of the waiting time X to next event from t is

$$S_t(x) = \exp\left[-\int_t^{t+x} \lambda(u) \, du\right]. \tag{93}$$

To continue our proof, we require the following theorem: If X is a continuous random variable with a strictly increasing distribution function F, then F has an inverse F^{-1} defined on the open interval (0,1), and $F(X)$ and $1 - F(U)$ are both uniformly distributed on (0,1). Moreover, if U is a uniform r.v. on $(0, 1)$, then $F^{-1}(U)$ is a r.v. with the distribution function F.

Thus, $S_t(X) = \exp\left[-\int_t^{t+X} \lambda(u) \, du\right]$ is a uniform r.v. on $(0, 1)$. Since $Y = \int_0^t \lambda(u) \, du$, we have $S_t(X) = \exp[-Y]$, i.e., Y is exponentially distributed with rate 1.

This method can be used to test the goodness-of-fit of the model. If the fitted model $\hat{\lambda}(t)$ is close to the true model, then $\left\{\hat{\tau}_i = \int_0^{t_i} \hat{\lambda}(u) \, du : i = 1, 2, \cdots, \right\}$ is close to the standard Poisson model. The departure of the rate in the transformed time sequence from the unit rate indicates either increased activity or quiescence, relative to the seismic rate expected by the original model.

Example 8 (Continuation of Example 5.) A comparison between the cumulative numbers of earthquakes (black curve) and the corresponding fit values (red curve, Figure 5) shows that quiescence may have started on the 17th day after the mainshock. This is also confirmed in the plot of the transformed time domain (Figure 8).
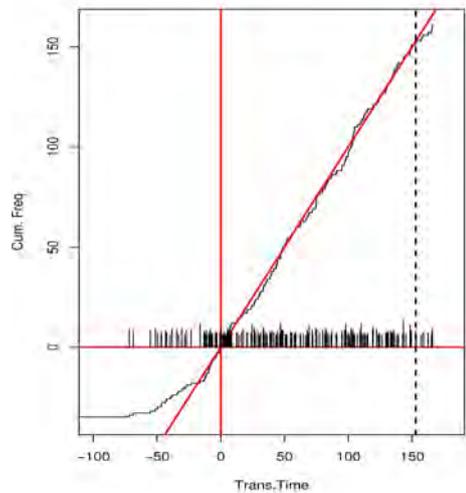
Fig. 8 Comparison of the observed and modeled cumulative number of $M \geq 4$-aftershocks of the Wenchuan earthquake: observation (black) and Omori-Utsu formula (red) in the transformed time domain.

6 Related software

There are a couple of versions of available software for fitting the models discussed in this paper. The first is IASPEI software SASeis implemented by *Utsu and Ogata* (1997) which is implemented using FORTRAN code, and later is revised as SASeis2006 by *Ogata* (2006) (URL: http://www.ism.ac.jp/∼ogata/Ssg/softwares.html). An alternative is provided by *Harte* (2010, 2012) in the R package "PtProcess" (URL: http://cran.r-project.org/web/packages/PtProcess/), which uses an object orientated approach. R is a statistical computing language developed by *R Development Core Team* (2012).

7 Summary

In this article, we have illustrated a developmental approach to modeling earthquake data using statistical models. By applying a sequence of progressively more sophisticated and possibly competing models, we can refine our understanding of the earthquake process. Starting from the model of complete randomness, the Poisson model, new models are constructed by adding deterministic factors of physical hypotheses or empirical observations relations. For example, the hypothesis of characteristic recurrence of earthquakes leads to the renewal models, the elastic-rebound theory leads to the stress release model, and empirical Omori-Utsu formula leads to the ETAS model. Many interesting studies, extensions and applications of the

above-mentioned models exist. We cannot go through all of them in the scope of our discussions. Instead, we provide the reader with a general overview of some of the most important point-process models of seismicity and the relevant methods for model fitting and inference.

Acknowledgements The authors thank the editor, Jeremy D. Zechar, and two anonymous reviewers for their constructive comments.

References

Akaike, H. (1974), A new look at the statistical model identification, *Automatic Control, IEEE Transactions on*, *19*(6), 716 – 723, doi:10.1109/TAC.1974.1100705. 8

Bebbington, M., and D. Harte (2001), On the statistics of the linked stress release process, *Journal of Applied Probability*, *38A*, 176–187. 17

Bebbington, M., and D. Harte (2003), The linked stress release model for spatio-temporal seismicity: formulations, procedures and applications, *Geophysical Journal International*, *154*, 925–946. 17

Cornell, C. A. (1968), Engineering seismic risk analysis, *Bullitin of the Seismogological Society of America*, *58*(5), 1583–1606. 8

Daley, D. D., and D. Vere-Jones (2003), *An Introduction to Theory of Point Processes – Volume 1: Elementary Theory and Methods (2nd Edition)*, Springer, New York, NY. 17, 33

Davis, P. M., D. D. Jackson, and Y. Y. Kagan (1989), The longer it has been since the last earthquake, the longer the expected time till the next?, *Bullitin of the Seismogological Society of America*, *79*(5), 1439–1456. 11

Ellsworth, W. L., M. V. Matthews, R. M. Nadeau, S. P. Nishenko, P. A. Reasenberg, and R. W. Simpson (1999), A physically-based earthquake recurrence model for estimation of long-term earthquake probabilities, *U.S. Geological Survey Open-File Report*, pp. 99–522. 13

Enescu, B., S. Hainzl, and Y. Ben-Zion (2009), Correlations of seismicity patterns in southern california with surface heat flow data, *Bullitin of the Seismogological Society of America*, *99*, 3114 – 3123, doi:10.1785/0120080038. 33

Field, E. H. (2007), A summary of previous Working Groups on California Earthquake Probabilities, *Bullitin of the Seismogological Society of America*, *97*(4), 1033–1053, doi:10.1785/0120060048. 11

Gardner, J. K., and L. Knopoff (1974), Is the sequence of earthquakes in southern California, with aftershocks removed, Poissonian?, *Bullitin of the Seismogological Society of America*, *64*(5), 1363–1367. 8

Gutenberg, B., and C. F. Richter (1944), Frequency of earthquakes in California, *Bull. Seis. Soc. Am.*, *34*, 184–188. 19

Hainzl, S., and Y. Ogata (2005), Detecting fluid signals in seismicity data through statistical earthquake modeling, *Journal of Geophysical Research*, *110*(B05), B05S07, doi:10.1029/2004JB003247. 33

Harte, D. (2010), PtProcess: An R package for modelling marked point processes indexed by time, *Journal of Statistical Software*, *35*(8), 1–32. 35

Harte, D. (2012), *PtProcess: Time Dependent Point Process Modelling*, R package version 3.3-1. 35

Hawkes, A. G. (1971a), Spectra of some self-exciting and mutually exciting point processes, *Biometrika*, *58*(1), 83–90, doi:10.1093/biomet/58.1.83. 27

Hawkes, A. G. (1971b), Point spectra of some mutually exciting point processes, *J. Royal Stat. Soc. Series B (Meth.)*, *33*(3), 438–443. 27

Hawkes, A. G., and D. Oakes (1974), A cluster process representation of a self-exciting process, *J. of Appl. Prob.*, *11*(3), 493–503. 27

Helmstetter, A., and D. Sornette (2002), Subcritical and supercritical regimes in epidemic models of earthquake aftershocks, *Journal of Geophysical Research*, *107*(B10), 2237. 33

Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2006), Comparison of short-term and time-independent earthquake forecast models for Southern California, *Bullitin of the Seismogological Society of America*, *96*(1), 90–106, doi: 10.1785/0120050067. 33

Kagan, Y. Y., and D. D. Jackson (1995), New seismic gap hypothesis: Five years after, *J. Geophys. Res.*, *100*(B3), 3943–3959. 11, 24

Kagan, Y. Y., and L. Knopoff (1987), Random stress and earthquake statistics - time-dependence, *Geophysical Journal of the Royal Astronomical Soiceity*, *88*(3), 723–731. 13

Kagan, Y. Y., and F. Schoenberg (2001), Estimation of the upper cutoff parameter for the tapered pareto distribution, *Journal of Applied Probability*, *38A*, 158–175. 19

Liu, J., D. Vere-Jones, L. Ma, Y. Shi, and J. Zhuang (1998), The principal of coupled stress release model and its application, *Acta Seismologica Sinica*, *11*, 273–281. 17

Lombardi, A. M., M. Cocco, and W. Marzocchi (2010), On the increase of background seismicity rate during the 1997–1998 Umbria-Marche, central Italy, sequence: apparent variation or fluid-driven triggering?, *Bullitin of the Seismogological Society of America*, *100*(3), 1138–1152, doi:10.1785/0120090077. 33

Lu, C., and D. Vere-Jones (2000), Application of linked stress release model to historical earthquake data: comparison between two kinds of tectonic seismicity, *Pure and Applied Geophysics*, *157*, 2351–2364, doi: 10.1007/PL00001087. 17

Lu, C., D. Harte, and M. Bebbington (1999), A linked stress release model for historical japanese earthquakes: Coupling among major seismic regions,, *Earth Planets Space*, *51*, 907–918. 17

Ma, L., and D. Vere-Jones (1997), Application of m8 and lin-lin algorithms to new zealand earthquake data, *New Zealand Journal of Geology and Geophysics*, *40*, 77–89. 11

MacFadden, J. A., and W. Weissblum (1965), Higher-oder properties of a stationary point process, *Journal of the Royal Statistical Society, Ser. B (Methodological)*, *25*, 413–431. 17

Matthews, M. V., W. L. Ellsworth, and P. A. Reasenberg (2002), A Brownian model for recurrent earthquakes, *Bulletin of the Seismological Society of America*, *92*(6), 2233–2250, doi:10.1785/0120010267. 13

Nishenko, S. P. (1991), Circum-pacific seismic potential 1989-1999, *Pure Appl. Geophys.*, *135*, 169–259. 11

Nishenko, S. P., and R. Buland (1987), A generic recurrence interval distribution for earthquake forecasting, *Bulletin of the Seismological Society of America*, *77*(4), 1382–1399. 11

Nomura, S., Y. Ogata, F. Komaki, and S. Toda (2011), Bayesian forecasting of recurrent earthquakes and predictive performance for a small sample size, *Journal of Geophysical Research*, *116*, B04,315, doi:10.1029/2010JB007917. 17

Ogata, Y. (1988), Statistical models for earthquake occurrences and residual analysis for point processes, *Journal of the American Statistical Association*, *83*, 9 – 27. 26, 27, 33

Ogata, Y. (2002), Slip-size-dependent renewal processes and bayesian inferences for uncertainties, *Journal of Geophysical Research*, *107*(B11), 2268, doi:10.1029/2001JB000668. 16, 17

Ogata, Y. (2005), Detection of anomalous seismicity as a stress change sensor, *Journal of Geophysical Research*, *110*, B05S06, doi:10.1029/2004JB003245. 33

Ogata, Y. (2006), *Statistical Analysis of Seismicity - Updated Version (SASeis2006)*, Computer Science Monograph, *33*, 1-28. The Institute of Statistical Mathematics, Tokyo, Japan. 27, 35

Ogata, Y., and K. Katsura (2004), Point-process models with linearly parameterized intensity for application to earthquake data, in *Essays in Time Series and Allied Processes (Papers in honour of E.J. Hannan)*, vol. 23A, edited by J. Gani and M. B. Priestley, pp. 291–310, Journal of Applied Probability, New York. 11

Ogata, Y., and K. Shimazaki (1984), Transition from aftershock to normal activity, *Bulletin of the Seismological Society of America*, *74*(5), 1757–1765. 25, 26

Omori, F. (1894), On the aftershocks of earthquakes, *Journal of the College of Science, Imperial University of Tokyo*, *7*, 111–200. 22

Papangelou, F. (1972), Integrability of expected increments of point processes and a related random change of scale, *Trans. Amer. Math. Soc.*, *164*, 438–506. 33

R Development Core Team (2012), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0. 35

Reasenberg, P. A., and L. M. Jones (1989), Earthquake hazard after a mainshock in California, *Science*, *243*, 1173 – 1176. 23

Reasenberg, P. A., and L. M. Jones (1994), Earthquake aftershocks: Update, *Science*, *265*, 1251 – 1252. 23

Reid, H. (1910), The Mechanics of the Earthquake, The California Earthquake of April 18, 1906, Report of the State Investigation Commission, Vol. 2, *Carnegie Institution of Washington, Washington, DC*, pp. 16–28. 11, 17

Ross, S. M. (2003), *Introduction to Probability Models, Eighth Edition*, 773 pp., Academic Press. 8

Saichev, A., and D. Sornette (2007), Theory of earthquake recurrence times, *J. Geophys. Res.*, *112*(B04313), doi: 10.1029/2006JB004536. 33

Scholz, C. H. (2002), *The mechanics of earthquakes and faulting (2nd ed.)*, Cambridge University Press. 18

Shi, Y., J. Liu, D. Vere-Jones, J. Zhuang, , and L. Ma (1998), Application of mechanical and statistical models to study of seismicity of synthetic earthquakes and the prediction of natural ones, *Acta Seismol. Sinica*, *11*, 421–430. 17

Sykes, L. R., and W. Menke (2006), Repeat times of large earthquakes: Implications for earthquake mechanics and long-term prediction, *Bull. Seismol. Soc. Am.*, *96*(5), 1569–1596, doi:10.1785/0120050083. 11

Utsu, T. (1957), Magnitude of earthquakes and occurrence of their aftershocks, *Zisin (J. Seismol. Soc. Jap.)*, *10*, 35–45 (in Japanese). 23

Utsu, T. (1970), Aftershock and earthquake statistics (ii) – further investigation of aftershocks and other earthquake sequences based on a new classification of earthquake sequences, *Journal of the Faculty the Science, Hokkaido Univesity, Ser. VII (Geophysics)*, *3*, 197 – 266. 25

Utsu, T. (1999), *Seismicity Studies: A Comprehensive Review*, 876 pp., Univ. of Tokyo Press. 16

Utsu, T., and Y. Ogata (1997), Statistical analysis of seismicity., in *Algorithms for Earthquake Statistics and Prediction. International Association of Seismology and Physics of the Earth's Interior (IASPEI) Library Volume 6.*, edited by J. Healy, V. Keilis-Borok, and W. Lee, pp. 13–94, IASPEI, Menlo Park CA. 35

Utsu, T., Y. Ogata, and R. S. Matsu'ura (1995), The centenary of the Omori formula for a decay law of aftershock activity, *Journal of Physics of the Earth*, *43*, 1–33. 23

Vere-Jones, D., R. Robinson, and W. Yang (2001), Remarks on the accelerated moment release model: problems of model formulation, simulation and estimation, *Geophysical Journal International*, *144*(3), 517–531. 19

Weibull, W. (1951), A statistical distribution function of wide applicability, *Journal of Applied Mechanics-Transactions of the ASME*, *18*(3), 293–297. 13

Zheng, X., and D. Vere-Jones (1991), Application of stress release models to historical earthquakes from North China, *Pure and Applied Geophysics*, *135*(4), 559–576, doi:10.1007/BF01772406. 17

Zheng, X., and D. Vere-Jones (1994), Further applications of the stochastic stress release model to historical earthquake data, *Tectonophysics*, *229*, 101–121. 11, 17, 19, 20, 22

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. 33

Zhuang, J., M. 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. 33

Appendix: List of aftershocks of the Wenchuan earthquake

Table 5: DATA FILE: *example4.data*

#	year	month	day	hour	minute	lat.	long.	mag.	days since mainshock
1	2008	5	12	14	28	31	103.4	8	0
2	2008	5	12	14	43	31	103.5	6	0.0104
3	2008	5	12	15	34	31	103.5	5	0.0458
4	2008	5	12	15	40	31	103.6	4.7	0.05
5	2008	5	12	16	10	31.2	103.4	4.8	0.0708
6	2008	5	12	16	21	31.3	104.1	5.2	0.0785
7	2008	5	12	16	26	31.5	103.8	4.3	0.0819
8	2008	5	12	16	35	31.4	103.5	4.6	0.0882
9	2008	5	12	16	36	31	103.2	4.3	0.0889
10	2008	5	12	16	47	32.2	105.4	4.8	0.0965
11	2008	5	12	16	50	32.6	105.2	4.5	0.0986
12	2008	5	12	17	1	32.2	104.7	4.1	0.1062
13	2008	5	12	17	7	31.3	103.8	5	0.1104
14	2008	5	12	17	23	32.3	104.8	5.1	0.1215
15	2008	5	12	17	30	32.4	104.9	4.1	0.1264
16	2008	5	12	17	42	31.4	104	5.2	0.1347
17	2008	5	12	17	54	31	103.2	4.3	0.1431
									Continued on next page ...

(Continuation from previous page ...)

#	year	month	day	hour	minute	lat.	long.	mag.	days since mainshock
18	2008	5	12	18	2	32.2	105.1	4.8	0.1486
19	2008	5	12	18	23	31	103.3	4.9	0.1632
20	2008	5	12	19	10	31.4	103.6	6	0.1958
21	2008	5	12	19	33	32.6	105.4	4.5	0.2118
22	2008	5	12	19	41	32.4	105.1	4.6	0.2174
23	2008	5	12	19	45	32.4	105	4	0.2201
24	2008	5	12	19	52	32.6	105.4	4.7	0.225
25	2008	5	12	20	4	32.6	105.2	4.2	0.2333
26	2008	5	12	20	6	32.2	105.5	4.1	0.2347
27	2008	5	12	20	11	31.4	103.8	4.5	0.2382
28	2008	5	12	20	15	32	104.4	4.9	0.241
29	2008	5	12	20	23	32.7	105.3	4.5	0.2465
30	2008	5	12	20	29	31.4	103.9	4.1	0.2507
31	2008	5	12	20	33	31.4	104.1	4.2	0.2535
32	2008	5	12	20	54	31.3	103.4	4.3	0.2681
33	2008	5	12	21	2	31.1	103.5	4.6	0.2736
34	2008	5	12	21	7	31	103.4	4.3	0.2771
35	2008	5	12	21	32	31.2	103.9	4.3	0.2944
36	2008	5	12	21	36	32.9	105.5	4	0.2972
37	2008	5	12	21	40	31	103.5	5.1	0.3
38	2008	5	12	21	55	32	104.3	4.2	0.3104
39	2008	5	12	22	6	32.5	105.1	4	0.3181
40	2008	5	12	22	9	31.9	104.7	4.5	0.3201
41	2008	5	12	22	15	32.2	104.9	4.4	0.3243
42	2008	5	12	22	26	31.3	103.9	4	0.3319
43	2008	5	12	22	37	32.2	104.5	4.3	0.3396
44	2008	5	12	22	46	32.7	105.5	5.1	0.3458
45	2008	5	12	22	55	32.4	105	4.2	0.3521
46	2008	5	12	23	5	31.3	103.6	5	0.359
47	2008	5	12	23	5	31.3	103.5	5.2	0.359
48	2008	5	12	23	16	30.9	103.2	4.1	0.3667
49	2008	5	12	23	28	31	103.5	5	0.375
50	2008	5	13	0	28	31.2	103.8	4.4	0.4167
51	2008	5	13	0	34	32.5	105	4.2	0.4208
52	2008	5	13	1	1	30.9	103.4	4	0.4396
53	2008	5	13	1	29	31.3	103.4	4.6	0.459
54	2008	5	13	1	54	31.3	103.4	5	0.4764
55	2008	5	13	2	46	32.4	105	4.4	0.5125
56	2008	5	13	2	55	31.9	105.1	4.4	0.5188
57	2008	5	13	3	53	31.3	103.6	4.3	0.559
58	2008	5	13	4	8	31.4	104	5.7	0.5694
59	2008	5	13	4	45	31.7	104.5	5.2	0.5951
60	2008	5	13	4	51	32.4	105.2	4.7	0.5993
61	2008	5	13	5	8	31.3	103.2	4.5	0.6111
62	2008	5	13	5	51	32.5	105.3	4.6	0.641
63	2008	5	13	6	19	31.9	104.2	4.1	0.6604
64	2008	5	13	6	24	32.2	105	4	0.6639
65	2008	5	13	6	47	31.3	103.4	4.5	0.6799
66	2008	5	13	7	38	31.9	104.5	4	0.7153
67	2008	5	13	7	46	31.2	103.4	5.3	0.7208
68	2008	5	13	7	54	31.3	103.6	5.1	0.7264
69	2008	5	13	8	22	31.3	104	4.2	0.7458
70	2008	5	13	8	54	32.6	105.2	4.3	0.7681

Continued on next page ...

(Continuation from previous page ...)

#	year	month	day	hour	minute	lat.	long.	mag.	days since mainshock
71	2008	5	13	9	7	31.4	103.7	4	0.7771
72	2008	5	13	10	15	31.6	103.9	4.5	0.8243
73	2008	5	13	10	33	31.3	103.6	4.2	0.8368
74	2008	5	13	10	59	31	103.3	4.3	0.8549
75	2008	5	13	11	0	31.2	103.5	4.7	0.8556
76	2008	5	13	11	48	31.2	103.7	4.5	0.8889
77	2008	5	13	12	45	31	103.3	4.1	0.9285
78	2008	5	13	12	50	31.3	103.4	4.1	0.9319
79	2008	5	13	13	25	32.6	105.2	4.2	0.9562
80	2008	5	13	13	36	32.4	105.2	4.4	0.9639
81	2008	5	13	13	37	31	103.5	4.6	0.9646
82	2008	5	13	14	38	31.4	103.8	4.3	1.0069
83	2008	5	13	15	7	30.9	103.4	6.1	1.0271
84	2008	5	13	15	19	32.3	105	4.8	1.0354
85	2008	5	13	15	51	32.5	105.3	4.4	1.0576
86	2008	5	13	15	53	32.3	105	4.8	1.059
87	2008	5	13	16	11	32.5	105.2	4.1	1.0715
88	2008	5	13	16	20	31.4	103.9	4.8	1.0778
89	2008	5	13	17	41	32.1	104.4	4.1	1.134
90	2008	5	13	18	16	31.8	104.3	4	1.1583
91	2008	5	13	18	36	31.3	103.6	4.1	1.1722
92	2008	5	13	20	51	32.3	105	4.6	1.266
93	2008	5	13	21	13	32.5	105.5	4.4	1.2812
94	2008	5	13	21	31	32.4	105.1	4.5	1.2937
95	2008	5	13	23	10	32.6	105.5	4.1	1.3625
96	2008	5	13	23	54	32.1	104.9	4.5	1.3931
97	2008	5	14	0	23	31.7	104.3	4	1.4132
98	2008	5	14	3	30	31.1	103.3	4.1	1.5431
99	2008	5	14	3	51	31	103.3	4.2	1.5576
100	2008	5	14	6	3	31.3	103.6	4.2	1.6493
101	2008	5	14	8	8	31.1	103.4	4	1.7361
102	2008	5	14	9	9	31.4	103.8	4.8	1.7785
103	2008	5	14	9	56	31.1	103.5	4.1	1.8111
104	2008	5	14	10	54	31.3	103.4	5.6	1.8514
105	2008	5	14	13	54	31.9	104.2	4.9	1.9764
106	2008	5	14	14	33	31.4	103.9	4.1	2.0035
107	2008	5	14	17	26	31.4	104	5.1	2.1236
108	2008	5	14	17	51	32.4	104.2	4.8	2.141
109	2008	5	14	17	57	32.3	104.8	4.2	2.1451
110	2008	5	14	18	0	32.2	104.7	4.7	2.1472
111	2008	5	14	18	11	32.2	104.5	4.2	2.1549
112	2008	5	14	18	18	32.4	105.1	4.8	2.1597
113	2008	5	14	18	30	32.4	105.2	4.9	2.1681
114	2008	5	14	21	29	32.3	105.1	4.5	2.2924
115	2008	5	15	1	17	31.5	103.8	4.8	2.4507
116	2008	5	15	1	33	31.4	103.5	4.6	2.4618
117	2008	5	15	3	59	31.1	103.5	4.2	2.5632
118	2008	5	15	5	1	31.6	104.2	5	2.6063
119	2008	5	15	6	10	31.2	103.6	4.6	2.6542
120	2008	5	15	8	9	31.8	104.4	4.3	2.7368
121	2008	5	15	8	50	31.3	103.4	4	2.7653
122	2008	5	15	12	27	31.3	103.7	4.1	2.916
123	2008	5	15	13	27	32	104.3	4.6	2.9576

(Continuation from previous page ...)

#	year	month	day	hour	minute	lat.	long.	mag.	days since mainshock
124	2008	5	15	20	10	31.4	103.8	4.2	3.2375
125	2008	5	15	21	4	32.6	105.6	4.4	3.275
126	2008	5	16	5	55	32.3	104.7	4.5	3.6437
127	2008	5	16	6	10	31.4	103.9	4.6	3.6542
128	2008	5	16	6	34	31.9	104.4	4.2	3.6708
129	2008	5	16	11	10	32.5	105.1	4	3.8625
130	2008	5	16	11	34	31.4	104.1	4.9	3.8792
131	2008	5	16	13	25	31.4	103.2	5.9	3.9562
132	2008	5	16	14	34	32.4	105.2	4.3	4.0042
133	2008	5	16	18	17	31.3	103.5	4.3	4.159
134	2008	5	16	18	20	32.5	105.1	4	4.1611
135	2008	5	16	18	51	31.4	103.6	4.2	4.1826
136	2008	5	16	21	21	31.9	104.2	4	4.2868
137	2008	5	17	0	14	31.2	103.5	5.1	4.4069
138	2008	5	17	1	22	31.2	103.6	4.5	4.4542
139	2008	5	17	3	59	31	103.5	4.3	4.5632
140	2008	5	17	4	0	32.6	105.4	4.1	4.5639
141	2008	5	17	4	15	32.2	104.4	4.8	4.5743
142	2008	5	17	4	16	31.3	103.5	5	4.575
143	2008	5	17	4	29	31.4	103.3	4.5	4.584
144	2008	5	17	6	33	32.2	105.1	4.2	4.6701
145	2008	5	17	7	23	31.3	103.8	4.2	4.7049
146	2008	5	17	8	28	31.6	104	4.1	4.75
147	2008	5	17	8	38	32	104	4	4.7569
148	2008	5	17	15	38	32	104.4	4.1	5.0486
149	2008	5	17	21	32	32.2	104.7	4.7	5.2944
150	2008	5	18	1	8	32.1	105	6	5.4444
151	2008	5	18	4	26	31.2	103.5	4.1	5.5819
152	2008	5	18	8	45	31.8	104	4.2	5.7618
153	2008	5	18	9	4	31.1	103.5	4.2	5.775
154	2008	5	18	11	51	31	103.4	4.2	5.891
155	2008	5	18	17	25	31.2	103.1	4	6.1229
156	2008	5	18	20	37	31.3	103.2	4.2	6.2562
157	2008	5	19	12	8	32.1	105	4.8	6.9028
158	2008	5	19	14	6	32.5	105.3	5.4	6.9847
159	2008	5	20	1	52	32.3	104.9	5	7.475
160	2008	5	20	8	57	31.7	104	4.1	7.7701
161	2008	5	20	11	42	32.6	105.4	4	7.8847
162	2008	5	20	12	17	30.8	103.3	4.3	7.909
163	2008	5	20	14	54	31.8	104.2	4	8.0181
164	2008	5	21	0	38	30.9	103.3	4	8.4236
165	2008	5	21	16	40	31.4	103.3	4.1	9.0917
166	2008	5	21	17	33	32.3	105.2	4.5	9.1285
167	2008	5	21	21	59	31.4	103.9	4.5	9.3132
168	2008	5	21	23	29	32.4	105.1	4.3	9.3757
169	2008	5	22	4	36	32.2	104.8	4.6	9.5889
170	2008	5	22	15	18	31.2	103.6	4.7	10.0347
171	2008	5	22	19	22	32.6	105.4	4	10.2042
172	2008	5	22	23	0	31.9	104.3	4.3	10.3556
173	2008	5	23	1	37	31.3	103.6	4.6	10.4646
174	2008	5	23	8	5	31.2	103.6	4.7	10.734
175	2008	5	23	9	23	31.2	103.5	4.2	10.7882
176	2008	5	23	11	12	31.2	103.2	4.2	10.8639

(Continuation from previous page ...)

#	year	month	day	hour	minute	lat.	long.	mag.	days since mainshock
177	2008	5	24	0	10	32.2	105	4	11.4042
178	2008	5	24	1	53	32.5	105.2	4.1	11.4757
179	2008	5	24	11	0	31.1	103.4	4	11.8556
180	2008	5	25	12	27	32	104.6	4.2	12.916
181	2008	5	25	16	21	32.6	105.4	6.4	13.0785
182	2008	5	25	17	34	33	104.9	4.7	13.1292
183	2008	5	26	8	39	30.8	103.3	4.3	13.7576
184	2008	5	27	16	3	32.7	105.6	5.4	15.066
185	2008	5	27	16	37	32.8	105.6	5.7	15.0896
186	2008	5	27	21	59	32.5	105.2	4.7	15.3132
187	2008	5	28	0	46	32.2	104.6	4.2	15.4292
188	2008	5	28	1	35	32.7	105.4	4.7	15.4632
189	2008	5	29	12	48	32.6	105.5	4.6	16.9306
190	2008	5	29	15	10	31.4	103.7	4.5	17.0292
191	2008	5	31	14	22	32.4	105	4	18.9958
192	2008	5	31	15	34	32.6	105.4	4	19.0458
193	2008	6	1	11	23	31.6	104	4.5	19.8715
194	2008	6	3	11	9	32	104.5	4.3	21.8618
195	2008	6	5	1	26	32.3	105.1	4	23.4569
196	2008	6	5	5	21	31.2	103.4	4.2	23.6201
197	2008	6	5	12	41	32.3	105	5	23.9257
198	2008	6	5	14	2	32.7	105.5	4.3	23.9819