Estimation of Space–Time Branching Process Models in Seismology Using an EM–Type Algorithm

Alejandro VEEN and Frederic P. SCHOENBERG

Maximum likelihood estimation of branching point process models via numerical optimization procedures 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 expected complete data 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.

KEY WORDS: Branching process models; Earthquakes; Epidemic-type aftershock sequence model; Expectation-maximization algorithm; Maximum likelihood; Space–time point process models.

1. INTRODUCTION

Point process models have long been used to describe earthquake occurrences (Vere-Jones 1970, 1975). See Ogata (1999) for 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 branching process models in which all earthquakes can trigger aftershocks, and among these, the epidemic-type aftershock sequence (ETAS) model is considered to be one of 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 nonlinear, 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 in order to attain maximum likelihood estimates (MLEs) (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 complete data 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. Using an analogous expression, conventional ML maximizes the *incomplete data log-likelihood*.

In this article, we show how an EM-type algorithm can be combined with a partial information approach in certain steps. This relates to partial likelihood maximization, which was introduced by Cox (1975) and which was 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), we are able to present a highly robust and accurate estimation procedure that can be used to estimate even very complex branching process models. 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 conventional ML procedure using a numerical optimization approach.

Following a description in Section 2 of self-exciting point process models for earthquake occurrences, we describe some of the problems with conventional ML 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, and the article will conclude with a discussion in Section 6.

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, nondecreasing, $\mathcal{H}$-predictable process such that $N([0,t)) - \int \lambda(t|\mathcal{H}_t)\,dt$ is an $\mathbf{H}$-martingale. In this representation, $\mathbf{H}$ must contain the history of the process up to time t, denoted as $\mathcal{H}_t = \{t_i : t_i < t\}$ with t_i as the time event i occurs, but $\mathbf{H}$ may contain additional information as well. Because the finite-dimensional distributions 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 0 otherwise, and $\int_0^\infty g(v)\,dv < 1$ in order to ensure stationarity (Hawkes 1971a,b). Early applications of self-exciting point processes to earthquake occurrence models can be found in Hawkes and

Alejandro Veen is Researcher, IBM T. J. Watson Research Center, Yorktown Heights, NY 10598 (E-mail: *aveen@us.ibm.com*). Frederic P. Schoenberg is Professor, Department of Statistics, University of California, Los Angeles, CA 90095-1554 (E-mail: *frederic@stat.ucla.edu*). This material is based on work supported by the National Science Foundation under grant 0306526. We thank Yan Kagan, Yingnian Wu, Ilya Zaliapin, and anonymous referees for helpful comments, and the Southern California Earthquake Center for its generosity in sharing its data. All computations were performed in R.

© 0 American Statistical Association
Journal of the American Statistical Association
???? 0, Vol. 0, No. 00, Applications and Case Studies
DOI 10.1198/016214508000000148

Adamopoulos (1973), Lomnitz (1974, chap. 7), and Kagan and Knopoff (1987).

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:

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

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

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

where $\tau_i = t - t_i$ is the time elapsed since 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 "cutoff magnitude," that is, the lowest earthquake magnitude in the dataset under consideration. The term $K_0/(\tau_i + c)^{(1+\omega)}$ describing the temporal distribution of aftershocks is known 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 1993b, 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:

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

with triggering function

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

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 the history of the process up to time t is now defined as $\mathcal{H}_t = \{(t_i, x_i, y_i, m_i) : t_i < t\}$. One characteristic of this model is that the aftershock zone does not scale with the magnitude of the triggering event. It has been suggested that aftershock zones increase with main shock magnitudes, and some recent research seems to support this view (Kagan 2002b). Examples of how this can be incorporated into the ETAS model can be found in Kagan and Knopoff (1987) and Ogata (1998).

Summing things up, the ETAS model can be described as a branching process with immigration (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 models the temporal or space–time distribution of the triggered aftershocks.

3. MAXIMUM LIKELIHOOD ESTIMATION OF THE ETAS MODEL

Conventional ML estimation attempts to maximize

$$\ell(\boldsymbol{\theta}) = \sum_i \log\bigl(\lambda(t_i, x_i, y_i|\mathcal{H}_{t_i})\bigr) - \int_0^T \int_{y_0}^{y_1} \int_{x_0}^{x_1} \lambda(t, x, y|\mathcal{H}_t)\, dx\, dy\, dt, \quad (3)$$

the incomplete data log-likelihood function of model (1), 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 dataset (x_i, y_i, t_i, m_i) is observed (Ogata 1998; Daley and Vere-Jones 2003, chap. 7). The term "incomplete data" signifies that $\ell(\boldsymbol{\theta})$ does not incorporate any information about the branching structure, and it is used to distinguish it from the (in practice unattainable) *complete data log-likelihood function* introduced in Section 4. Typically, (3) is maximized by using a numerical optimization routine, because 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. To distinguish the conventional MLE computed by numerical maximization from the one based on the EM-type algorithm presented later, we will denote the former by $\hat{\boldsymbol{\theta}}_{\text{num}}$ and the latter by $\hat{\boldsymbol{\theta}}_{\text{EM}}$. Similarly, the vector components will be denoted by $\hat{\omega}_{\text{num}}$, $\hat{\omega}_{\text{EM}}$, and so on.

An illustration may be helpful to demonstrate some of the difficulties encountered when directly maximizing (3) using numerical methods. Figure 1 shows a simulated earthquake catalog of 638 events, using model (1). The space–time window used for this simulation is similar to the Southern California dataset described in Section 5. Background earthquakes are simulated on an area of 8° of longitude by 5° of latitude over a period of 7,500 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 discussions 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):

$$f_{\text{GR}}(m) = \frac{\beta e^{-\beta(m-M_0)}}{1 - e^{-\beta(M_{\text{GR}}^{\max} - M_0)}}, \quad (4)$$

where $f_{\text{GR}}(m)$ is the probability density function, $\beta = \log(10)$, and $M_0 = 2 \le m \le M_{\text{GR}}^{\max} = 8$, 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 2002a, 2003) and the upper threshold is the approximate magnitude of the strongest California earthquakes in historic times, the 1857 Fort Tejon earthquake and the "great" San Francisco earthquake of 1906.

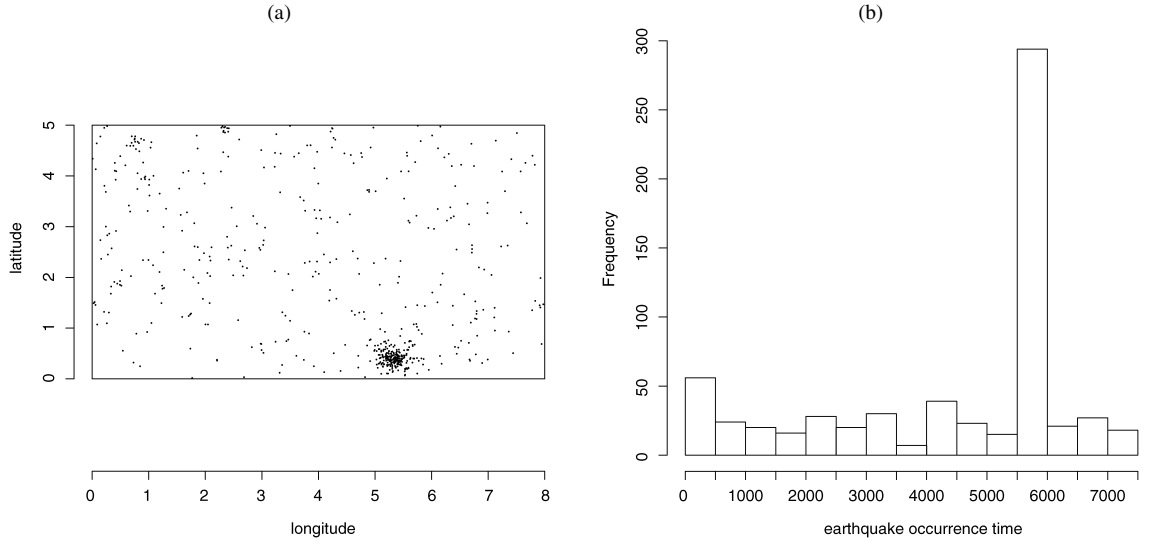

Figure 1. Simulated earthquake process using a space–time–magnitude ETAS model. This figure shows a simulated earthquake catalog using model (1) with a parameterization as described in Table 1. The simulated catalog consists of 638 earthquakes (241 background events and 397 aftershocks). The spatial distribution is presented in (a), although 32 (triggered) earthquakes are not shown because they are outside the specified space–time window. The temporal distribution is shown in (b). The spike in activity starting on day 5,527 is caused by a magnitude 5.33 earthquake and its aftershocks and corresponds to the cluster on the lower right side of (a).

The use of numerical methods to maximize (3) 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 incomplete data log-likelihood for variations of each component of the parameter vector by up to 50% around the MLE. The function is quite flat around $\hat{\boldsymbol{\theta}}_{\text{num}}$, 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 others remain constant at their MLEs. Again, (3) 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.

Note that flatness in the log-likelihood function does not necessarily imply that accurate estimation of the parameters in the ETAS model is unimportant. Some of the ETAS parameters have physical interpretations that can be used by seismologists to characterize earthquake catalogs (see Ogata 1998, and the references therein). Further, the accurate estimation of ETAS parameters is important for comparisons of certain parameters for different catalogs as well as for studies of bias in ETAS parameter estimation.

In cases where the log-likelihood function is extremely flat, the choice of starting values can influence the results. In Figure 4, conventional ML estimation is performed using eight different starting values and a standard Newton–Raphson optimization routine. The true values of Table 1 are used as starting values 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 the true $\boldsymbol{\theta}$. In four 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

Table 1. Specification of the space–time–magnitude ETAS model (1) used for simulation

Parameter	Value	Space–time window for background events
$\mu(x, y)$	.0008	
K_0	.0000305	
a	2.3026	$[0°, 8°] \times [0°, 5°] \times [0, 7{,}500$ days$]$
c	.01	
ω	.5	Parameters of the magnitude distribution
d	.015	M_0 2
ρ	.8	$M_{\text{GR}}^{\max}$ 8

NOTE: This specification uses a homogeneous background intensity (measured in events per day per squared degree) that does not depend on the location. The time is measured in days; spatial distances are measured in degrees. A truncated exponential distribution (4) is used to simulate magnitudes.

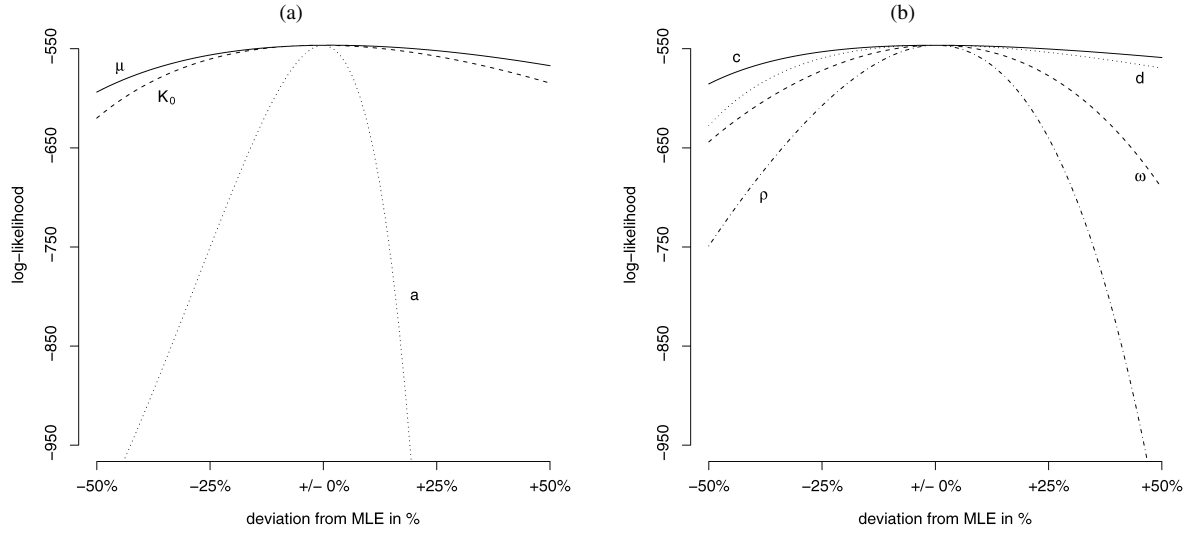

Figure 2. Flatness of the incomplete data log-likelihood function for varying one parameter at a time. This figure demonstrates the relative flatness of the incomplete data log-likelihood function (3) with respect to the components of the parameter estimate $\hat{\boldsymbol{\theta}}_{\text{num}}$ (two 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.

the true values, as the computation of gradients may actually be more accurate in locations farther away from the true parameter value.

Another problem often encountered in practice is that the log-likelihood (3) 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 $\hat{\boldsymbol{\theta}}_{\text{num}}$ can be very far from the true parameter value. This makes it difficult to perform simulation studies of bias and asymptotic properties for which an automatic procedure would be desirable.

While the main focus of this article is to present how branching process models can be estimated using an EM-type algorithm, a few remarks will be added here on how to improve conventional ML estimation via a numerical maximization of the incomplete data log-likelihood. One approach is to maximize (3) with respect to $\log(c)$ and $\log(d)$, in place of c and d. This procedure is computationally more stable because c and d can be quite ill constrained. Also, the simplex algorithm as

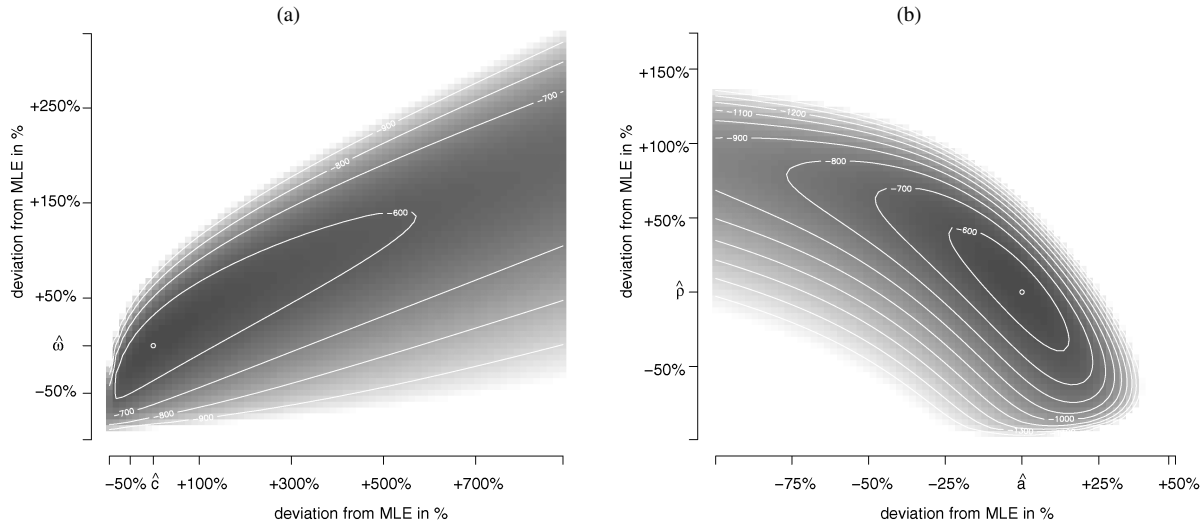

Figure 3. Flatness of the incomplete data log-likelihood in multidimensional settings. The problem of flatness of the incomplete data log-likelihood function (3) (shown in gray levels) can be aggravated in a multidimensional context. In this analysis, pairs of components of $\boldsymbol{\theta}$ [c, ω in (a) and a, ρ in (b)] are varied around the MLE $\hat{\boldsymbol{\theta}}_{\text{num}}$ (small white circle), while all the other components remain fixed. Along certain trajectories, even large deviations reduce the log-likelihood only marginally.

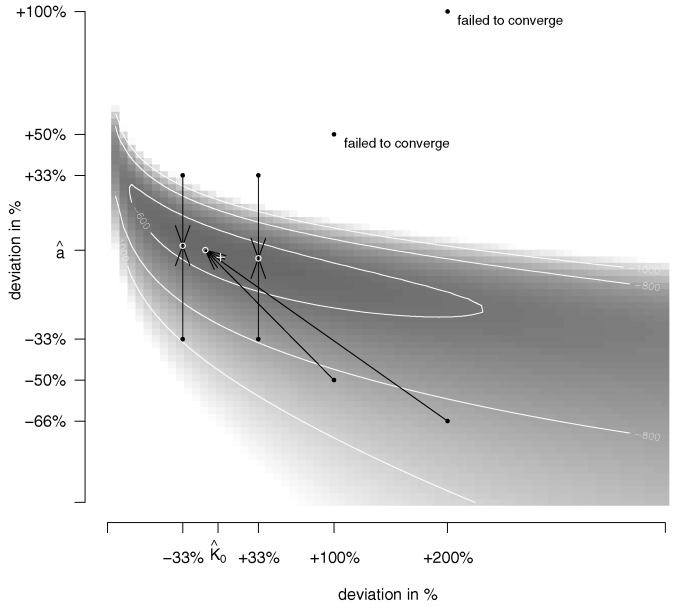

Figure 4. Difficulties in estimating ETAS parameters using conventional ML. Except for K_0 and a, the starting values (black dots) for the components of $\boldsymbol{\theta}$ are set to the true values (Table 1). The white circles show the estimation results of conventional ML ($\hat{\boldsymbol{\theta}}_{\text{num}}$), 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 incomplete data log-likelihood function (shown in gray-levels) with respect to variations of K_0.

well as genetic algorithms have proved to be useful and efficient alternatives to the Newton–Raphson procedure used in our study. Finally, Ogata (1993a) presented a "fast" algorithm applicable to Markovian conditional intensities that replaces the first term of (3) [which is effectively a double sum because $\lambda(t_i, x_i, y_i|\mathcal{H}_{t_i})$ is itself a sum; see (1)] with a single sum. This procedure replaces the computation of the inner sum with a numerical integration procedure.

4. ETAS ESTIMATION USING AN EM–TYPE ALGORITHM

In their seminal article, Dempster et al. (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 an 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. The branching structure of the ETAS model has previously been used as a computationally efficient simulation procedure for earthquake catalogs (Felzer, Becker, Abercrombie, Ekström, and Rice 2002; Zhuang, Ogata, and Vere-Jones 2004). It has also facilitated a probabilistic "declustering" method using a "stochastic reconstruction" of earthquake catalogs (Zhuang et al. 2002, 2004). For an early version of probabilistic declustering, see Kagan and Knopoff (1976). In the works of Zhuang et al., probabilities of earthquakes being background events or otherwise triggered by preceding events are used to improve spatial background intensity estimates. Here, 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 $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_\kappa)$. 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:

$$\nu_k = \mu_k \cdot (\text{area of cell } k) \cdot (\text{length of time window}), \qquad (5)$$

where the length of the time window is T [see (3)]. Depending on the situation, either μ_k or ν_k will be used in the formulas, because 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), that is, if the unobserved quantities u_i were known for all i, the *complete data log-likelihood* $\ell_c(\boldsymbol{\theta})$ would be written as

$$\begin{aligned}\ell_c(\boldsymbol{\theta}) = &\sum_{k=1}^{\kappa}\{-\log(n_k!) - \nu_k + n_k \log(\nu_k)\} \\ &+ \sum_i \big\{-\log(l_i!) - G_i(\boldsymbol{\theta}) + l_i \log(G_i(\boldsymbol{\theta}))\big\} \\ &+ \sum_{i\,:\,u_i \neq 0} \big\{\log(\omega) + \omega\log(c) + \log(\rho) + \rho\log(d) \\ &- (1+\rho)\log\big((x_i - x_{u_i})^2 + (y_i - y_{u_i})^2 + d\big) \\ &- (1+\omega)\log\big(t_i - t_{u_i} + c\big) - \log(\pi)\big\}, \qquad (6)\end{aligned}$$

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 (2), $G_i(\boldsymbol{\theta})$ can be derived as

$$\begin{aligned}G_i(\boldsymbol{\theta}) &= \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty g(t - t_i, x - x_i, y - y_i)\, dx\, dy\, dt \\ &= K_0 \pi \frac{d^{-\rho} c^{-\omega}}{\rho\omega} e^{a(m_i - M_0)}. \qquad (7)\end{aligned}$$

The third sum of (6) 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 $g(t-t_i, x-x_i, y-y_i)$ in (2) by the expected number of aftershocks $G_i(\boldsymbol{\theta})$ in (7).

Complete Data Maximum Likelihood Estimation. Generally, the complete data log-likelihood $\ell_c(\boldsymbol{\theta})$ cannot be maximized in practice, because 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; using them allows us to maximize $\ell_c(\boldsymbol{\theta})$, that is, find the (in practice unattainable) *complete data MLE*, denoted as $\tilde{\boldsymbol{\theta}}$. Partial derivatives of $\ell_c(\boldsymbol{\theta})$ with respect to each component of $\boldsymbol{\theta}$ will be used, which are generally more compact than those of the incomplete data log-likelihood $\ell(\boldsymbol{\theta})$ in (3).

Given the quantities u_i for all i, the parameters for the background intensity can be estimated using $0 \stackrel{!}{=} \partial\ell_c(\boldsymbol{\theta})/\partial\nu_k = -1 + \frac{n_k}{\nu_k}$ with $k = 1, 2, \ldots, \kappa$, where the notation $\stackrel{!}{=}$ indicates that the partial derivatives are set to 0 in order to maximize the respective function, in this case $\ell_c(\boldsymbol{\theta})$. The complete data MLE is $\tilde{\nu}_k = n_k$, and $\tilde{\mu}_k$ can then be computed using (5).

The partial derivatives with respect to the parameters c and ω describing the temporal aftershock distribution are

$$0 \stackrel{!}{=} \frac{\partial\ell_c(\boldsymbol{\theta})}{\partial c} = \sum_{i\,:\,u_i\neq 0}\left(\frac{\omega}{c} - \frac{1+\omega}{t_i - t_{u_i} + c}\right) + \frac{\omega}{c}\sum_i (G_i(\boldsymbol{\theta}) - l_i), \quad (8)$$

$$0 \stackrel{!}{=} \frac{\partial\ell_c(\boldsymbol{\theta})}{\partial\omega} = \sum_{i\,:\,u_i\neq 0}\left(\frac{1}{\omega} + \log(c) - \log\bigl(t_i - t_{u_i} + c\bigr)\right) + \left(\frac{1}{\omega} + \log(c)\right)\sum_i (G_i(\boldsymbol{\theta}) - l_i). \quad (9)$$

Note that (8) and (9) would only depend on c and ω if the terms after the first sums in each of the equations could be ignored. This is indeed possible, because $0 \stackrel{!}{=} \partial\ell_c(\boldsymbol{\theta})/\partial K_0 = -1/K_0\sum_i (G_i(\boldsymbol{\theta}) - l_i)$ will subsequently guarantee that $\sum_i (G_i(\boldsymbol{\theta}) - l_i)$ equals 0. Setting only the first sums of (8) and (9) to 0 can be interpreted as using only the time passed between the triggering event and the aftershock for the estimation of the temporal parameters while ignoring the contribution of the *number* of aftershocks triggered by each earthquake to the log-likelihood. In this sense, a partial information approach is used to estimate c and ω in this step. This implies the following:

$$\frac{\tilde{\omega}}{(1+\tilde{\omega})\tilde{c}} = \frac{1}{L}\sum_{i\,:\,u_i\neq 0}\frac{1}{t_i - t_{u_i} + \tilde{c}}, \quad (10)$$

$$\frac{1}{\tilde{\omega}} + \log(\tilde{c}) = \frac{1}{L}\sum_{i\,:\,u_i\neq 0}\log\bigl(t_i - t_{u_i} + \tilde{c}\bigr), \quad (11)$$

where L is the number of triggered earthquakes. This equation system can be solved by choosing a strictly positive starting value for $\tilde{c}$ and then iterating between (a) computing the right sides of (10) and (11) and (b) updating the current values for $\tilde{c}$ and $\tilde{\omega}$ by solving the equation system using the quantities computed in step (a) for the right sides.

The spatial parameters d and ρ can be estimated analogously, because $\partial\ell_c(\boldsymbol{\theta})/\partial d$ and $\partial\ell_c(\boldsymbol{\theta})/\partial\rho$ are structurally identical to (8) and (9), respectively, with the squared distance between aftershock and triggering event replacing the time elapsed between the two events.

Finally, once the parameters of the space–time distribution of aftershocks are known, the remaining parameters $\boldsymbol{\theta}$, K_0, and a, which govern the number of triggered aftershocks, can be estimated as parameters of a Poisson regression in which the number of triggered earthquakes depends on the magnitude of the triggering event through $G_i(\boldsymbol{\theta})$ in (7). A similar conditional likelihood approach was investigated by Ogata and Akaike (1982) for a temporal self-exciting point process model for earthquake occurrences. Ogata and Akaike presented a two-step procedure, in which they held the equivalent of parameter a constant in order to estimate the other parameters of their model [step (a)]. They then updated the equivalent of a using Akaike's (1974, 1977) information criterion (AIC) [step (b)] and iterated between (a) and (b). Our approach, as outlined in the following section, 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.

In summary, an algorithm for complete data ML would estimate (a) $\boldsymbol{\mu}$, (b) c, ω, d, and ρ using partial information ML, and (c) K_0 and a by Poisson regression conditioning on the estimates obtained in (b).

Maximum Likelihood Estimation Using the EM Algorithm. Because the quantities u_i are unknown (and, in fact, unobservable), complete data ML estimation as outlined previously cannot be implemented 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 $\hat{\boldsymbol{\theta}}_{\text{EM}}^{(n)}$:

$$\text{prob}^{(n+1)}(u_i = j) = \bigl(g\bigl(t_i - t_j, x_i - x_j, y_i - y_j, m_j|\hat{\boldsymbol{\theta}}_{\text{EM}}^{(n)}\bigr)\bigr)\Big/\left(\hat{\mu}_{k\,:\,i\in\text{cell }k}^{(n)} + \sum_{r=1}^{i-1} g\bigl(t_i - t_r, x_i - x_r, y_i - y_r, m_r|\hat{\boldsymbol{\theta}}_{\text{EM}}^{(n)}\bigr)\right). \quad (12)$$

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

$$
\begin{aligned}
&E_{\hat{\boldsymbol{\theta}}_{\mathrm{EM}}^{(n)}}[\ell_c(\boldsymbol{\theta})] \\
&\quad = \sum_k \bigl\{-\log\bigl(\Gamma\bigl(\hat{n}_k^{(n)}+1\bigr)\bigr) - \nu_k + \hat{n}_k^{(n)}\log(\nu_k)\bigr\} \\
&\qquad + \sum_i \bigl\{-\log\bigl(\Gamma\bigl(\hat{l}_i^{(n)}+1\bigr)\bigr) - G_i(\boldsymbol{\theta}) + \hat{l}_i^{(n)}\log(G_i(\boldsymbol{\theta}))\bigr\} \\
&\qquad + \sum_{i\geq 2}\sum_{j=1}^{i-1} \mathrm{prob}^{(n)}(u_i = j) \\
&\qquad \times \bigl\{\log(\omega) + \omega\log(c) + \log(\rho) + \rho\log(d) \\
&\qquad - (1+\rho)\log\bigl((x_i - x_j)^2 + (y_i - y_j)^2 + d\bigr) \\
&\qquad - (1+\omega)\log(t_i - t_j + c) - \log(\pi)\bigr\}, \qquad (13)
\end{aligned}
$$

where the gamma function with $x! = \Gamma(x+1)$ replaces the factorials used in (6) allowing for noninteger arguments.

The M step of the EM algorithm maximizes (13). In principle, this can be done as outlined in the previous section for the (hypothetical) complete data ML procedure. For this purpose, the quantities n_k and l_i have to be replaced with their counterparts in expectation, that is, $\hat{n}_k^{(n)}$ and $\hat{l}_i^{(n)}$, and the expressions for the partial derivatives have to reflect the estimated triggering probabilities (12), because it is unknown which earthquake triggered which aftershock. The equations (10) and (11) used to estimate the temporal parameters c and ω, for instance, take on the following form:

$$
\begin{aligned}
\frac{\hat{\omega}^{(n)}}{(1+\hat{\omega}^{(n)})\hat{c}^{(n)}} &= \frac{1}{\hat{L}^{(n)}}\sum_{s\geq 2}\sum_{r=1}^{s-1}\mathrm{prob}^{(n)}(u_s = r) \\
&\quad \times \frac{1}{t_s - t_r + \hat{c}^{(n)}}, \qquad (14)
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{\hat{\omega}^{(n)}} + \log\bigl(\hat{c}^{(n)}\bigr) &= \frac{1}{\hat{L}^{(n)}}\sum_{s\geq 2}\sum_{r=1}^{s-1}\mathrm{prob}^{(n)}(u_s = r) \\
&\quad \times \log\bigl(t_s - t_r + \hat{c}^{(n)}\bigr), \qquad (15)
\end{aligned}
$$

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

Algorithm 1.

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

Step 1 (E Step). Based on $\hat{\boldsymbol{\theta}}_{\mathrm{EM}}^{(n)}$, estimate the triggering probabilities $\mathrm{prob}^{(n+1)}(u_i = j)$ for all i, j as shown in (12).

Step 2 (M Step). Maximize (13), that is, find $\hat{\boldsymbol{\theta}}_{\mathrm{EM}}^{(n+1)} = \arg\max_{\boldsymbol{\theta}} E_{\hat{\boldsymbol{\theta}}_{\mathrm{EM}}^{(n)}}[\ell_c(\boldsymbol{\theta})]$:

(a) Find $\hat{\boldsymbol{\mu}}_{\mathrm{EM}}^{(n+1)}$ using $\hat{\nu}_{k_{\mathrm{EM}}}^{(n+1)} = \hat{n}_k^{(n+1)}$ and (5).

(b) Find $\hat{c}_{\mathrm{EM}}^{(n+1)}$ and $\hat{\omega}_{\mathrm{EM}}^{(n+1)}$ as outlined in the previous section but using (14) and (15). Compute $\hat{d}_{\mathrm{EM}}^{(n+1)}$ and $\hat{\rho}_{\mathrm{EM}}^{(n+1)}$ analogously.

(c) Fix all components of $\boldsymbol{\theta}$, except for K_0 and a, at their current estimates; that is, let $\breve{\boldsymbol{\theta}}^{(n+1)} = (\hat{\boldsymbol{\mu}}_{\mathrm{EM}}^{(n+1)}, K_0, a, \hat{c}_{\mathrm{EM}}^{(n+1)}, \hat{\omega}_{\mathrm{EM}}^{(n+1)}, \hat{d}_{\mathrm{EM}}^{(n+1)}, \hat{\rho}_{\mathrm{EM}}^{(n+1)})$. Find $\hat{K}_{0_{\mathrm{EM}}}^{(n+1)}$ and $\hat{a}_{\mathrm{EM}}^{(n+1)}$ by solving for K_0 and a numerically:

$$
\begin{aligned}
0 &\stackrel{!}{=} \partial E_{\hat{\boldsymbol{\theta}}_{\mathrm{EM}}^{(n)}}\bigl[\ell_c\bigl(\breve{\boldsymbol{\theta}}^{(n+1)}\bigr)\bigr]/\partial K_0 \\
&= -1/K_0 \sum_i \bigl(G_i\bigl(\breve{\boldsymbol{\theta}}^{(n+1)}\bigr) - \hat{l}_i^{(n+1)}\bigr), \\
0 &\stackrel{!}{=} \partial E_{\hat{\boldsymbol{\theta}}_{\mathrm{EM}}^{(n)}}\bigl[\ell_c\bigl(\breve{\boldsymbol{\theta}}^{(n+1)}\bigr)\bigr]/\partial a \\
&= -\sum_i \bigl(\bigl(G_i\bigl(\breve{\boldsymbol{\theta}}^{(n+1)}\bigr) - \hat{l}_i^{(n+1)}\bigr)(m_i - M_0)\bigr).
\end{aligned}
$$

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

Dempster et al. (1977) showed that, under general conditions, estimates obtained using the EM algorithm are consistent just as conventional (i.e., incomplete data) MLEs are. In a way, the EM estimates are "expected MLEs" because they maximize the expected complete data log-likelihood. Like other optimization routines, 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.

The robustness and accuracy of the algorithm introduced in this work becomes evident when directly comparing it to conventional ML estimation. Using the same starting values as those in 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 synthetic earthquake catalogs are simulated using model (1) and the parameters in Table 1 and then estimated using the proposed EM-type algorithm as well as using conventional ML with our EM-type estimates as starting values. Table 2 shows that the EM-type algorithm generally produces estimates that are less biased (the only exception being the parameter $\boldsymbol{\mu}$). Apart from the bias, the sampling distributions are quite similar for all components of $\boldsymbol{\theta}$; Figure 5 serves as an example and presents the sampling distributions of $\hat{\omega}_{\mathrm{EM}}$ and $\hat{\omega}_{\mathrm{num}}$.

A possible explanation for why the EM-type algorithm yields superior estimates is that most theoretical results relating to ML estimation only hold asymptotically (see, for instance, Ferguson 1996, for a general treatment of this matter, and Ogata 1978, who derived analogous results for point process models). While both $\hat{\boldsymbol{\theta}}_{\mathrm{num}}$ and $\hat{\boldsymbol{\theta}}_{\mathrm{EM}}$ are consistent, in practice, with a sample consisting of a limited number of observations, EM estimates and conventional MLEs may, in fact, be different as $\ell(\hat{\boldsymbol{\theta}}_{\mathrm{num}}) \geq \ell(\hat{\boldsymbol{\theta}}_{\mathrm{EM}})$ and $E_{\hat{\boldsymbol{\theta}}_{\mathrm{EM}}}[\ell_c(\hat{\boldsymbol{\theta}}_{\mathrm{EM}})] \geq E_{\hat{\boldsymbol{\theta}}_{\mathrm{EM}}}[\ell_c(\hat{\boldsymbol{\theta}}_{\mathrm{num}})]$ [see (3) and (13)]. In our simulations, the number of simulated earthquakes may not be large enough for reaching an asymptotic regime in the numerical maximization of the incomplete data log-likelihood function. However, the number of simulated earthquakes might be large enough for the EM procedure to produce accurate results, because the expected complete data log-likelihood incorporates additional structural information. In

Table 2. Bias of the parameter estimates using the EM-type algorithm and conventional ML

	$\hat{\mu}$ ($\times 10^{-4}$)	$\hat{K}_0$ ($\times 10^{-5}$)	$\hat{a}$	$\hat{c}$	$\hat{\omega}$	$\hat{d}$	$\hat{\rho}$
True values	8.000	3.050	2.303	.01000	.500	.01500	.800
Conventional MLEs	8.011	2.993	2.275	.01086	.519	.01625	.841
Standard errors	(.527)	(.708)	(.113)	(.00283)	(.058)	(.00409)	(.103)
Bias in % of true value	+.14%	−1.86%	−1.22%	+8.56%	+3.80%	+8.35%	+5.13%
EM-type estimates	7.925	2.993	2.296	.01019	.501	.01564	.824
Standard errors	(.516)	(.708)	(.109)	(.00265)	(.056)	(.00423)	(.112)
Bias in % of true value	−.94%	−1.85%	−.27%	+1.91%	+.20%	+4.30%	+3.00%

NOTE: The bias of the proposed EM-type algorithm and conventional ML is compared by applying both procedures to 100 simulated processes. For most parameters, the EM-type algorithm yields results that are closer to the true parameter values described in Table 1.

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.

To demonstrate the robustness of the proposed EM-type algorithm with respect to starting values, 10 earthquake catalogs are simulated with the same parameterization as before and then estimated using 100 different starting values for $\boldsymbol{\theta}$. The starting values are sampled from a uniform distribution whose range is one-fifth of the parameter value to five 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 .5% of the true parameter value, the average being less than .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 EARTHQUAKE OCCURRENCES IN SOUTHERN CALIFORNIA

The methodology introduced in the previous section will now be used to estimate the ETAS model (1) using seismological data compiled by the Southern California Earthquake Center (SCEC). The data 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 consist of 6,796 earthquakes occurring in a rectangular area around Los Angeles between longitudes −122° and −114°

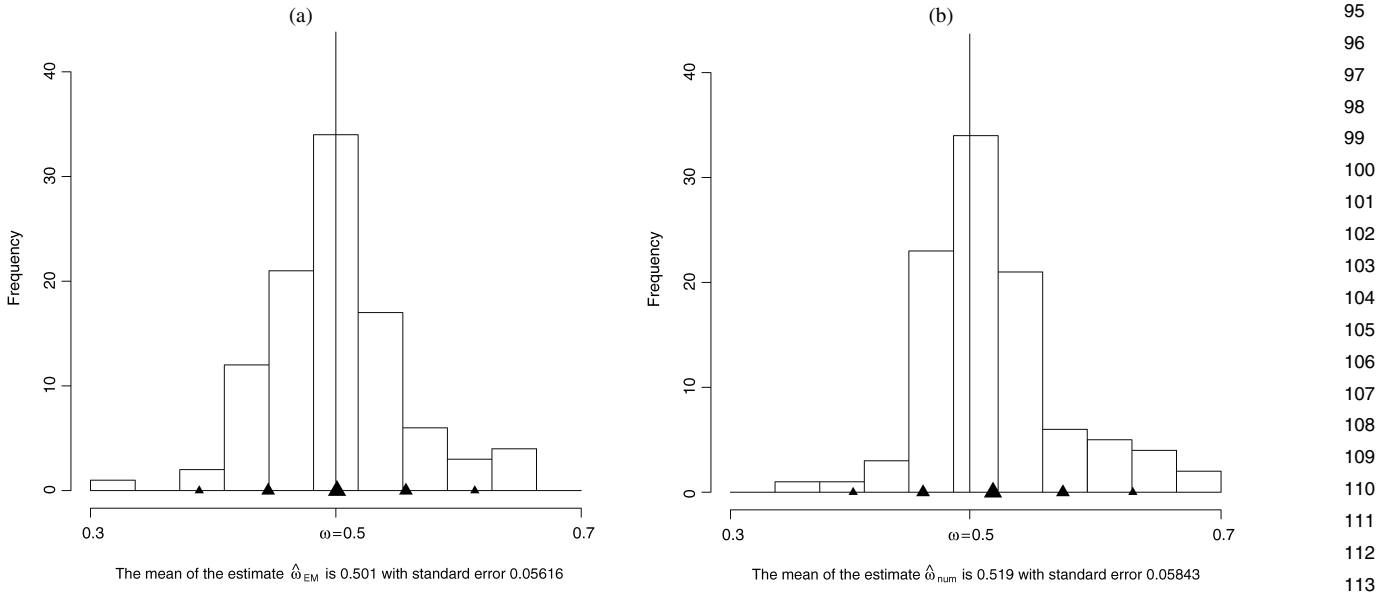

Figure 5. Sampling distributions of parameter estimates. The sampling distributions for the different components of $\boldsymbol{\theta}$ are quite similar for EM-type (a) and conventional ML (b) estimation, though the former is less biased. Shown here are the sampling distributions of $\hat{\omega}_{\mathrm{EM}}$ and $\hat{\omega}_{\mathrm{num}}$. Using the parameter values of Table 1, 100 earthquake catalogs were simulated and the model then estimated using the two procedures. The vertical lines on each histogram indicate the true parameter value. The location of the mean based on the 100 estimates is shown as the large triangle on the bottom of each histogram. One and two standard error intervals are represented by the mid-sized and small triangles, respectively.

and latitudes 32° and 37° (733 km × 556 km) between January 1, 1984, and June 17, 2004, and are considered complete above $M_0 = 3$ (Kagan 2002a, 2003).

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 declustering method introduced by Zhuang et al. (2002, 2004, 2005), who used a spatial kernel density smoother as an estimate of $\mu(x, y)$ in (1). 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 identified 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).

The space–time ETAS model (1) 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, because 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, because it seems to require a substantially larger sample size in order to attain the asymptotic regime guaranteeing consistent estimation.

Table 3. Estimation results of space–time ETAS model (1) for Southern California, 1984–2004, $M_0 = 3$

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}	1,357	3.26	SE; southern section of San Andreas fault system
3	3.921×10^{-3}	476	6.14	NW; creeping section of San Andreas fault system
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}	2,822	3.43	Mojave block, eastern transverse ranges
7	6.130×10^{-3}	637	2.75	Western Great Basin
0	$.019 \times 10^{-3}$	38	12.52	Rest

$\hat{K}_0$	$\hat{a}$	$\hat{c}$	$\hat{\omega}$	$\hat{d}$	$\hat{\rho}$
4.823×10^{-5}	1.034	.01922	.222	4.906×10^{-5}	.497

NOTE: The estimated background intensities 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 Fig. 6).

The estimation results are presented in Table 3 and Figure 6. The declustered background intensities for six of the seven re-

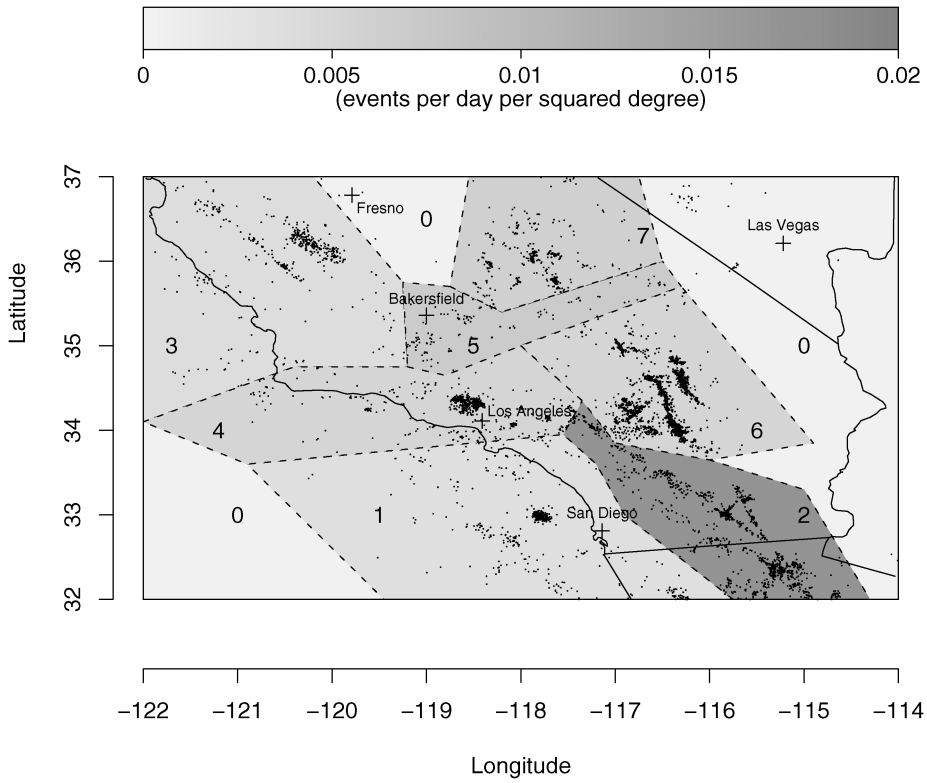

Figure 6. Declustered background intensity estimates for Southern California. Background intensities are estimated for geologically distinct regions in Southern California using the space–time ETAS model (1). The declustered background intensities are quite similar for six of the seven regions including 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).

gions 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, because strong earthquakes with large aftershock clusters were observed in the western transverse ranges (region 4) and the Mojave block (region 6) with 817 and 2,822 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, has a substantially higher background intensity of 16.945×10^{-3} events per day per squared degree.

Some parameter estimates are not in line with geophysical theory. The model parameter $\hat{a} = 1.034$, for instance, is commonly thought to be closer to $\log(10) \approx 2.3$ (Felzer, Abercrombie, and Ekström 2004; Helmstetter, Kagan, and Jackson 2005, 2006). However, several empirical studies estimate this parameter to a value closer to unity (see, for instance, Ogata 1998; Helmstetter et al. 2006). Helmstetter et al. (2006) attributed this downward bias to data errors in earthquake catalogs (incompleteness and measurement errors concerning earthquake locations and magnitudes) as well as model misspecifications. Also, many seismologists consider the parameter estimate $\hat{c} = .01922$ to be too large (or even an artifact) due to catalog incompleteness of small earthquakes after strong events (Kagan 2004; Helmstetter et al. 2005).

The fact that the southern section of the San Andreas fault system is highly active is well established in the literature. For example, Fialko (2006) found a high level of seismic strain accumulation in this region. Given that the last two major California earthquakes in 1857 and 1906 ruptured the middle and northern sections of the San Andreas fault, 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 the references therein). This is supported by paleoseismological 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 concluded 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 directly imply an increased risk for a large-magnitude event. However, an increased incidence of strong earthquakes can be anticipated simply because more earthquakes are expected in general. In this sense an elevated seismic hazard may exist in the southern section of the San Andreas fault system compared to all other regions in Southern California.

6. DISCUSSION AND CONCLUDING REMARKS

We presented an EM-type algorithm that 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, because 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, because 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 incomplete data 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, for example, in order to estimate the bias or variance of certain parameter estimates. Such repetitions are very difficult using conventional ML procedures 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 et al. (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 is substantially higher than in other regions of Southern California. In this sense, an elevated seismic risk exists in this area.

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 aftershocks usually occur on faults with fault orientation varying over different regions. Moreover, as pointed out at the end of Section 2, there is evidence that aftershock regions scale with the magnitude of the main shock. The investigation and application to California seismicity of ETAS models with noncircular and magnitude-scaling aftershock distributions are important directions for further research.

[Received July 2006. Revised April 2007.]

REFERENCES

Akaike, H. (1974), "A New Look at Statistical Model Identification," *IEEE Transactions on Automatic Control*, 19, 716–722.

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

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

Daley, D., and Vere-Jones, D. (2003), *An Introduction to the Theory of Point Processes* (2nd ed.), New York: Springer.

Dempster, A., Laird, N., and Rubin, D. (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, Ser. B, 39, 1–38.

Felzer, K. R., Abercrombie, R. E., and Ekström, G. (2004), "A Common Origin for Aftershocks, Foreshocks, and Multiplets," *Bulletin of the Seismological Society of America*, 94, 88–98.

Felzer, K. R., Becker, T. W., Abercrombie, R. E., Ekström, G., and Rice, J. R. (2002), "Triggering of the 1999 M_W 7.1 Hector Mine Earthquake by Aftershocks of the 1992 M_W 7.3 Landers Earthquake," *Journal of Geophysical Research*, 107, 2190.

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

Fialko, Y. (2006), "Interseismic Strain Accumulation and the Earthquake Potential on the Southern San Andreas Fault System," *Nature*, 441, 968–971.

Gutenberg, B., and Richter, C. F. (1944), "Frequency of Earthquakes in California," *Bulletin of the Seismological Society of America*, 34, 185–188.

Hawkes, A. G. (1971a), "Point Spectra of Some Mutually Exciting Point Processes," *Journal of the Royal Statistical Society*, Ser. B, 33, 438–443.

——— (1971b), "Spectra of Some Self-Exciting and Mutually Exciting Point Processes," *Biometrika*, 58, 83–90.

Hawkes, A. G., and Adamopoulos, L. (1973), "Cluster Models for Earthquakes—Regional Comparisons," *Bulletin of the International Statistical Institute*, 45, 454–461.

Helmstetter, A., Kagan, Y. Y., and Jackson, D. D. (2005), "Importance of Small Earthquakes for Stress Transfers and Earthquake Triggering," *Journal of Geophysical Research*, 110, B05S08.

——— (2006), "Comparison of Short-Term and Long-Term Earthquake Forecast Models for Southern California," *Bulletin of the Seismological Society of America*, 96, 90–106.

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

——— (1994), *Fault Activity Map of California and Adjacent Areas, Scale* 1 : 750,000, Sacramento, CA: California Division of Mines and Geology.

Kagan, Y. Y. (2002a), "Aftershock Zone Scaling," *Bulletin of the Seismological Society of America*, 92, 641–655.

——— (2002b), "Modern California Earthquake Catalogs and Their Comparison," *Seismological Research Letters*, 73, 921–929.

——— (2003), "Accuracy of Modern Global Earthquake Catalogs," *Physics of the Earth and Planetary Interiors*, 135, 173–209.

——— (2004), "Short-Term Properties of Earthquake Catalogs and Models of Earthquake Source," *Bulletin of the Seismological Society of America*, 94, 1207–1228.

Kagan, Y. Y., and Knopoff, L. (1976), "Statistical Search for Nonrandom Features of the Seismicity of Strong Earthquakes," *Physics of the Earth and Planetary Interiors*, 12, 291–318.

——— (1987), "Statistical Short-Term Earthquake Prediction," *Science*, 236, 1563–1567.

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

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

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.

——— (1993a), "Fast Likelihood Computation of Epidemic Type Aftershock Sequence Model," *Geophysical Research Letters*, 20, 2143–2146.

——— (1993b), "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 Akaike, H. (1982), "On Linear Intensity Models for Mixed Doubly Stochastic Poisson and Self-Exciting Point Processes," *Journal of the Royal Statistical Society*, Ser. B, 44, 102–107.

R Development Core Team (2007), *R: A Language and Environment for Statistical Computing*, Vienna: R Foundation for Statistical Computing.

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

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

Zaliapin, I., Keilis-Borok, V., and Axen, G. (2002), "Premonitory Spreading of Seismicity Over the Fault's Network in Southern California: Precursor Accord," *Journal of Geophysical Research*, 107, 2221.

Zhuang, J., Ogata, Y., and Vere-Jones, D. (2002), "Stochastic Declustering of Space–Time Earthquake Occurrences," *Journal of American Statistical Association*, 97, 369–380.

——— (2004), "Analyzing Earthquake Clustering Features by Using Stochastic Reconstruction," *Journal of Geophysical Research*, 109, B05301.

——— (2005), **???**.