Title

Estimation of Space-time Branching Process Models in Seismology using an EM-type Algorithm

Authors

Alejandro Veen
Frederic Paik Schoenberg

Publication Date

2011-10-25

Estimation of space-time branching process models in seismology using an EM-type algorithm

Alejandro Veen
IBM T.J. Watson Research Center
1101 Kitchawan Road, Route 134
Yorktown Heights, NY 10598
aveen@us.ibm.com

Frederic Paik Schoenberg
UCLA Department of Statistics
8125 Math Sciences Building, Box 951554
Los Angeles, CA 90095-1554
frederic@stat.ucla.edu

June 26, 2006

Abstract

The estimation of branching point process models by maximum likelihood can be unstable and computationally intensive. We explore an alternative estimation method based on the Expectation-

Maximization algorithm. The method involves viewing the estimation of such branching processes as analogous to incomplete data problems. Using an application from seismology, we show how the Epidemic-type Aftershock Sequence (ETAS) model can in fact be estimated this way and we propose a computationally efficient procedure to maximize the log-likelihood function. Using a space-time ETAS model, we demonstrate that this method is extremely robust and accurate and use it to estimate declustered background seismicity rates of geologically distinct regions in Southern California. All regions show similar declustered background intensity estimates except for the one covering the Southern section of the San Andreas fault system to the East of San Diego in which a substantially higher intensity is observed.

Keywords: earthquakes, epidemic-type aftershock sequence model, ETAS model, space-time point process models, branching process models

1 Introduction

Point process models have long been used to describe earthquake occurrences (Vere-Jones 1970, 1975). Ogata (1999) provides a nice review. Some of the early applications fitted Neyman-Scott-type models in which main shocks are viewed as cluster centers, each of which may trigger a random number of aftershocks with magnitudes not larger than the main shock (Vere-Jones 1970, Hawkes and Adamopoulos 1973). More recent work has favored the use of sub-critical branching process models with immigration in which all

earthquakes can trigger aftershocks, and among these, the Epidemic-type Aftershock Sequence (ETAS) model is considered to be among the standard models in seismology (Ogata 1988, 1998).

ETAS and other branching process models are commonly estimated using Maximum Likelihood (ML). However, closed form solutions are usually not available and numerical maximization algorithms must be employed. In such situations, computational difficulties can arise especially if the models are complex, multidimensional, and non-linear, as this often leads to multimodal or extremely flat log-likelihood functions.

The view of branching process models as incomplete data problems suggests the use of the Expectation-Maximization (EM) algorithm established by Dempster, Laird, and Rubin (1977). In this context, the information about which event 'triggers' each other event is unobservable and can be described probabilistically. The EM algorithm involves maximizing the *expected* log-likelihood, which in the context of branching point process models is based on the probabilistic incorporation of the branching structure and is usually easier to maximize.

In this work, we show how an EM-type algorithm may be used to maximize the log-likelihood using a partial information approach in certain steps. This relates to partial likelihood maximization which was introduced by Cox (1975) and which is briefly discussed by Ogata and Akaike (1982) in the context of branching processes in seismology. By coupling two well-established estimation methods (EM and a partial information approach in certain steps)

we are able to present a highly robust and accurate estimation procedure which can be used to estimate even very complex branching process models. In order to demonstrate the properties of our proposed method, we use a space-time ETAS model to simulate earthquake catalogs and then compare the results of the EM-type estimation algorithm to the traditional ML procedure.

Following a description in Section 2 of self-exciting point process models for earthquake occurrences, we describe some of the problems with conventional maximum likelihood estimation of such models in Section 3. Section 4 describes a proposed alternative estimation method, based on the EM algorithm, and shows its robustness and accuracy using simulations. The EM-type algorithm is then used in Section 5 to estimate background seismicity rates for Southern California.

2 Self-exciting point processes and the ETAS model

Consider a simple, temporal point process N on $[0, \infty)$ adapted to a filtration $\mathcal{H}_t$. Assuming it exists, the conditional intensity $\lambda(t|\mathcal{H}_t)$ is defined as the unique, non-decreasing, $\mathcal{H}$-predictable process such that $N([0,t)) - \int \lambda(t|\mathcal{H}_t)\mathrm{d}t$ is an $\boldsymbol{H}$-martingale. In this representation, $\boldsymbol{H}$ must contain the history of the process up to time t, denoted as $\mathcal{H}_t = \{t_i : t_i < t\}$, but may contain additional information as well. Since the finite-dimensional distri-

butions of such a point process are uniquely determined by its conditional intensity (Daley and Vere-Jones 2003), one way to model a point process is via its conditional intensity.

In self-exciting point processes, the conditional intensity is given by

$$\lambda(t|\mathcal{H}_t) = \mu + \sum_{i:t_i<t} g(t - t_i),$$

where $\mu > 0$, $g(v) \geq 0$ for nonnegative v and equals zero otherwise, and $\int_0^\infty g(v)dv < 1$ in order to ensure stationarity (Hawkes 1971b, Hawkes 1971a). Early applications of self-exciting point processes to earthquake occurrence models can be found in Hawkes and Adamopoulos (1973) as well as in Lomnitz (1974, Chapter 7). Modeling earthquake occurrences using a self-exciting point process implies the separation of the seismicity into a long-term background component μ and a short-term clustering component $\sum_{i:t_i<t} g(t - t_i)$ which represents aftershock activity. An interesting historical note is that Hawkes and Adamopoulos (1973), who were the first to explore the use of a self-exciting point processes to model seismicity, concluded in their work that these types of models do not fit the data very well and claimed that Neyman-Scott models provide superior fit.

A particularly important example of a self-exciting point process is the ETAS model, which was first introduced by Ogata (1988) and is widely used to describe earthquake occurrences. Early forms of the ETAS model (Ogata

1988) only took magnitudes and earthquake occurrence times into account:

$$\lambda(t|\mathcal{H}_t) = \mu + \sum_{i:t_i<t} g(t - t_i, m_i), \tag{1}$$

where the history of the process also includes earthquake magnitudes, $\mathcal{H}_t = \{(t_i, m_i) : t_i < t\}$, μ is the (in this case constant) background intensity of earthquake occurrences, and $g(\cdot)$ is the so-called "triggering function", since it describes how earthquakes trigger aftershocks. One possible triggering function suggested in (Ogata 1988) is

$$g(\tau_i, m_i) = \frac{K_0}{(\tau_i + c)^{(1+\omega)}} \, \mathrm{e}^{a(m_i - M_0)}, \tag{2}$$

where $\tau_i = t - t_i$ is the time elapsed since earthquake i, m_i is the magnitude of earthquake i, $K_0 > 0$ is a normalizing constant governing the expected number of direct aftershocks triggered by earthquake i, the parameters $c, a, \omega > 0$, and M_0 is the "cut-off magnitude", i.e. the lowest earthquake magnitude in the data set under consideration. The term $K_0/(\tau_i + c)^{(1+\omega)}$ describing the temporal distribution of aftershocks is known in seismology as the modified Omori-Utsu law. While the literature in seismology usually lets $\omega > -1$, the interpretation of the modified Omori-Utsu law as a probability density function requires strictly positive values for ω.

The ETAS model has since been extended to describe the space-time-magnitude distribution of earthquake occurrences (Ogata 1993, 1998). A

version suggested in Ogata (1998) uses circular aftershock regions where the squared distance between an aftershock and its triggering event follows a Pareto distribution:

$$\lambda(t, x, y|\mathcal{H}_t) = \mu(x, y) + \sum_{i:t_i<t} g(t - t_i, x - x_i, y - y_i, m_i), \tag{3}$$

with triggering function

$$g(t - t_i, x - x_i, y - y_i, m_i) = \frac{K_0 \cdot \mathrm{e}^{a(m_i - M_0)}}{(t-t_i+c)^{(1+\omega)}\,((x-x_i)^2+(y-y_i)^2+d)^{(1+\rho)}}, \tag{4}$$

where (x_i, y_i) represents the epicenter of earthquake i, $d > 0$ and $\rho > 0$ are parameters describing the spatial distribution of triggered seismicity, and where the history of the process up to time t is now defined as $\mathcal{H}_t = \{(t_i, x_i, y_i, m_i) : t_i < t\}$.

Summing things up, the ETAS model can be described as a subcritical branching process with immigration. The process is subcritical because the expected number of aftershocks is less than unity (thus ensuring stationarity) and it is characterized by immigration through spontaneous background earthquakes. The aftershock activity is modeled through a triggering function consisting of two terms, one of which models the expected number of aftershocks for earthquake i while the other part models the temporal or space-time distribution of the triggered aftershocks.

3 Conventional maximum likelihood estimation of the ETAS model

The log-likelihood of model (3) is given by

$$\ell(\boldsymbol{\theta}) = \sum_i \lambda(t_i, x_i, y_i|\mathcal{H}_{t_i}) - \int_0^T \int_{y_0}^{y_1} \int_{x_0}^{x_1} \lambda(t, x, y|\mathcal{H}_t)\, \mathrm{d}x\mathrm{d}y\mathrm{d}t, \qquad (5)$$

where $\boldsymbol{\theta} = (\mu, K_0, a, c, \omega, d, \rho)$ is the parameter vector and $[x_0, x_1] \times [y_0, y_1] \times [0, T]$ is the space-time window in which the data set (x_i, y_i, t_i, m_i) is observed (Ogata 1998, Daley and Vere-Jones 2003, ch. 7). Let $\hat{\boldsymbol{\theta}}$ denote the value that maximizes (5), i.e. the Maximum Likelihood Estimate (MLE). Typically, $\hat{\boldsymbol{\theta}}$ is obtained by using a numerical optimization routine, since no closed form solution is generally available. Unfortunately, in cases where the log-likelihood function is extremely flat in the vicinity of its maximum, such optimization routines can have convergence problems and can be substantially influenced by arbitrary choices of starting values. In order to distinguish the MLE computed by numerical maximization from the one based on the EM-type algorithm presented later, we will denote the former $\hat{\boldsymbol{\theta}}_{num}$ and the latter $\hat{\boldsymbol{\theta}}_{EM}$. Similarly, components of the parameter vector estimates will be denoted $\hat{\mu}_{num}$, $\hat{\mu}_{EM}$, $\hat{K_0}_{num}$, $\hat{K_0}_{EM}$, etc.

An illustration may be helpful in order to demonstrate some of the difficulties encountered when directly maximizing the log-likelihood function (5) using numerical methods. Figure 1 shows a simulated earthquake catalog of

638 events, using the model (3). The space-time window used for this simulation is similar to the Southern California data set described in Section 5. Background earthquakes are simulated on an area of 8° of longitude by 5° of latitude over a period of 7500 days (approximately 20 years). Parameter values, as shown in Table 1, are chosen to approximate those used in descriptions of earthquake catalogs, based on Ogata (1998) as well as discussion with UCLA seismologists Yan Y. Kagan and Ilya Zaliapin. For simplicity, a truncated exponential distribution is used to model earthquake magnitudes in accordance with the Gutenberg-Richter law (Gutenberg and Richter 1944):

$$f_{GR}(m) = \frac{\beta \, \mathrm{e}^{-\beta(m-M_0)}}{1 - \mathrm{e}^{-\beta\left(M_{GR}^{max} - M_0\right)}}, \tag{6}$$

where $f_{GR}(m)$ is the probability density function, $\beta = \log(10)$, and $M_0 = 2 \leq m \leq M_{GR}^{max} = 8$, where the lower threshold is the approximate current threshold (since 2001) above which catalogs of the Southern California Seismic Network (SCSN) are believed to be complete (Kagan 2002, 2003), and the upper threshold is the approximate magnitude of the strongest Californian earthquakes in historic times, the 1857 Fort Tejon earthquake and the 'great' San Francisco earthquake of 1906.

The use of numerical methods to maximize the log-likelihood function (5) can be problematic in cases where the log-likelihood is extremely flat, unless some supervision is imposed and intelligent starting values are used. Figure 2, for instance, shows the log-likelihood for variations of each component of the

parameter vector by up to 50% around the MLE $\hat{\boldsymbol{\theta}}$. The log-likelihood is quite flat around $\hat{\boldsymbol{\theta}}$, especially with regard to the parameters μ, K_0, c, and d; as a result these parameters are difficult to estimate and they generally are associated with rather large standard errors as well as numerical challenges during the estimation procedure. The parameters a, ω, and ρ, on the other hand, show much more peaked log-likelihood functions and can hence be estimated more stably.

The issue of log-likelihood flatness can be aggravated in a multidimensional context. In Figure 3, two parameters are varied while the other components of $\hat{\boldsymbol{\theta}}$ remain constant at their MLEs. Again, the log-likelihood function can stay extremely flat along certain trajectories, even for large deviations from the MLEs. The parameter c, for instance, can be increased to more than four times its MLE and ω increased to double its MLE, yet the log-likelihood function is reduced only very slightly. The problem of log-likelihood flatness becomes increasingly severe as more and more parameters are varied at once. In more realistic settings, where ETAS models are estimated for actual earthquake catalogs, none of the parameters would be known in advance.

In cases where the log-likelihood function is extremely flat, the choice of starting values can influence the results. In Figure 4, numerical ML estimation is performed for eight different starting values for $\boldsymbol{\theta}$ using a standard Newton-Raphson optimization routine. The 'true' values of Table 1 are chosen for all parameters except K_0 and a, which are varied as indicated in the figure. In two cases, the estimation results are quite close to $\boldsymbol{\theta}$, while in four

other cases, the algorithm finds reasonable estimates for most parameters but fails to find an acceptable estimate for K_0: in fact, the algorithm does not change the starting value for K_0 at all, even though it is roughly 33% off its 'true' value. In two of the cases, the algorithm fails to converge.

As shown in Figure 4, even if the starting values are close to the MLE, any departure from an optimal choice of tolerance levels and stopping criteria for the numerical maximization procedure can lead to poor convergence results. In this example, the Newton-type algorithm used to find the MLEs actually yields better results if the starting values for K_0 are not too close to the 'true' values, as the computation of gradients may actually be more accurate in locations further away from the true parameter value.

Another problem often encountered in practice is that the log-likelihood function can be multimodal (see e.g. Ogata and Akaike (1982)). Whenever the numerical optimization routine converges to a solution, it is quite difficult to determine whether it has converged to a local maximum or to the global maximum. Even if the log-likelihood is unimodal, in cases where the log-likelihood is flat there can be numerical multimodality due to rounding errors, the way these errors affect intermediate and final results, and the way values are stored in memory. In cases where the log-likelihood surface is extremely flat such as those shown in Figures 3 and 4, such numerical problems can explain why the MLE can be very far from the true parameter value. This makes it difficult to use ML in simulation studies of bias and asymptotic properties for which an automatic procedure would be desirable.

4 ETAS estimation using an EM-type algorithm

In their seminal paper, Dempster, Laird, and Rubin (1977) established the EM algorithm as the estimation method of choice for incomplete data problems. It has been extended in various ways and adapted to a wide range of applications. A good overview of this algorithm and its extensions is provided by McLachlan and Krishnan (1996).

The estimation of the ETAS model can be viewed as incomplete data problem in which the unobservable quantity u_i identifies whether an earthquake is a background event ($u_i = 0$) or whether it was triggered by a preceding event, denoted as $u_i = j$ for the case that earthquake i was triggered by earthquake j. This view of the ETAS model is inspired by what Zhuang, Ogata, and Vere-Jones (2002, 2004) call "stochastic reconstruction". In their work, probabilities of earthquakes being background events or otherwise triggered by preceding events are used to improve spatial background intensity estimates. In our work, the branching structure of the model will help estimate all parameters of the ETAS model, not just the background intensity.

In the following, consider an ETAS model with an inhomogeneous background rate $\mu(x, y)$. While the EM methodology allows for quite general forms of $\mu(x, y)$, here we model an inhomogeneous background rate by subdividing the (in this case rectangular) spatial observation window $[x_0, x_1] \times$

$[y_0, y_1]$ into κ cells each with constant intensity μ_k, $k \in \{1, \ldots, \kappa\}$. The background intensities for the κ cells can then be collected in the vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_\kappa)$. In order to simplify some of the upcoming expressions, it is helpful to define the expected number of background earthquakes in cell k, denoted as ν_k:

$$\nu_k = \mu_k \cdot (\text{area of cell } k) \cdot (\text{length of time window}), \tag{7}$$

where the length of the time window is T (see (5)). Depending on the situation, either μ_k or ν_k will be used in the formulas, since one is simply a fixed multiple of the other. The actual number of background events in cell k will be denoted as n_k and is modeled as a Poisson random variable with expectation ν_k.

If the complete branching structure of an observed ETAS process were known (including whether an event is a background event or a triggered event), i.e. if the unobserved quantities u_i were known for all i, the *complete data log-likelihood*, $\ell_c(\boldsymbol{\theta})$, could be written as

$$\begin{aligned}
\ell_c(\boldsymbol{\theta}) \;=\; & \sum_{k=1}^{\kappa} \{-\log(n_k!) - \nu_k + n_k \log(\nu_k)\} \\
& + \sum_{i} \{-\log(l_i!) - G_i(\boldsymbol{\theta}) + l_i \log(G_i(\boldsymbol{\theta}))\} \\
& + \sum_{i:u_i \neq 0} \Big\{ \log(\omega) + \omega \log(c) + \log(\rho) + \rho \log(d) \\
& \qquad -(1+\rho) \log\left((x_i - x_{u_i})^2 + (y_i - y_{u_i})^2 + d\right)
\end{aligned}$$

$$-(1+\omega)\log(t_i - t_{u_i} + c) - \log(\pi)\Big\}, \tag{8}$$

where the first sum relates to the actual number of background events in each of the κ cells. The second sum relates to the number of direct aftershocks l_i (in this case triggered by earthquake i) which also follows a Poisson distribution and whose expectation will be denoted as $G_i(\boldsymbol{\theta})$. Using the triggering function $g(\cdot)$ defined in (4), $G_i(\boldsymbol{\theta})$ can be derived as

$$\begin{aligned} G_i(\boldsymbol{\theta}) &= \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty g(t - t_i, x - x_i, y - y_i)\,\mathrm{d}x\mathrm{d}y\mathrm{d}t \\ &= K_0 \pi \frac{d^{-\rho} c^{-\omega}}{\rho\ \omega} e^{a(m_i - M_0)} \end{aligned}$$

The third sum of (8) is due to the space-time distribution of aftershocks (relative to their triggering events) and its density can be figured by dividing the triggering function $g(t - t_i, x - x_i, y - y_i)$ by the expected number of aftershocks $G_i(\boldsymbol{\theta})$.

Complete Data Maximum Likelihood Estimation.

Generally, the complete data log-likelihood $\ell_c(\boldsymbol{\theta})$ cannot be maximized in practice, since the branching structure is unobservable. However, the use of the EM algorithm will allow a probabilistic incorporation of the branching structure. To aid in the exposition of the implementation of the EM algorithm, assume for the moment that the quantities u_i are indeed known for all i. In the following, we will present an algorithm which maximizes

$\ell_c(\boldsymbol{\theta})$ using the unobservable quantities and hence produces the (in practice unattainable) *complete data MLE*, denoted as $\tilde{\boldsymbol{\theta}}$. In order to find $\tilde{\boldsymbol{\theta}}$ consider the partial derivatives of $\ell_c(\boldsymbol{\theta})$ with respect to each of the components of $\boldsymbol{\theta}$:

$$
\begin{aligned}
0 \stackrel{!}{=} \frac{\partial \ell_c(\boldsymbol{\theta})}{\partial \nu_k} &= -1 + \frac{n_k}{\nu_k}, && (9)\\
0 \stackrel{!}{=} \frac{\partial \ell_c(\boldsymbol{\theta})}{\partial K_0} &= -\frac{1}{K_0}\sum_i (G_i(\boldsymbol{\theta}) - l_i), && (10)\\
0 \stackrel{!}{=} \frac{\partial \ell_c(\boldsymbol{\theta})}{\partial a} &= -\sum_i ((m_i - M_0)G_i(\boldsymbol{\theta}) - (m_i - M_0)l_i), && (11)\\
0 \stackrel{!}{=} \frac{\partial \ell_c(\boldsymbol{\theta})}{\partial c} &= \sum_{i:u_i \neq 0} \left(\frac{\omega}{c} - \frac{1+\omega}{t_i - t_{u_i} + c} \right) + \frac{\omega}{c}\sum_i (G_i(\boldsymbol{\theta}) - l_i), && (12)\\
0 \stackrel{!}{=} \frac{\partial \ell_c(\boldsymbol{\theta})}{\partial \omega} &= \sum_{i:u_i \neq 0} \left(\frac{1}{\omega} + \log(c) - \log(t_i - t_{u_i} + c) \right) \\
&\qquad + \left(\frac{1}{\omega} - \log(c) \right) \sum_i (G_i(\boldsymbol{\theta}) - l_i), && (13)\\
0 \stackrel{!}{=} \frac{\partial \ell_c(\boldsymbol{\theta})}{\partial d} &= \sum_{i:u_i \neq 0} \left(\frac{\rho}{d} - \frac{1+\rho}{(x_i - x_{u_i})^2 + (y_i - y_{u_i})^2 + d} \right) \\
&\qquad + \frac{\rho}{d}\sum_i (G_i(\boldsymbol{\theta}) - l_i), && (14)\\
0 \stackrel{!}{=} \frac{\partial \ell_c(\boldsymbol{\theta})}{\partial \rho} &= \sum_{i:u_i \neq 0} \left(\frac{1}{\rho} + \log(d) - \log\left((x_i - x_{u_i})^2 + (y_i - y_{u_i})^2 + d \right) \right) \\
&\qquad + \left(\frac{1}{\rho} - \log(d) \right) \sum_i (G_i(\boldsymbol{\theta}) - l_i). && (15)
\end{aligned}
$$

The parameters for the background intensity can be derived from (9), which in fact represents several equations with $k = 1, 2, \ldots, \kappa$. The complete data MLE is $\tilde{\nu}_k = n_k$ and $\tilde{\mu}_k$ can then be computed as in (7). In principle, all other parameters have to be estimated simultaneously by solving (10)–(15),

which can be extremely computationally intensive. While we will abstain from formally deriving the properties of the following stepwise algorithm, we will provide an intuition for why it converges to the maximum of $\ell_c(\boldsymbol{\theta})$ and hence to the complete data MLE of $\boldsymbol{\theta}$.

Consider the equations (12) and (13) which relate to the temporal parameters c and ω of the aftershock distribution. Note that (12) and (13) depend only on c and ω if the terms after the first sums in each of the equations are ignored. In this stepwise algorithm, this may be reasonable, as (10) will subsequently guarantee that $\sum_i (G_i(\boldsymbol{\theta}) - l_i)$ equals zero. Setting only the first sums of (12) and (13) to zero can be interpreted as using only the time passed between the triggering event and aftershock in order to estimate the temporal parameters while ignoring the contribution to that part of the log-likelihood which relates to the *number* of aftershocks triggered by each earthquake. In this sense, a partial information approach is used to estimate c and ω in this step.

Setting only the first terms of (12) and (13) to zero implies the following system of equations:

$$\frac{\omega}{(1+\omega)c} = \frac{1}{L}\sum_{i:u_i\neq 0}\frac{1}{t_i - t_{u_i} + c}, \tag{16}$$

$$\frac{1}{\omega} + \log(c) = \frac{1}{L}\sum_{i:u_i\neq 0}\log(t_i - t_{u_i} + c), \tag{17}$$

where L is the number of triggered earthquakes, i.e. the total number of earthquakes minus the number of background events. This system of equa-

tions may be solved iteratively, suggesting the following procedure:

Algorithm 1

Step 0 Set $\tilde{c}_{\text{current}}$ to some strictly positive value.

Step 1 Compute the right hand sides of (16) and (17) using $\tilde{c}_{\text{current}}$.

Step 2 Find $\tilde{c}_{\text{new}}$ and $\tilde{\omega}_{\text{new}}$ by solving (16) and (17) using the quantities computed in the previous step.

Step 3 If $\Delta\tilde{c} = \tilde{c}_{\text{new}} - \tilde{c}_{\text{current}}$ and $\Delta\tilde{\omega} = \tilde{\omega}_{\text{new}} - \tilde{\omega}_{\text{current}}$ are smaller than some stopping criterion, stop. Otherwise, the new estimates become the current estimates and repeat steps 1–3.

Notice that Step 2 requires a numerical approach to solving (16) and (17). However, finding zeros is a computationally simpler problem than maximizing a function such as (5). Moreover, only two parameters have to be estimated in this part of the estimation procedure as opposed to six if solving (10)–(15) simultaneously.

The spatial parameters d and ρ can be estimated analogously, as (14) and (15) are structurally identical to (12) and (13), respectively, with the squared distance between aftershock and triggering event replacing the time elapsed between the two events.

Finally, if the parameters of the space-time distribution of aftershocks were known, the remaining parameters of $\boldsymbol{\theta}$, K_0 and a, which govern the number of triggered aftershocks, could be estimated as the parameters of a

Poisson regression. In fact, (10) and (11) are the estimating equations of a Poisson regression in which the number of triggered earthquakes (dependent variable) is Poisson with expectation depending on the magnitude of the triggering event through $G_i(\boldsymbol{\theta})$. In order to take advantage of this circumstance in an iterative algorithm, one can condition on the current parameter estimates of c, ω, d, and ρ in (10) and (11) and then iterate between estimating those parameters on the one hand and K_0 and a on the other hand.

A similar conditional likelihood approach has been investigated previously by Ogata and Akaike (1982) for a temporal self-exciting point process model for earthquake occurrences. Ogata and Akaike hold the parameter which is equivalent to the parameter a in the model discussed here constant and maximize the resulting conditional log-likelihood to estimate the other parameters of the model. They then update a using Akaike's (1974, 1977) Information Criterion (AIC) and iterate between these two steps. Our approach applies the conditional likelihood methodology in both steps of this procedure and does not require the use of any information criteria to direct the algorithm. The following algorithm summarizes the abovementioned procedure:

Algorithm 2

Step 0 Set $n = 1$ and set each component of $\tilde{\boldsymbol{\theta}}^{(1)}$ to some strictly positive value.

Step 1 Estimate the background intensity vector $\tilde{\boldsymbol{\mu}}$ as described above.

Step 2.1 Fix the space-time parameters at their current estimates $\tilde{c}^{(n)}$,

$\tilde{\omega}^{(n)}$, $\tilde{d}^{(n)}$, and $\tilde{\rho}^{(n)}$. Iteratively solve (10) and (11) in order to find the estimates $\tilde{K}_0^{(n+1)}$ and $\tilde{a}^{(n+1)}$.

Step 2.2 Find $\tilde{c}^{(n+1)}$ and $\tilde{\omega}^{(n+1)}$ as described in Algorithm 1.

Step 2.3 Find $\tilde{d}^{(n+1)}$ and $\tilde{\rho}^{(n+1)}$ analogously to Algorithm 1.

Step 2.4 If $\Delta\tilde{\boldsymbol{\theta}}^{(n+1)} = \tilde{\boldsymbol{\theta}}^{(n+1)} - \tilde{\boldsymbol{\theta}}^{(n)}$ is smaller than some stopping criterion stop. Otherwise, increase n by one and repeat steps 2.1–2.4.

Maximum Likelihood Estimation Using the EM Algorithm.

Since the quantities u_i are unknown (and in fact unobservable), the complete data log-likelihood cannot be expressed (nor maximized) directly in practice. However, the *expected* complete data log-likelihood can be computed (E-step) and then maximized (M-step). The E-step of the EM algorithm requires estimating the triggering probabilities $\text{prob}^{(n+1)}(u_i = j)$ for all i, j based on a current estimate $\hat{\boldsymbol{\theta}}_{EM}^{(n)}$:

$$\text{prob}^{(n+1)}(u_i = j) = \frac{g\left(t_i - t_j, x_i - x_j, y_i - y_j, m_j \mid \hat{\boldsymbol{\theta}}_{EM}^{(n)}\right)}{\hat{\mu}_{k:\, i \in \text{ cell } k}^{(n)} + \sum_{r=1}^{i-1} g\left(t_i - t_r, x_i - x_r, y_i - y_r, m_r \mid \hat{\boldsymbol{\theta}}_{EM}^{(n)}\right)}. \tag{18}$$

These probabilities allow finding expressions for the expected number of background events in cell k, $\hat{n}_k^{(n+1)}$ and the expected number of direct aftershocks triggered by each earthquake i, $\hat{l}_i^{(n+1)}$:

$$\hat{n}_k^{(n+1)} = \sum_{i \in \text{ cell } k,\, i \geq 2} \left(1 - \sum_{j=1}^{i-1} \text{prob}^{(n+1)}(u_i = j)\right),$$

$$\hat{l}_i^{(n+1)} = \sum_{s \geq i+1} \text{prob}^{(n+1)}(u_s = i).$$

The expected complete data log-likelihood can then be written as

$$\begin{aligned}
E_{\hat{\boldsymbol{\theta}}_{EM}^{(n)}}[\ell_c(\boldsymbol{\theta})] &= \sum_k \Big\{ -\log(\hat{n}_k^{(n)}!) - \nu_k + \hat{n}_k^{(n)} \log(\nu_k) \Big\} \\
&\quad + \sum_i \Big\{ -\log(\hat{l}_i^{(n)}!) - G_i(\boldsymbol{\theta}) + \hat{l}_i^{(n)} \log(G_i(\boldsymbol{\theta})) \Big\} \\
&\quad + \sum_{i \geq 2} \sum_{j=1}^{i-1} \text{prob}^{(n)}(u_i = j) \\
&\quad \cdot \Big\{ \log(\omega) + \omega \log(c) + \log(\rho) + \rho \log(d) \\
&\qquad -(1+\rho) \log\left((x_i - x_j)^2 + (y_i - y_j)^2 + d \right) \\
&\qquad -(1+\omega) \log(t_i - t_j + c) - \log(\pi) \Big\}. \qquad (19)
\end{aligned}$$

The M-step of the EM algorithm maximizes (19). In principle, this maximization can be performed using Algorithm 2 after replacing the quantities n_k and l_i with their counterparts in expectation, that is $\hat{n}_k^{(n)}$ and $\hat{l}_i^{(n)}$, and modifying (12)–(15) to reflect the estimated probabilities as shown in (19). The equations (16) and (17) which are used to estimate the temporal parameters c and ω take on the following form:

$$\frac{\omega}{(1+\omega)c} = \frac{1}{\hat{L}^{(n)}} \sum_{s \geq 2} \sum_{r=1}^{s-1} \text{prob}^{(n)}(u_s = r) \cdot \frac{1}{t_s - t_r + c}, \qquad (20)$$

$$\frac{1}{\omega} + \log(c) = \frac{1}{\hat{L}^{(n)}} \sum_{s \geq 2} \sum_{r=1}^{s-1} \text{prob}^{(n)}(u_s = r) \cdot \log(t_s - t_r + c), \qquad (21)$$

where $\hat{L}^{(n)}$ is the expected number of triggered earthquakes in the data set, $\sum_i \hat{l}_i^{(n)}$. The corresponding equations for estimating d and ρ are analogous. The following summarizes our proposed algorithm to estimate the parameters of the ETAS model:

Algorithm 3

Step 0 Set $n = 1$ and set each component of $\hat{\boldsymbol{\theta}}_{EM}^{(1)}$ to some strictly positive value.

Step 1 (E-Step) Based on $\hat{\boldsymbol{\theta}}_{EM}^{(n)}$, estimate the triggering probabilities as shown in (18).

Step 2 (M-Step) Maximize (19) using a variant of Algorithm 2 where n and l_i are replaced by their current expectations $\hat{n}_k^{(n)}$ and $\hat{l}_i^{(n)}$ for all k, i, respectively and using (20) and (21) instead of (16) and (17) in order to update the temporal parameter estimates for c and ω. Proceed analogously in order to update the spatial parameter estimates for d and ρ. The new estimate will be denoted as $\tilde{\boldsymbol{\theta}}_{EM}^{(n+1)}$.

Step 3 If $\Delta\tilde{\boldsymbol{\theta}}_{EM}^{(n+1)} = \tilde{\boldsymbol{\theta}}_{EM}^{(n+1)} - \tilde{\boldsymbol{\theta}}_{EM}^{(n)}$ is smaller than some stopping criterion stop. Otherwise, increase n by one and repeat steps 1–3.

Variations aimed at improving the speed of the algorithm are possible. In the M-Step, for instance, Algorithm 2 may run only once (i.e. Step 2.4 of Algorithm 2 may be ignored) since Step 3 of Algorithm 3 already checks whether a further run of the algorithm could improve the estimates in any substantial way. Other variations are possible.

Dempster, Laird, and Rubin (1977) showed that under general conditions, estimates obtained using the EM algorithm converge to the ML estimates. As with other optimization algorithms, EM may converge to a local maximum or saddle point, but the incorporation of the probabilistic branching structure often leads to unique maxima, sometimes with closed form solutions.

Figure 5 demonstrates the robustness and accuracy of the algorithm introduced in this work. Using the same starting values as the ones used to present the difficulties of ML estimation (see Figure 4), the EM-type algorithm converges to an estimate of $\boldsymbol{\theta}$ very close to the 'true' value in all eight situations.

In a more systematic approach, 100 earthquake catalogs are simulated according to the model (3) using the parametrization in Table 1 and then estimated using the proposed EM-type algorithm and conventional ML, in which the results of our proposed algorithm are used as starting values. Figures 6 and 7 show that the sampling distributions for all components of $\boldsymbol{\theta}$ are quite similar for both methods. However, the EM-type algorithm seems to provide estimates which are considerably less biased, as shown in Table 2 (the only exception is the background parameter $\boldsymbol{\mu}$).

A possible explanation for why the EM-type algorithm yields superior results is that most theoretical results relating to ML estimation only hold asymptotically (see for instance Ferguson (1996) for a genreal treatment of this matter and Ogata (1978) who derived analogous results for point process models). In our simulations, the number of simulated earthquakes may

not be large enough for reaching an asymptotic regime in the numerical maximization of the likelihood function. However, the number of simulated earthquakes might be large enough for the EM-procedure to produce accurate results. In fact, for the numerical maximization routine the theoretical standard errors based on the Hessian matrix are substantially larger than the ones derived by simulation, which also suggests that the asymptotic regime assumed in the theoretical framework has not been attained.

In order to demonstrate the robustness of the proposed EM-type algorithm with respect to starting values, ten earthquake catalogs are simulated with the same parametrization as above, and then estimated using 100 different starting values for $\boldsymbol{\theta}$. The starting values are sampled from a uniform distribution whose range is $\frac{1}{5}$ of the parameter value to 5 times the parameter value. The results indicate that the parameter estimates are affected only minimally by even substantial offsets in starting values. The largest observed difference between the smallest and the largest parameter estimate for a component of $\boldsymbol{\theta}$ is less than 0.5% of the true parameter value, the average being less than 0.1%. It may be relevant to note that the small variability associated with different starting values can be controlled by the convergence criterion used. In this implementation, our proposed algorithm was halted as soon as each component of $\boldsymbol{\theta}$ had converged to four significant digits.

5 Application to earthquakes occurrences in Southern California

The methodology introduced in the previous Section can be used to estimate background seismicity rates in Southern California using the ETAS model (3). Data on earthquakes in Southern California are compiled by the Southern California Earthquake Center (SCEC) and include occurrence times, magnitudes, and locations based on measurements taken by a network of almost 400 seismographic stations throughout Southern California. The catalog is maintained by the Southern California Seismic Network (SCSN), a cooperative project of the California Institute of Technology and the United States Geological Survey and is publicly available at http://www.data.scec.org. The data set used in this work spans the time period between 01/01/1984 and 06/17/2004 and is considered complete above $M_0 = 3$ (Kagan (2002), Kagan (2003)). The catalog consists of 6796 earthquakes occurring in a rectangular area around Los Angeles between longitudes $-122°$ and $-114°$ and latitudes 32° and 37° (approximately 733km$\times$556km).

One problem of interest is to estimate spatial background intensities, which represent the occurrence rate of spontaneous, untriggered earthquakes. One way to approach this problem is to use the stochastic decustering method introduced by Zhuang, Ogata, and Vere-Jones (2002, 2004, 2005), who use a spatial kernel density smoother as an estimate of $\mu(x, y)$ in (3). As an

alternative, one may instead incorporate known geological features of Southern California and estimate background intensities in the ETAS model for geologically distinct regions within the study area. We use a regionalization proposed by Zaliapin, Keilis-Borok, and Axen (2002) who identify seven distinct regions based on fault orientation, historical slip, and tectonic setting. The resulting regions are in agreement with the main geological and fault activity maps for California (Jennings 1977, 1994). Region 1 covers the Southwestern offshore continental borderland of Southern California where Northwest striking right-lateral faults dominate. Region 2 covers the Southeast portion of the study area, from the Salton Trough well into Northern Mexico. This region covers the Southern section of the San Andreas fault system where it branches out into several Northwest striking right-lateral faults. The Northwestern part of Southern California is characterized by the creeping section of the San Andreas fault (region 3). Region 4, in which the 1994 Northridge earthquake occurred, includes the Western Transverse Ranges with East striking left-lateral and normal faults covering the major part of the Los Angeles metropolitan area. The narrow region 5 includes the left-lateral striking Garlock fault system whose orientation is East-West. Region 6 mainly covers the Mojave block with mostly Northwest striking right-lateral faults and also the Eastern Transverse Ranges, where East striking left-lateral faults and thrust faults are present. The two largest earthquakes in the space-time area studied in this work occurred here (the 1992 Landers and 1999 Hector Mine earthquakes). The Western Great Basin, Northeast of

the Garlock fault system, is characterized by North and Northwest striking right-lateral and normal faults (region 7). Finally, region 0 incorporates areas of relative seismic stability not covered in the regions outlined above. For a more detailed description of this regionalization, see Zaliapin, Keilis-Borok, and Axen (2002).

The space-time ETAS model (3) with $\mu(x, y)$ varying across seismic regions is estimated using the EM-type algorithm introduced in the previous section. A conventional maximization of the log-likelihood could be challenging, as the parameter estimates can be heavily influenced by a poor choice of starting values as shown in Section 3. Moreover, judging from the simulations described in Section 4, the conventional ML approach may have an increased bias compared to the EM-type algorithm, since it seems to require a substantially larger sample size in order to attain the asymptotic regime guaranteeing consistent estimation.

The estimation results are presented in Table 3 and Figure 8. The declustered background intensities for six of the seven regions are quite similar and range between 3.500×10^{-3} and 6.970×10^{-3} events per day per squared degree. This is a remarkable result, as strong earthquakes with large aftershock clusters were observed in the Western Transverse Ranges (region 4) and the Mojave block (region 6) with 817 and 2822 earthquakes, respectively, while the Garlock fault system (region 5) had very few (148) observed seismic events. Nevertheless, our declustered background seismicity estimates suggest that the occurrence rate of spontaneous, untriggered earthquakes is

similar in six of the seven regions. Only region 2, the area covering the Southern section of the San Andreas fault system to the East of San Diego, has a substantially higher background intensity of 16.945×10^{-3} events per day per squared degree.

In a recently published article, Fialko (2006) finds that the Southern section of the San Andreas fault system shows a high level of seismic strain accumulation. Given that the last two great Californian earthquakes in 1857 and 1906 ruptured the middle and Northern sections of the San Andreas fault, respectively, it is believed that faults in the Southern section of the fault system currently pose the highest seismic risk in California (see Fialko (2006) and references therein). This is supported by palaeoseismological evidence estimating the average recurrence time of large earthquakes in that area to be between 200 and 300 years together with the fact that no such large event (magnitude 7 or larger) has been observed in the last 250 years. Fialko concludes that the Southern San Andreas fault system may be in the late phase of its interseismic recurrence.

The ETAS model usually assumes that the earthquake magnitude distribution is separable from the space-time features of the model. Therefore, in this framework, an increased background seismicity rate in a particular area does not imply an increased risk of a large-magnitude event other than to the extent that more strong earthquakes may occur simply because more earthquakes are expected in general. However, our analysis does show an elevated seismic activity in the region covering the Southern section of the

San Andreas fault system and in this sense an elevated seismic risk exists in this region compared to all other regions in Southern California.

6 Discussion and concluding remarks

In this work, we presented an EM-type algorithm which maximizes the expected complete data log-likelihood function. The advantages of this algorithm compared to conventional ML estimation were substantial for the case of estimating the space-time ETAS model, in terms of convergence, bias, and robustness to choice of starting values.

This methodology should also work well for other specifications of ETAS models. In fact, it is applicable to all kinds of branching process models where the information of which event 'triggers' which other event is not observable but can be described probabilistically, as this allows the incorporation of the branching structure in expectation.

As the simulations in Section 3 illustrate, if the number of observations is limited, our proposed estimation procedure may in fact yield more accurate results, as the asymptotic regime under which the desirable properties of conventional ML are well-known may not be reached. In addition, even in cases where a direct numerical maximization of the log-likelihood may be preferred, the EM-type algorithm proposed here may be a useful way of obtaining starting values for the optimization routine. Further, the reliability with which the EM-type procedure converges to a reasonable estimate may

be especially attractive for simulation studies in which one is interested in repeatedly simulating and estimating a model, e.g. in order to estimate the bias or variance of certain parameter estimates. Such repetitions are very difficult using ML due to the required oversight involved and occasional lack of convergence.

Our seismological application involved the use of geologically distinct regions within Southern California, incorporating the known morphological and tectonic conditions present in this region. However, it should be noted that the EM-type algorithm could just as well be combined with a spatial kernel smoothing method as used by Zhuang, Ogata, and Vere-Jones (2002, 2004, 2005) in order to estimate continuous background intensities.

Our results suggest that the declustered background intensity in the area covering the Southern section of the San Andreas fault system to the East of San Diego is substantially higher than in other regions of Southern California. While our analysis does not directly imply an increased risk of a strong earthquake occurring in this area as suggested for instance by Fialko (2006) and references therein, it does single out this region as markedly different from all other regions in Southern California.

Note, however, that while Southern California appears to be rather naturally divided into distinct seismic regions, a subjective element in outlining the regions remains. Changes to the borders of the regions could in fact change the background intensity estimates, as could misspecification of the ETAS model. For instance, the space-time model employed in this work uses

circular aftershock regions, whereas spatial aftershock distributions observed in Southern California are rather elliptical with mainly North-Northwest and East-West orientations. The investigation and application to California seismicity of ETAS models with non-circular aftershock distributions are important directions for further research.

7 Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 0306526. We thank Yan Kagan, Yingnian Wu, and Ilya Zaliapin for helpful comments, and the Southern California Earthquake Center for its generosity in sharing their data. All computations were performed in R.

References

Akaike, H. (1974): "A new look at statistical model identification," *IEEE Transactions on Automatic Control*, AU–19, 716–722.

——— (1977): "On Entropy Maximization principle," in *Application of Statistics*, ed. by P. R. Krishnaiah, pp. 27–41. North-Holland.

Cox, D. R. (1975): "Partial likelihood," *Biometrika*, 62, 269–276.

Daley, D., and D. Vere-Jones (2003): *An Introduction to the Theory of Point Processes*. Springer, New York, second edn.

Dempster, A., N. Laird, and D. Rubin (1977): "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, 39(1), 1–38.

Ferguson, T. S. (1996): *A Course in Large Sample Theory*. Chapman & Hall, London.

Fialko, Y. (2006): "Interseismic strain accumulation and the and the earthquake potential on the southern San Andreas fault system," *Nature*, 441.

Gutenberg, B., and C. F. Richter (1944): "Frequency of Earthquakes in California," *Bulletin of the Seismological Society of America*, 34.

Hawkes, A. G. (1971a): "Point spectra of some mutually exciting point processes," *Journal of the Royal Statistical Society, Series B*, 33(3), 438–443.

——— (1971b): "Spectra of some self-exciting and mutually exciting point processes," *Biometrika*, 58, 83–90.

Hawkes, A. G., and L. Adamopoulos (1973): "Cluster models for earthquakes – regional comparisons," *Bulletin of the International Statistical Institute*, 45(3), 454–461.

Jennings, C. W. (1977): *Geological map of California, scale 1:750,000* California Division of Mines and Geology, Sacramento, CA.

——— (1994): *Fault activity map of California and adjacent areas, scale 1:750,000* California Division of Mines and Geology, Sacramento, CA.

Kagan, Y. Y. (2002): "Modern California earthquake catalogs and their comparison," *Seismological Research Letters*, 73(6), 921–929.

——— (2003): "Accuracy of modern global earthquake catalogs," *Physics of The Earth and Planetary Interiors*, 135, 173–209.

Lomnitz, C. (1974): *Global Tectonics and Earthquake Risk*. Elsevier, London.

McLachlan, G. J., and T. Krishnan (1996): *The EM Algorithm and Extensions*. Wiley-Interscience.

Ogata, Y. (1978): "The asymptotic behaviour of maximum likelihood estimators for stationary point processes," *Annals of the Institute of Statistical Mathematics*, 30, 243–261.

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

——— (1993): "Space-time modeling of earthquake occurrences," *Bulletin of the International Statistical Institute*, 55, 249–250.

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

——— (1999): "Seismicity Analysis through Point-process Modeling: A Review," *Pure and Applied Geophysics*, 155, 471–507.

Ogata, Y., and H. Akaike (1982): "On Linear Intensity Models for Mixed Doubly Stochastic Poisson and Self-Exciting Point Processes," *Journal of the Royal Statistical Society, Series B (Methodological)*, 44(1), 102–107.

R Development Core Team (2005): *R: A language and environment for statistical computing* R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, http://www.R-project.org.

Vere-Jones, D. (1970): "Stochastic Models for Earthquake Occurrence," *Journal of the Royal Statistical Society, Series B*, 32(1), 1–62.

——— (1975): "Sochastic Models for Earthquake Sequences," *Geophysical Journal of the Royal Astronomical Society*, 42, 811–826.

Zaliapin, I., V. Keilis-Borok, and G. Axen (2002): "Premonitory spreading of seismicity over the fault's network in southern California: Precursor Accord," *Journal of Geophysical Research*, 107, (B10), 2221, doi:10.1029/2000JB000034.

Zhuang, J., Y. Ogata, and D. Vere-Jones (2002): "Stochastic declustering of space-time earthquake occurrences," *Journal of American Statistical Association*, 97(458), 369–380.

——— (2004): "Analyzing earthquake clustering features by using stochastic reconstruction," *Journal of Geophysical Research*, 109, B05301, doi:10.1029/2003JB002879.

Table 1: **Specification of the space-time-magnitude ETAS model used for simulation.** The ETAS model is described in (3). Note that a homogeneous background intensity (measured in events per day per squared degree) is used which does not depend on the location. The time is measured in days, spatial distances are measured in degrees. A truncated exponential distribution (6) is used to simulate magnitudes.

parameter	value
$\mu(x, y)$	0.0008
K_0	0.0000305
a	2.3026
c	0.01
ω	0.5
d	0.015
ρ	0.8

space-time window for background events
$[0°, 8°] \times [0°, 5°] \times [0, 7500 \text{ days}]$

parameters of the magnitude distribution	
M_0	2
M_{GR}^{max}	8

Table 2: **Bias of the parameter estimates using the EM-type algorithm and a numerical maximization of the log-likelihood.** The bias of the proposed EM-type algorithm and the conventional ML procedure is compared by simulating 100 processes and estimating them using these two methods. For most parameters, the EM-type algorithm yields results that are closer to the 'true' parameter values described in Table 1.

	$\hat{\mu}$ $[\times 10^{-4}]$	$\hat{K}_0$ $[\times 10^{-5}]$	$\hat{a}$	$\hat{c}$	$\hat{\omega}$	$\hat{d}$	$\hat{\rho}$
'true' values	8.000	3.050	2.303	0.01000	0.500	0.01500	0.800
numerical MLEs	8.011	2.993	2.275	0.01086	0.519	0.01625	0.841
standard errors	(0.527)	(0.708)	(0.113)	(0.00283)	(0.058)	(0.00409)	(0.103)
bias in % of 'true' value	+0.14%	-1.86%	-1.22%	+8.56%	+3.80%	+8.35%	+5.13%
EM-type estimates	7.925	2.993	2.296	0.01019	0.501	0.01564	0.824
standard errors	(0.516)	(0.708)	(0.109)	(0.00265)	(0.056)	(0.00423)	(0.112)
bias in % of 'true' value	-0.94%	-1.85%	-0.27%	+1.91%	+0.20%	+4.30%	+3.00%

Table 3: **Estimation results of space-time ETAS model (3) for Southern California, 1984-2004, $\mathbf{M_0 = 3}$.** The estimated background intensities (measured in events per day per squared degree) are between 3.5×10^{-3} and 6.97×10^{-3} in six of the seven regions, including the two in which major earthquakes occurred. Region 2, however, has a substantially higher declustered background intensity of 16.945×10^{-3}. See Figure 8.

k	$\hat{\mu}_k$ [events per day per degree^2]	number of earthquakes in region	area of region [degree^2]	description of region
1	3.500×10^{-3}	501	6.37	SW; coastal and offshore region
2	16.945×10^{-3}	1357	3.26	SE; S. section of S.A. fault system
3	3.921×10^{-3}	476	6.14	NW; creeping section of S.A.
4	5.216×10^{-3}	817	3.85	Western transverse ranges
5	6.970×10^{-3}	148	1.68	Garlock fault system
6	5.294×10^{-3}	2822	3.43	Mojave block, E. transverse ranges
7	6.130×10^{-3}	637	2.75	Western Great Basin
0	0.019×10^{-3}	38	12.52	Rest

$\hat{K}_0$	$\hat{a}$	$\hat{c}$	$\hat{\omega}$	$\hat{d}$	$\hat{\rho}$
4.823×10^{-5}	1.034	0.01922	0.222	4.906×10^{-5}	0.497

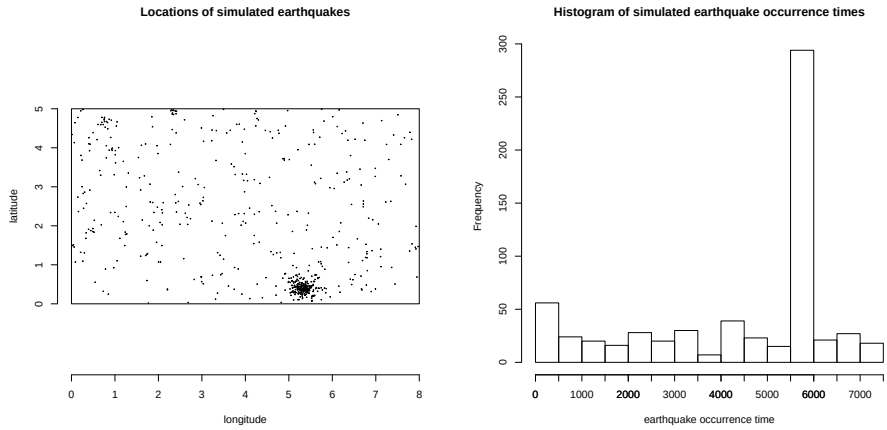

Figure 1: **Simulated earthquake process using a space-time-magnitude ETAS model.** This figure shows a simulated earthquake catalog using model (3) with a parametrization as described in Table 1. The simulated catalog consists of 638 earthquakes (241 background events and 397 aftershocks). The spatial distribution is presented on the left, although 32 (triggered) earthquakes are not shown because they are outside the specified space-time window. The temporal distribution is shown on the right. The spike in activity starting on day 5527 is caused by a magnitude 5.33 earthquake and its aftershocks and corresponds to the cluster on the lower right side of the spatial plot.

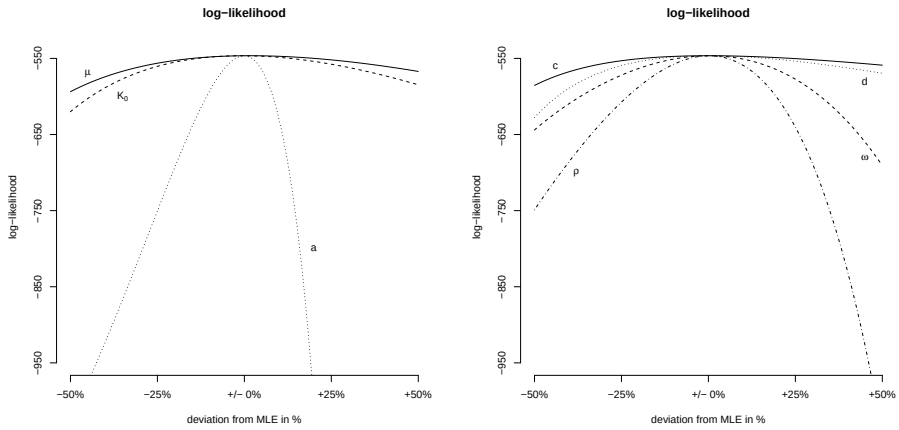

Figure 2: **Flatness of the log-likelihood function for varying one parameter at a time.** This figure demonstrates the relative flatness of the log-likelihood function for some components of the parameter vector $\hat{\boldsymbol{\theta}}$ (two graphs are shown to improve the legibility). The log-likelihood stays very flat when μ, K_0, c, and d are varied around their MLEs. This indicates potentially high standard errors and/or numerical challenges for the estimation of these parameters. The parameters a, ω, and ρ have much more peaked log-likelihood functions and can hence be estimated more easily.

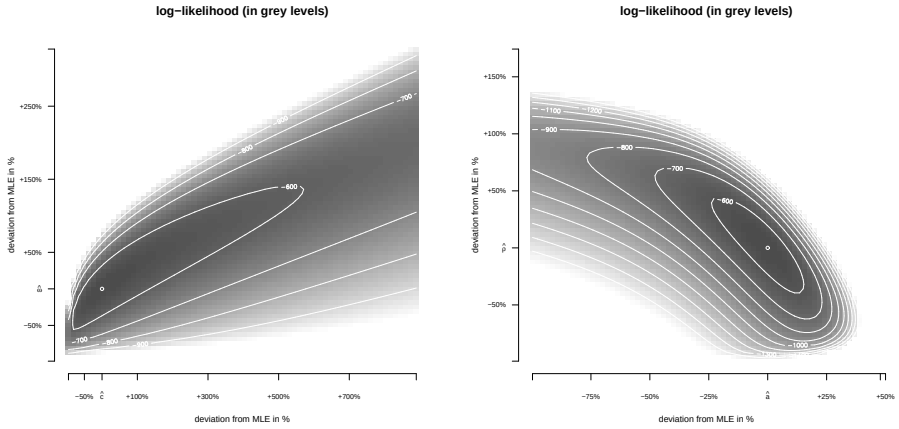

Figure 3: **Flatness of the log-likelihood in multidimensional settings.** The problem of flatness of the log-likelihood (shown in grey levels) function can be aggravated in a multidimensional context. In this analysis, pairs of components of $\hat{\boldsymbol{\theta}}$ ($\hat{c}$, $\hat{\omega}$ on the left and $\hat{a}$, $\hat{\rho}$ on the right) are varied, while all the other components are fixed at their MLEs (small white circle). Along certain trajectories, even large deviations from the 'true' values reduce the log-likelihood only marginally.

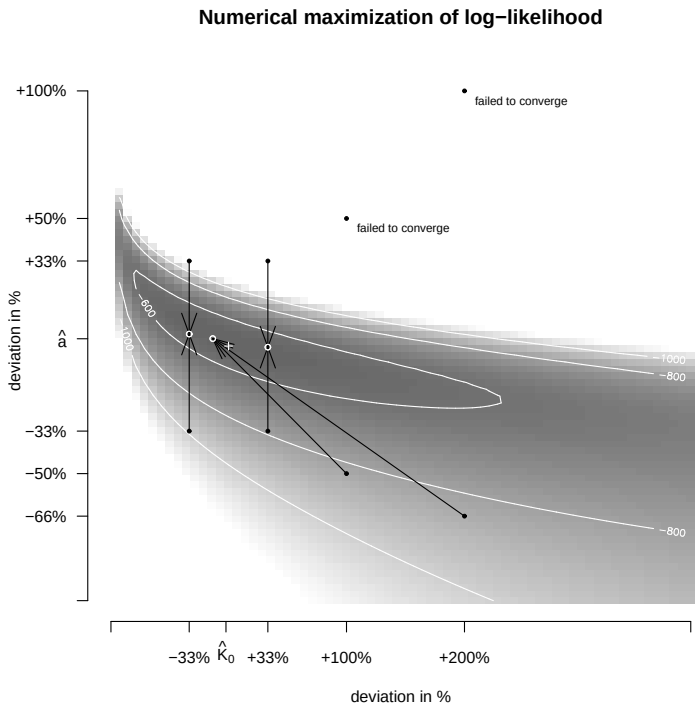

Figure 4: **Difficulties in estimating ETAS parameters when maximizing the log-likelihood numerically.** Except for K_0 and a, the starting values (black dots) for the components of $\boldsymbol{\theta}$ are set to the 'true' values (see Table 1). The white circles show the estimation results of the numerical maximization routine ($\hat{\boldsymbol{\theta}}_{num}$) and the '+' symbol depicts the location of the 'true' K_0 and a. The numerical maximization routine converges to an estimate close to $\boldsymbol{\theta}$ in only two cases. It fails to converge in two cases and seems incapable of improving the K_0 estimate for the four starting configurations in which K_0 is modified by 33%. This could be due to the relative flatness of the log-likelihood function (shown in grey-levels) with respect to variations of K_0.

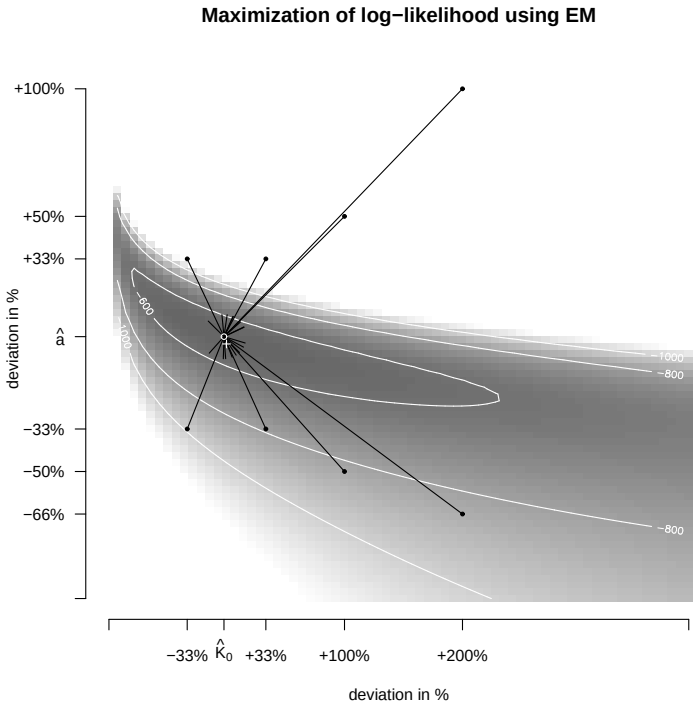

Figure 5: **Estimation of ETAS parameters using the EM-type algorithm.** Using the same starting values as in Figure 4 (black dots), the EM-type algorithm converges to an estimate close to the 'true' value in all eight situations. The white circles show the estimation results and the '+' symbol indicates the location of the 'true' $\boldsymbol{\theta}$.

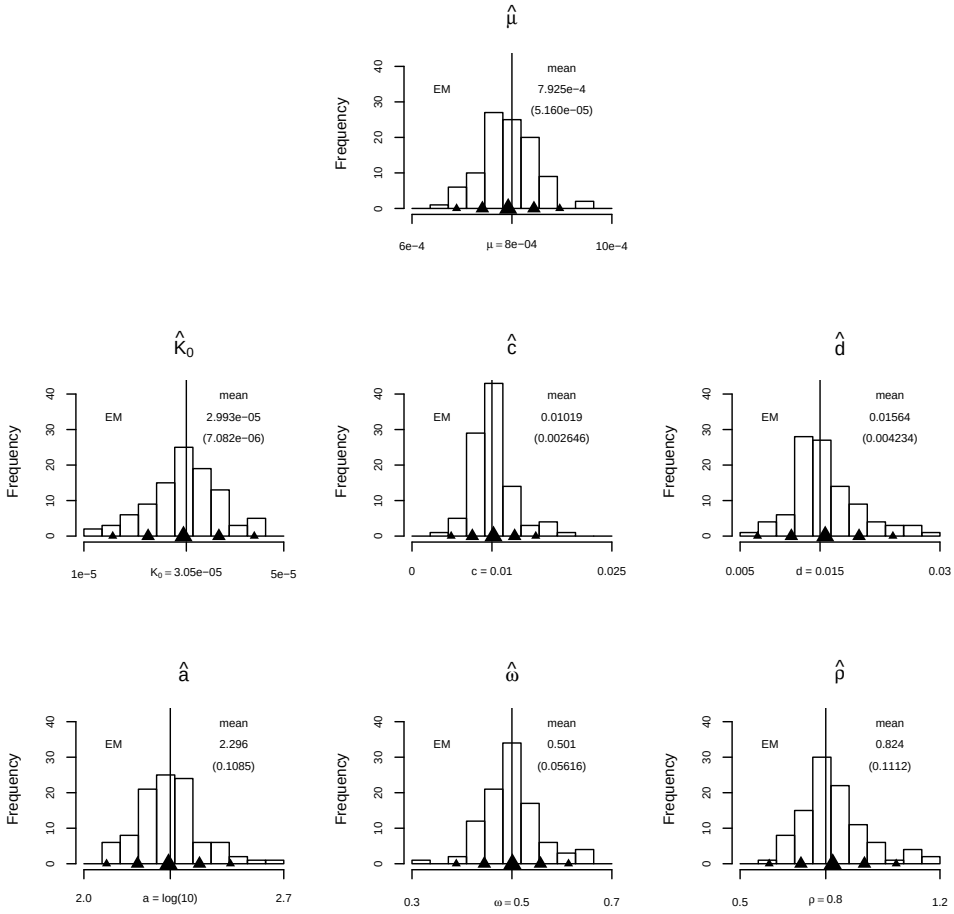

Figure 6: **Sampling distributions of parameter estimates using the EM-type algorithm.** Using the parameter values of Table 1, 100 earthquake catalogs were simulated and then estimated using the EM-type algorithm proposed in this work. The estimates do not seem to be substantially biased and standard errors are reasonably small. The vertical lines on each histogram indicate the 'true' parameter values. The location of the mean based on the 100 estimates is shown as the large triangle on the bottom of each histogram. The standard error is shown in parentheses and one as well as two standard error intervals are represented by the middle and small triangles, respectively.

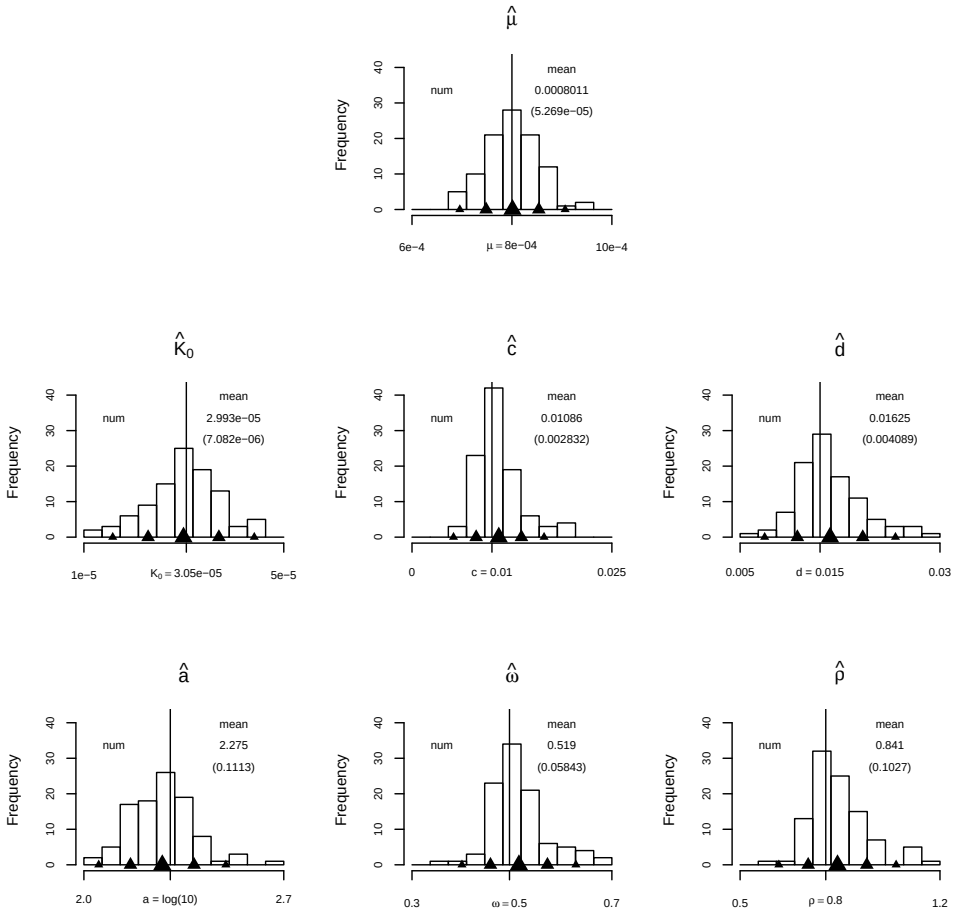

Figure 7: **Sampling distributions of parameter estimates when numerically maximizing the log-likelihood.** The ETAS parameters for the same 100 simulated earthquake catalogs as in Figure 6 are estimated using ML, where the results of the EM-type algorithm serve as starting values. The sampling distributions are quite similar but the estimates seem to be slightly more biased.

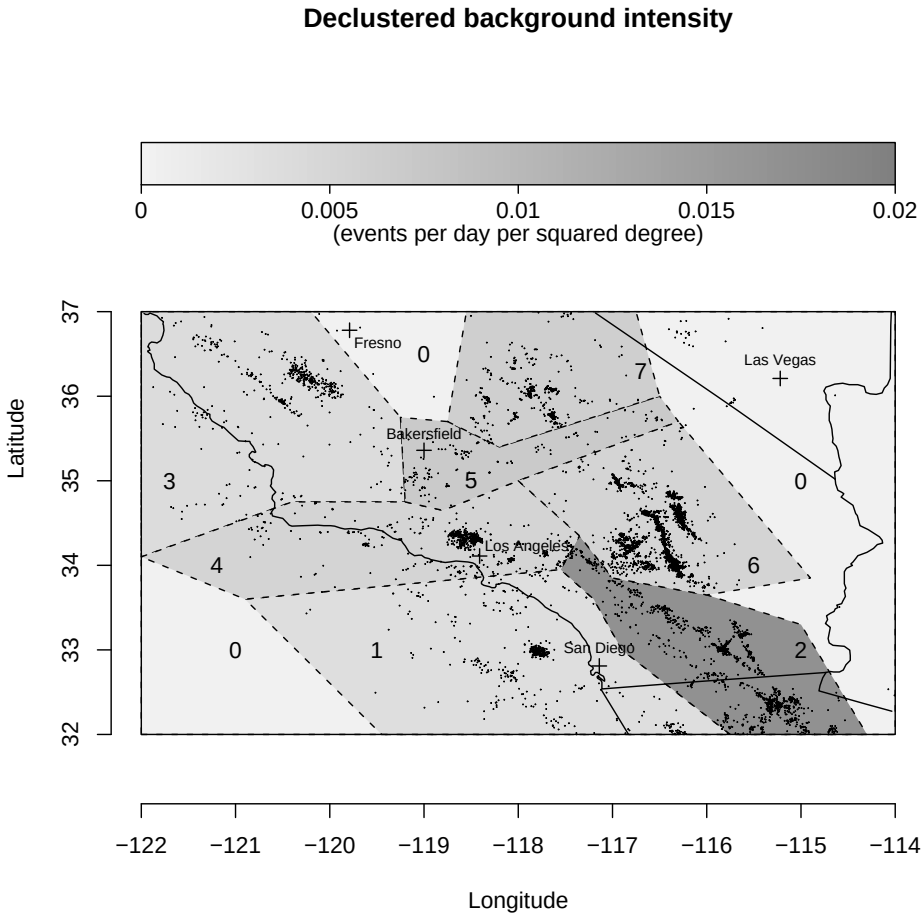

Figure 8: **Declustered background intensity estimates for Southern California.** Background intensities are estimated for geologically distinct regions in Southern California using the space-time ETAS model (3). The declustered background intensities (measured in events per day per squared degree) are quite similar for six of the seven regions including the ones in which the largest earthquakes and aftershock clusters were observed. The only region with a substantially larger background rate is the Southern section of the San Andreas fault system to the East of San Diego (region 2).