Comparison of short-term and long-term earthquake forecast models for southern California

A. Helmstetter, Yan Y. Kagan, David D. Jackson

▶ **To cite this version:**

A. Helmstetter, Yan Y. Kagan, David D. Jackson. Comparison of short-term and long-term earthquake forecast models for southern California. Bulletin of the Seismological Society of America, 2006, 96 (1), pp.90-106. 10.1785/0120050067 . hal-00195371

HAL Id: hal-00195371
https://hal.science/hal-00195371

Submitted on 10 Dec 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PostScript file created: April 17, 2005

Comparison of short-term and long-term earthquake forecast models for southern California

Agnès Helmstetter[1,3], Yan Y. Kagan[2] and David D. Jackson[2]

[1] Institute of Geophysics and Planetary Physics, University of California, Los Angeles
[2] Department of Earth and Space Sciences, University of California, Los Angeles
[3] Now at Lamont-Doherty Earth Observatory, Columbia University, New York

Abstract

We consider the problem of forecasting earthquakes on two different time scales: years, and days. We evaluate some published forecast models on these time scales, and suggest improvements in forecast models on both time scales. For time scales of several years, we modify the smoothed seismicity method of Kagan and Jackson [1994; 2000] by smoothing the locations of magnitude 2 and larger earthquakes. Kagan and Jackson used only magnitude 5 and larger. The new long-term model outperforms the best known published models in a retrospective test on magnitude 5 and larger, primarily because it has much higher spatial resolution. We have also developed a model to estimate daily earthquake probabilities in southern California, using the Epidemic Type Earthquake Sequence model [Kagan and Knopoff, 1987; Ogata, 1988; Kagan and Jackson, 2000]. The forecasted seismicity rate is the sum of a constant background seismicity and of the aftershocks of all past earthquakes. The background rate is estimated by smoothing past seismicity. Each earthquake triggers aftershocks with a rate that increases exponentially with its magnitude and decreases with time following Omori's law. We use an isotropic kernel to model the spatial distribution of aftershocks for small ($m \leq 5.5$) mainshocks, and by smoothing the location of early aftershocks for larger mainshocks. The model also assumes that all earthquake magnitudes follow the Gutenberg-Richter law with a uniform b-value. We use a maximum likelihood method to estimate the model parameters and test the short-term and long-term forecasts. A retrospective test using a daily update of the forecasts between 1985/1/1 and 2004/3/10 shows that the short-term model increases the average probability of an earthquake occurrence by a factor 11.5 compared to the long-term time-independent forecast.

Introduction

The best indicator of earthquake probability is the occurrence of recent earthquakes. Several studies show that many earthquakes are triggered in part by preceding events. Aftershocks are the most obvious examples, but many large earthquakes are preceded by smaller ones. Also, many events are apparently triggered through a cascade process in which triggered quakes trigger others in turn. Typically, the seismicity rate just after and close to a large $m \geq 7$ earthquake can increase by a factor 10^4 and more, and stay above the background level for several decades especially in regions of low tectonic deformation. These large fluctuations of seismicity rate provide a way to improve earthquake forecasting significantly compared to time-independent forecasts. Small earthquakes also have a significant contribution in earthquake triggering because they are much more numerous than larger ones [Helmstetter, 2003; Helmstetter et al., 2005]. As a consequence, many large earthquakes are triggered by previous smaller earthquakes ("foreshocks"), which can be useful to improve the forecasting of large earthquakes [Kagan and Knopoff, 1987; Reasenberg and Jones, 1989; Kagan, 1991; Reasenberg, 1999].

Short-term clustering has been recognized for some time, and several quantitative models of clustering have been proposed [Kagan and Knopoff, 1987; Ogata, 1988; Reasenberg and Jones, 1989; Kagan, 1991; Kagan and Jackson, 2000; Helmstetter and Sornette, 2003a; Ogata, 2004]. Short-term forecasts based on earthquake clustering have already been developed. Kagan and Knopoff [1987] performed retrospective tests using the California seismicity. Jackson and Kagan [1999] and Kagan and Jackson [2000] calculate in real time short-term hazard estimates for the northwest and southwest Pacific regions since 1999 (see http://scec.ess.ucla.edu/∼ykagan/predictions_index.html). More recently, Gerstenberger et al. [2004] developed a method to provide daily forecasts of seismic hazard in southern California. However, these forecasts have not yet been fully optimized and tested.

We have devised and implemented a new method for issuing daily earthquake forecasts for southern California. Our forecasts are updated daily and available on-line (http://moho.ess.ucla.edu/∼helmstet/-forecast.html). Short-term effects may be viewed as temporary perturbations to long-term earthquake potential. This long-term forecast could be a Poissonian process or a time-dependent process including, for example, stress shadows. It can include any geologic information based on fault geometry and slip rate, as well as data from geodesy or paleoseismicity. As a first step, we have measured the long-term seismic activity using historical seismicity only. We show that this simple model performs better than a more sophisticated model which incorporates geology data and characteristic earthquakes [Frankel et al., 1997].

We use the Epidemic Type Earthquake Sequence (ETES) model [Kagan and Knopoff, 1987; Ogata, 1988] to obtain short-term earthquake forecast including foreshocks and aftershocks. This model is usually called "Epidemic Type Aftershock Sequence" (ETAS), but in addition to aftershocks this model also describes background seismicity, mainshocks and foreshocks, using the same laws for all earthquakes. We use a maximum likelihood approach to estimate the parameters by maximizing the forecasting skills of the model [Kagan and Knopoff, 1987; Schorlemmer et al., 2005].

Time-independent forecasts

Definition of the model

We have developed a method to estimate the long-term probability of an earthquake as a function of space and magnitude, from historical seismicity. We estimate the density of seismicity $\mu(\vec{r})$ by declustering and smoothing past seismicity. We use the composite seismicity catalog from the Advanced National Seismic System (ANSS), available at http://-quake.geo.berkeley.edu/anss/catalog-search.html. We selected earthquakes above $m_0 = 3$ for the time period 1932-1979 and above $m_0 = 2$ since 1980. We computed the background density of earthquakes on a grid which covers southern California with a resolution of $0.05° \times 0.05°$. The boundary of the grid is $[32.45°N - 36.65°N]$ in latitude and $[121.55°W - 114.45°W]$ in longitude. Before computing the density of seismicity, we need to decluster the catalog, to remove the largest clusters of seismicity, which would give large peaks of seismicity which do not represent the long-term average. We used Reasenberg's [1985] declustering algorithm with parameters $r_{\rm fact} = 20$, $x_{\rm meff} = 2.00$, $p_1 = 0.99$, $\tau_{\rm min} = 1.0$ day, $\tau_{\rm max} = 10.0$ days, and with a minimum cluster size of 5 events.

The declustered catalog is shown in Figure 1. Note that there is a better method of declustering, which do not need to specify a space-time window to define aftershocks, and uses an ETES-type model to estimate

the probability that each earthquake is an aftershock [Kagan and Knopoff, 1976; Kagan, 1999; Zhuang et al., 2004]. This method is however more complex to use and time-consuming for a large number of earthquakes.

The background density $\mu(\vec{r})$ at point $\vec{r}$ is given by the sum over all past earthquakes $1 \leq i \leq N$ of the kernel $K_d(\vec{r})$ of each earthquake

$$\mu(\vec{r}) = \sum_{i=1} K_{d_i}(\vec{r} - \vec{r_i}) \,. \quad (1)$$

We use an isotropic kernel $K_d(r)$ given by

$$K_d(r) = \frac{C(d)}{(r^2 + d^2)^{1.5}} \,, \quad (2)$$

where $C(d)$ is a normalizing constant, and d is the bandwidth of the kernel.

We use adaptive kernels with a bandwidth d which depends on the density of earthquakes, so that we have a better resolution (smaller d) where the density is higher, but a larger d were the density of seismicity is smaller. Specifically, the bandwidth of the kernel for earthquake i is given by [Izenman, 1991]

$$d_i = \frac{d_0}{\sqrt{\mu(\vec{r_i})}} \quad \text{and} \quad d_i > d_{\min} \,. \quad (3)$$

The background density defined by (1) and (3) has spatial variations at scales smaller than the resolution of the grid (0.05°) used for the forecasts. Therefore, we need to integrate the value of $\mu(\vec{r})$ defined by (1) over each cell to obtain the background rate in this cell. The advantage of the function (2) is that we can compute analytically the integral of $K_d(x, y)$ over one dimension x or y, and then compute numerically the integral in the other dimension.

The function (3) has two adjustable parameters d_0 and the minimum value $d_{\min}$. We estimate these parameters by optimizing the likelihood of the model. We use the data from 1932 to 1995 to compute the density $\mu(r)$ on each cell, and the data since 1996 to evaluate the model. We estimate d_0 and $d_{\min}$ by maximizing the log likelihood of the model given by the sum over all cells

$$LL = \sum_{i_x=1}^{N_x} \sum_{i_y=1}^{N_y} \log p_\mu(i_x, i_y, n) \,, \quad (4)$$

where n is the number of events that occurred in the cell (i_x, i_y), and the probability $p_\mu(i_x, i_y, n)$ of having n events in the cell (i_x, i_y) is given by (assuming a Poisson process)

$$p_\mu(i_x, i_y, n) = [T\,\mu(i_x, i_x)]^n \, \frac{\exp[-\mu\,(i_x, i_y)T\,]}{n!} \,, \quad (5)$$

where $\mu(i_x, i_y)$ represents the density of seismicity in the cell i_x, i_y per unit time and T is the duration of the catalog used to test the model. For each value of the parameters d_0 and $d_{\min}$ we compute iteratively μ and d using expressions (1), (2) and (3), starting with $d = d_{\min}$ for all earthquakes. After each iteration we get a smoother density μ. We stop the iteration when the likelihood LL of the model starts to decrease. The optimization gives $d_0 = 0.0225$ and $d_{\min} = 0.5$ km (minimum allowed value equal to location accuracy). We used 3 iterations to compute iteratively μ and d before the likelihood decreased. The background density, estimated using the above parameters and the declustered catalog up to 2003, is shown in Figure 2.

Comparison with other long-term models

We have compared this model with other long-term forecasts for southern California, the model of Kagan and Jackson [1994] (see also http://moho.ess.ucla.edu/~kagan/s_cal_tbl_new.dat) and the model of Frankel et al. [1997].

Kagan and Jackson [1994; 2000] forecasts . Kagan and Jackson [1994; 2000] (KJ94) used a similar smoothing method to estimate the long-term seismicity rate in southern California (see also Kagan et al. [2003b]). The main differences between their algorithm and the present work are that

- KJ94 use larger $m \geq 5.5$ earthquakes since 1850, without declustering the catalog.
- KJ94 introduce a weight proportional to the logarithm of the moment of each earthquake, while we use the same weight for all earthquakes.
- KJ94 use an anisotropic power-law kernel (maximum density in the direction of the earthquake rupture), with a slower decay with distance $K(r) \approx 1/(r + R_{\min})$. Because the integral of $K(\vec{r})$ over the space does not converge, they truncate the kernel function after a distance $R_{\max} = 200$ km. Their kernel has a larger bandwidth $R_{\min} = 5$ km than the present model (smoother density).

- KJ94 add a constant uniform value to take into account "surprises": earthquakes that occur were no past earthquake occurred.
- KJ94 use extended sources instead of a point source: each $m \geq 6.5$ earthquake is replaced by the sum of smaller scale ruptures (see http://-moho.ess.ucla.edu/∼kagan/cal_fps2.txt).

We modified our model to compare with KJ94 model, in order to use the same grid with the same resolution of 0.1°. Both models were developed by using only data before 1990 to estimate the parameters and the long-term density of seismicity μ. We estimated the parameters d_0 and $d_{\min}$ (defined in (3)) of our long-term model by using the data until 1986/1/1 to compute μ and the data from 1986/1/1 to 1990/1/1 to estimate the likelihood LL of the model. The optimization of the LL gave the parameters values $d_0 = 0.02$ and $d_{\min} = 0.5$ km. We then used those parameters and the data until 1990/1/1 to estimate the long-term density $\mu(\vec{r})$.

We use the log likelihood defined in (4) to compare KJ94 model with our long-term model. Because we want to test only the spatial distribution of earthquakes, not the predicted total number, we normalized both models by the observed number of earthquakes ($N = 56$). We obtain $LL = -463$ for KJ94 model and $LL = -389$ for our model. Both models are shown in Figure 3 together with the observed $m \geq 5$ earthquakes since 1990. The present work thus improves the prediction of KJ94 by a factor (ratio of probabilities) $\exp((-389+463)/56) = 3.7$ per earthquake, despite being much simpler (isotropic and point-source model). This suggests that including small earthquakes ($m \geq 2$) to predict larger ones ($m \geq 5$) considerably improves the predictions, because large earthquakes occur generally at the same place as smaller ones [Kafka and Levin, 2000].

Frankel et al. [1997] forecasts. Frankel et al. [1997] (F97) model is a more complex model which includes both a smoothed historical seismicity (using $m \geq 4$ earthquakes since 1933 and $m \geq 6$ since 1850) and characteristic earthquakes on known faults, with a seismicity rate constrained by the geologic slip rate, and with a rupture length controlled by the fault length. The magnitude distribution follows the Gutenberg-Richter (GR) law with $b = 0.9$ for small magnitudes ($m \leq 6$) and a bump for $m > 6.2$ due to characteristic events. We adjusted our model to use only data before 1996 to build the model and the same grid as F97 with a resolution of 0.1°. We assumed a GR distribution with $b = 1$ and with an upper magnitude cut-off at $m = 8$. We used the ANSS catalog for the time period 1932-1995 to estimate the long-term rate of $m \geq 4$ earthquakes (without declustering the catalog). We then assumed a GR law with $b = 1$ to estimate the long-term rate of $m \geq 5$ earthquakes from the number of $m \geq 4$ events. We use $m \geq 5$ earthquakes in the ANSS catalog that occurred since 1996 to compare the models. Both models are illustrated on Figure 4, which represents the density of seismicity above $m = 5$.

We test how each model explains the number of observed events, as well as their location and magnitude, by comparing the likelihood of each model. The log likelihood is defined by

$$LL = \sum_{i_x=1}^{N_x} \sum_{i_y=1}^{N_y} \sum_{i_m=1}^{N_m} \log p_\mu(i_x, i_y, i_m, n)\,, \qquad (6)$$

where n is the number of events that occurred in the cell (i_x, i_y) and in the magnitude bin i_m (the magnitude range $[5.0 - 8.0]$ is divided in bins of 0.1 unit). The log likelihood is $LL = -155$ for our model and $LL = -161$ for F97 model. We thus obtain a probability gain of 1.5 per earthquake for our model compared to F97. Our model thus better predicts the observed earthquake occurrence since 1996 than F97 model. F97 however better predicts the observed number than our model, because the number of $m \geq 5$ earthquakes in the time period 1996-2004 was smaller than the long-term (1932-1995) rate (predicted number $N = 14.6$ for F97 and $N = 26.6$ for our model, compared to the observed number $N = 15$). The difference in likelihood between the two models in mainly due to the choice of the kernel and of the minimum magnitude used to estimate the rate of seismicity from historical seismicity. F97 use a smoother kernel, with a fixed characteristic smoothing distance of 10 km and with a $\sim 1/r$ decay, and use only $m \geq 4$ earthquakes for smoothing.

Focal mechanisms

In addition to the location and magnitude of earthquakes, seismic hazard estimation also requires the specification of earthquake focal mechanisms. Kagan and Jackson [1994] have developed a method to estimate focal mechanisms based on catalogs of focal mechanisms (see http://moho.ess.ucla.edu/∼kagan/-s_cal_tbl_new.dat for the forecasts for southern California). They used the same catalog ($m \geq 5$) and the same kernel to estimate the rate of seismicity and the

average focal mechanisms in each cell (see section). We would like to redo this analysis by applying the same smoothing kernel than for estimating the rate, including smaller $m \geq 2$ earthquakes, and optimizing the forecast using a maximum likelihood method. This analysis will be presented in another paper.

Time dependent forecasts

Definition of the ETES model

The ETES model is based on two empirical laws of seismicity, which can also be reproduced by a multitude of physical mechanisms, the Gutenberg-Richter law to model the magnitude distribution, and Omori's law to characterize the decay of triggered seismicity with the time since the mainshock [Kagan and Knopoff, 1987; Ogata, 1988; Kagan, 1991; Kagan and Jackson, 2000; Helmstetter and Sornette, 2003a; Rhoades and Evison, 2004]. This model assumes that all earthquakes may be simultaneously mainshocks, aftershocks and possibly foreshocks. Each earthquake triggers direct aftershocks with a rate increasing exponentially $\sim 10^{\alpha m}$ with the earthquake magnitude m, and that decays with time according to Omori's law $\sim 1/(t+c)^p$. We also assume that all earthquakes have the same magnitude distribution, which is independent of the past seismicity. Each earthquake has thus a finite probability of triggering a larger earthquake. An observed "aftershock" sequence in the ETES model is the sum of a cascade of events in which each event can trigger more events.

The global seismicity rate $\lambda(t, \vec{r}, m)$ is the sum of a background rate $\mu(\vec{r})$, usually taken as a spatially non-homogeneous Poisson process, and the sum of aftershocks of all past earthquakes

$$\lambda(t, \vec{r}, m) = \mu(\vec{r})\, P_m(m) + P_m(m) \sum_{t_i < t} \phi_{m_i}(\vec{r} - \vec{r}_i, t - t_i)\,, \tag{7}$$

where $P_m(m)$ is a time-independent magnitude distribution. The function $\phi_m(\vec{r}, t)$ gives the spatio-temporal distribution of triggered events at point $\vec{r}$ and at time t after an earthquake of magnitude m

$$\phi_m(\vec{r}, t) = \rho(m)\, \psi(t)\, f(\vec{r}, m)\,, \tag{8}$$

where $\rho(m)$ is the average number of earthquakes triggered directly by an earthquake of magnitude $m > m_0$

$$\rho(m) = K\, 10^{\alpha(m - m_0)}\,, \tag{9}$$

the function $\psi(t)$ is Omori's law normalized to 1 ($\int_0^\infty \psi(t) dt = 1$)

$$\psi(t) = \frac{(p-1)\, c^{p-1}}{(t+c)^p}\,, \tag{10}$$

and $f(\vec{r}, m)$ is the normalized distribution of horizontal distances $\vec{r}$ relative to the mainshock of magnitude m. We have tested different choices for $f(\vec{r}, m)$, which are described in section .

The exponent α has been found equal or close to 1.0 for the southern California seismicity [Felzer et al., 2004; Helmstetter et al., 2005], equal to the GR b-value, showing that small earthquakes are collectively as important as larger ones for seismicity triggering. Note that we consider in the sum in (7) only earthquakes above the detection magnitude m_0. Smaller undetected earthquakes may also have an important contribution to the rate of triggered seismicity. These undetected earthquakes may thus bias the parameters of the model, i.e., the parameters estimated by optimizing the likelihood of the models are "effective parameters", which more or less account for the influence of undetected small earthquakes with $m < m_0$.

The ETES model assumes that each primary aftershock may trigger its own aftershocks (secondary events). The secondary aftershocks may themselves trigger tertiary aftershocks and so on, creating a cascade process. The exponent p, which describes the time distribution of direct aftershocks, is larger than the observed Omori exponent, which characterizes the whole cascade of direct and secondary aftershocks [Helmstetter and Sornette, 2002].

As a first step, we use a simple GR magnitude distribution

$$P_m(m) \sim 10^{-b(m - m_0)}\,, \tag{11}$$

with a uniform b-value equal to 1.0, a upper cut-off at $m_{\max} = 8$ [Bird and Kagan, 2004], and a minimum magnitude $m_0 = 2$. If we want to predict relatively small $m \leq 4$ earthquakes, we must take into account the fact that small earthquakes are missing in the catalog after a large mainshock. The procedure described for correcting for undetected small earthquakes is described in section .

The background rate μ in (7) is estimated by smoothing the historical seismicity, using the long-term forecast described by equations (1) and (2) in section . Alternatively, our short-term forecast model can be included in a more complex interdisciplinary model, which can include information from geodesy, including GPS, paleo-seismicity and fault slip rates.

We can build a composite model by adding the short-term component to any long-term forecast. For this, we simply need to replace the background rate μ in (7) by the long-term forecast, multiplied by the fraction of time-independent events so that the total predicted seismicity rate (time-independent and triggered events) will be equal to the observed seismicity rate.

We use the QDDS (ftp://clover.wr.usgs.gov/pub/-QDDS/QDDS.html) and the QDM (ftp://clover.wr.usgs.gov/-pub/QDM/QDM.html) java applications to obtain the data (time, locations and magnitude) in real time from several regional networks (southern and northern California, Nevada) and to create a composite catalog. We automatically update our forecast each day and we put our results on our web site (http://-moho.ess.ucla.edu/∼helmstet/forecast.html). The model parameters are estimated by optimizing the prediction (maximizing the likelihood of the model) using retrospective tests. The inversion method and the results are presented in section .

The basic idea of our work is the same as the STEP model of Gerstenberger et al. [2004]. Both models are based on spatio-temporal clustering of earthquakes, but there are several significant differences between the two approaches:

- STEP model uses Frankel et al. [1997] forecasts for the background rate, while we estimate the background rate by declustering and smoothing the seismicity;
- In STEP model, the seismicity rate on each cell is the sum of the background rate plus the rate of aftershocks of only one mainshock, the previous earthquake which has the largest influence on this cell. If the rate of aftershocks of the mainshock is smaller than the background rate, then the seismicity rate is equal to the background rate. In our model, the seismicity rate on each cell is given by the sum of the background rate and of the aftershock rate of all past earthquakes.
- STEP model assumes than the rate of aftershocks scales as $\sim 10^{\alpha m}$ with $\alpha = b$, while we use $\alpha < b$;
- STEP model (in their more complex version) adjusts the parameters p, K (aftershocks productivity) and c of Omori's law for each aftershock sequence and for each cell (if the number of aftershocks is sufficient), using a fit by a maximum likelihood method. If the number of events is too small to adjust the parameters, the generic aftershock model of Reasenberg and Jones [1989] is used. In our model, we use the same parameters p, K and c for all earthquakes, which are estimated by optimizing the model using the whole catalog.
- STEP model introduces a variability of the GR b-value in time and in space, while we use a constant value $b = 1.0$. Bird and Kagan [2004] argue that the b-value has a uniform value $0.9 - 1.0$ for global $M \geq 5.6$ earthquakes in various tectonic provinces.

We will compare our model with STEP and other time-dependent models as part of the Regional Earthquake Likelihood Model (RELM) project to test in real time daily and long-term models for California [Kagan et al. 2003a; Jackson et al. 2004; Schorlemmer et al. 2005].

Application of ETES model for time-dependent forecasts

By definition, the ETES model provides the average instantaneous seismicity rate $\lambda(t)$ at time t given by (7), if we know all earthquakes that occurred up to time t. To forecast the seismicity between the present time t_p and a future time $t_p + T$, we cannot use directly expression (7), because a significant fraction of earthquakes that will occur between time t_p and time $t_p + T$ will be triggered by earthquakes that will occur between time t_p and time $t_p + T$ (see Figure 5). Therefore using expression (7) to provide short-term seismicity forecasts, with a time window T of 1 day, may significantly underestimate the number of earthquakes. Helmstetter and Sornette [2003a] have tested several methods of prediction on synthetic catalogs generated by the ETES model. They have shown that the use of (7) to predict the number of earthquakes in the next day estimated by $N_p = \int_{t_p}^{t_p+T} \lambda(t)\, dt$, using the exact parameters of the ETES model to compute $\lambda(t)$, may underestimate the number of earthquakes by a factor 2, because it does not take into account the contribution of yet unobserved seismicity in the seismicity rate $\lambda(t)$.

To solve this problem, Helmstetter and Sornette [2003a] proposed to generate synthetic catalogs with the ETES model in order to predict the seismicity for the next day, by averaging the number of earthquakes over all scenarios. This methods provides a much better estimation of the number of earthquakes than the

direct use of (7), but is much more complex to apply. It is practically not useful if we want to predict not only the number of earthquakes, but also their location and magnitude, because this would require to generate a huge number of scenarios. Helmstetter and Sornette [2003a] have shown that, for synthetic ETES catalogs, the use of $N_p = \int_{t_p}^{t_p+T} \lambda(t)\,dt$ to predict the number of earthquakes between t_p and t_p+T underestimates the number of actually occurred earthquakes by an approximate constant factor, independent of the future events number. This means that the effect of yet unobserved seismicity is to amplify the aftershock rate of past earthquakes by a constant factor.

This result suggests a simple solution to take into account the effect of yet unobserved earthquakes. We can use $N_p = \int_{t_p}^{t_p+T} \lambda(t)\,dt$ to predict the number of earthquakes between t_p and t_p+T but using effective parameters K, μ_s and α, which may be different from the true (i.e., $t_p = t$) parameters of the ETES model. Instead of using the likelihood of the ETES model to estimate these parameters, as done by Kagan [1991], we will estimate the parameters of the model by optimizing the likelihood of the forecasts, defined in section . These effective parameters will depend on the duration (horizon) T of the forecasts.

Threshold Magnitude

An important problem when modeling the occurrence of relatively small earthquakes $m \leq 4$ in California is that the catalog is not complete after large earthquakes [Kagan, 2004]. One effect of missing earthquakes is that the model will over-estimate the observed number of earthquakes, because small earthquakes are not detected. But another effect of missing early aftershocks is to underestimate the predicted seismicity rate, because we miss the contribution from these undetected small earthquakes in the future seismicity rate estimated from the ETES model (7). Indeed, secondary aftershocks (triggered by a previous aftershock) represent an important fraction of aftershocks [Felzer et al., 2003; Helmstetter and Sornette, 2003b].

We have developed a method to correct from both effects of undetected small aftershocks. We first need to estimate the threshold magnitude as a function of the time from the mainshock and of the mainshock magnitude. We analyzed all aftershock sequences of $m \geq 6$ earthquakes in southern California since 1985. We propose the following relation between the threshold magnitude $m_c(t,m)$ at time t (in days) after a mainshock of magnitude m

$$\begin{aligned} m_c(t,m) &= m - 4.5 - 0.76\log_{10}(t) \\ &\text{and} \quad m_c(t,m) \geq m_0 = 2\,. \end{aligned} \tag{12}$$

Of course, there are some fluctuations between one sequence and another one, but the above relation is correct within ≈ 0.2 magnitude units. The above relation is illustrated on Figure 6 for Joshua-Tree $m = 6.1$, San-Simeon $m = 6.5$ and Landers $m = 7.3$ aftershock sequences.

We use expression (12) to estimate the detection magnitude magnitude $m_c(t)$ at the time t of each earthquake. The time-dependent detection threshold $m_c(t)$ is larger than the usual threshold m_0 for earthquakes that occurred at short times after a large $m \geq 5$ earthquake. We select only earthquakes with $m_i > m_c(t_i)$ to estimate the likelihood of the forecasts (18).

We can also correct the forecasts for the second effect: missing contribution from undetected aftershocks in the sum (7). We can take into account the effect of earthquakes below the detection threshold $m_c(t)$ and above the minimum magnitude m_0 by adding a contribution to the number $\rho(m)$ of aftershocks of detected earthquakes $m > m_c(t)$, i.e., by replacing $\rho(m)$ in (9) by

$$\rho^*(m) = \rho(m) + \frac{Kb}{b-\alpha}\, 10^{b(m_c(t)-m_0)} \left[1 - 10^{-(b-\alpha)(m_c(t)-m_0)}\right]\,. \tag{13}$$

where $m_c(t)$ is the detection threshold at the time t of the earthquake, estimated by (12), due to the effect of all previous $m \geq 5$ earthquakes. The second contribution corresponds to the effect of all earthquakes with $m_0 < m < m_c(t)$ that occur on average for each detected earthquake. Practically, for a reasonable value of $\alpha \approx 0.8$, this correction (13) is of the same order as the contribution from observed earthquakes, because a large fraction of aftershocks are secondary aftershocks [Felzer et al., 2003], and because small earthquakes are collectively as important as larger ones for earthquake triggering if $\alpha = b$.

Spatial distribution of aftershocks

We have tested different choices for the spatial kernel $f(\vec{r}, m)$, which models the distribution of distances between a mainshock of magnitude m and its aftershocks. We used a power-law function

$$f_{pl}(\vec{r}, m) = \frac{C_{pl}}{[\,|\vec{r}|^2 + d(m)^2\,]^{1.5}}\,, \tag{14}$$

and a Gaussian distribution

$$f_{gs}(\vec{r},m) = C_{gs}\ \exp\left[-\frac{|\vec{r}|^2}{2\,d(m)^2}\right], \qquad (15)$$

where C_{pl} and C_{gs} are normalizing constants such that the integral of $f(\vec{r},m)$ over an infinite surface is equal to 1. The spatial regularization distance $d(m)$ accounts for the finite rupture size and for the location errors. We assume that $d(m)$ is given by

$$d(m) = 0.5 + f_d \times 0.01 \times 10^{0.5m}\ \text{km}, \qquad (16)$$

where the first term accounts for location accuracy and the second term represents the aftershock zone length of an earthquake of magnitude m. The parameter f_d is adjusted by optimizing the prediction and should be close to 1 if the aftershock zone size is equal to the rupture length as estimated by Wells and Coppersmith [1994].

The Gaussian kernel (15), which describes the density of earthquakes at point $\vec{r}$, is equivalent to the Rayleigh distribution $\sim \exp[-(r/d)^2/2]/r$ of distances $|\vec{r}|$ used by Kagan and Jackson [2000]. The choice of an exponent 1.5 in (14) is motivated by recent studies [Ogata, 2004; Console et al., 2003; Zhuang et al., 2004] who inverted this parameter in earthquake catalogs by maximizing the likelihood of the ETES model, and who all found an exponent close to 1.5. This choice is also convenient because the function (14) is integrable analytically. It predicts that the aftershock density decreases with the distance r from the mainshock as $1/r^3$ in the far field, proportionally to the static stress change.

For large earthquakes, which have a rupture length larger than the grid resolution of 0.05° (≈ 5 km) and a large number of aftershocks, we can improve the model by using a more complex anisotropic kernel, as done previously by Wiemer and Katsumata [1999], Wiemer [2000] and Gerstenberger et al. [2004]. We use the location of early aftershocks as a witness for estimating the mainshock fault plane, and the other active faults in the vicinity of the mainshock. We compute the distribution of aftershocks of large $m \geq 5.5$ mainshocks by smoothing the location of early aftershocks

$$f(\vec{r},m) = \sum_{i=1}^{N} f(|\vec{r}-\vec{r_i}|, m_i)\,, \qquad (17)$$

where the sum is on the mainshock and on all earthquakes that occurred within a distance $D_{\text{aft}}(m)$ from the mainshock, and at a time smaller than the present time t_p and not larger than T_{aft} from the mainshock. We took $D_{\text{aft}}(m) = 0.02 \times 10^{0.5m}$ km (approximately 2 rupture lengths) and $T_{\text{aft}} = 2$ days.

The kernel $f(r,m)$ in (17) used to smooth the location of early aftershocks, is either a power-law (14) or a Gaussian distribution (15), with an aftershock zone length given by (16) for the mainshock, but fixed to $d = 2$ km for the aftershocks. This way we recover the same expression for $f(\vec{r},m)$ for a large $m \geq 5.5$ earthquake as for a smaller earthquake if we compute this term just after the mainshock has occurred ($N = 1$ in (17)). At larger times after the mainshock, when a sufficient number of aftershocks have already occurred, $f(\vec{r},m)$ is given by the smoothed density of aftershocks. The density of aftershocks estimated using (17) is shown in Figure 7 for the Landers earthquake, using a power-law kernel (a) or a Gaussian kernel (b). The distribution of aftershocks that occurred after more than two hours after Landers (black dots) is in good agreement with the prediction based on aftershocks that occurred in the first two hours (white circles). In particular, the largest Big-Bear aftershock ($m = 6.4$, latitude=34.2°, longitude =−116.8°), which occurred about three hours after Landers, was preceded by other earthquakes in the first two hours after Landers, and is well predicted by our method. The Gaussian kernel (15) produces a density of aftershocks which is much more localized than with a power-law kernel.

The advantage of using the observed aftershocks to predict the spatial distribution of future aftershocks is that this method is completely automatic, fast, and uses only information from the time and location of aftershocks which are available a few minutes after the earthquake. It can provide an accurate prediction of the spatial distribution of future aftershocks after less than one hour after the mainshock when enough aftershocks have occurred. Our method has also the advantage of taking into account the geometry of the active fault network close to the mainshock, which is reflected by the spatial distribution of aftershocks.

Therefore, even if the spatial distribution of aftershock is controlled by the Coulomb stress change, it may be more accurate, much simpler and faster to use the method described above rather than to compute the Coulomb stress change. Indeed, Coulomb stress change calculation requires to know the mainshock fault plane and the slip distribution, which are available only several hours or days after a large earthquake [Scotti et al., 2003; Steacy et al., 2004]. Felzer et al. [2003] have already shown that a simple fore-

casting model (simplified ETES model), based on the time, location, and magnitudes of all previous aftershocks better predicts the location of future aftershocks than Coulomb stress change calculations.

Definition of the likelihood and estimation of the model parameters

We use a maximum likelihood method to test the forecasts and to estimate the parameters. For the temporal distribution of aftershocks, we choose to fix the c-value in Omori's law (10) equal to 0.0035 day (5 minutes). This parameter is not important as long as it is much smaller than the time window $T = 1$ day of the forecast. We estimate the parameter p of Omori's law, the exponent α and the constant K of the aftershock productivity $\rho(m)$ defined by (9). We write the background rate $\mu(\vec{r})$ as the product $\mu_s\,\mu_0(\vec{r})$, where $\mu_0(\vec{r})$ is the background density normalized to 1 (see section) and μ_s gives the total number of background events per day. We have thus 5 parameters to estimate: p (Omori exponent defined in (10)), K and α (see eq. (9)), μ_s, and f_d (parameter defined by (16), which describes the size of the aftershock zone).

The log likelihood (LL) of the forecasts is defined by Kagan and Jackson [2000], Kagan et al. [2003b], Schorlemmer et al. [2005]

$$LL = \sum_{i_t=1}^{N_t} \sum_{i_x=1}^{N_x} \sum_{i_y=1}^{N_y} \sum_{i_m=1}^{N_m} \log p_{N_p}(i_t, i_x, i_y, i_m, n)\,, \tag{18}$$

where n is the number of events that occurred in the bin (i_t, i_x, i_y, i_m). We take a step of $T = 1$ day in time, 0.05 degree in space and 0.1 in magnitude. The forecasts are updated each day at midnight Los Angeles time. We assume a Poisson process to estimate the probability $p_{N_p}(i_t, i_x, i_y, i_m, n)$ of having exactly n events in each bin from the predicted number $N_p(i_t, i_x, i_y, i_m)$ of events in this bin:

$$p_\lambda(i_t, i_x, i_y, i_m, n) = (N_p(i_t, i_x, i_y, i_m))^n\, \frac{\exp[-N_p(i_t, i_x, i_y, i_m)]}{n!} \tag{19}$$

where $N_p p(i_t, i_x, i_y, i_m)$ is given by the integral over each space-time-magnitude bin of the predicted seismicity rate $\lambda(\vec{r}, t, m)$

$$N_p(i_t, i_x, i_y, i_m) = \int_{i_m}^{i_m+1} \int_{i_t}^{i_t+1} \int_{i_x}^{i_x+1} \int_{i_y}^{i_y+1} \lambda(\vec{r}, t, m)\, dm\, dt\, dx\, dy\,. \tag{20}$$

We can simplify the expression of LL, by noting that we need to compute the seismicity rate only in the bins (i_x, i_y, i_m) that have a non-zero number of observed events n. We can rewrite (18) and (19) as

$$LL = \sum_{i_t=1}^{N_t} \Big\{ - N_p(i_t) + \sum_{i_x}^{n>0} \sum_{i_y}^{n>0} \sum_{i_m}^{n>0} n \log[N_p(i_t, i_x, i_y, i_m)] - \log(n!) \Big\}\,, \tag{21}$$

where $N_p(i_t)$ is the total predicted number of events between t_p and $t_p + T$

$$N_p(i_t) = \mu_s + \int_{i_t}^{i_t+1} \sum_{t_i<t_p} f_i\, \rho(m_i) \psi(t_p - t_i)\, dt\,. \tag{22}$$

The factor f_i in (22) is the integral of the spatial kernel $f_i(\vec{r} - \vec{r}_i)$ over the grid, which is smaller than 1 due to the finite size of the grid.

We maximize the log likelihood LL defined by (18) using a simplex algorithm [Press et al., 1992, p. 402], and using all earthquakes with $m \geq 2$ since 1985/1/1 and until 2004/3/10 to test the forecasts. We take into account in the seismicity rate (7) the aftershocks of all earthquakes with $m \geq 2$ since 1980/1/1 that occurred within the grid ($[32.45°N - 36.65°N]$ in latitude and $[121.55°W - 114.45°W]$ in longitude) or at less than 1° outside the grid. There are 65,664 target earthquakes above the threshold magnitude m_c in the time and space window used to compute the LL. We test different models for the spatial distribution of aftershocks, a power-law kernel (14) or a Gaussian (15).

We use the probability gain per earthquake G to quantify the performance of the short-term prediction by comparison to the long-term forecasts

$$G = \exp\left(\frac{LL - LL_{LT}}{N}\right)\,, \tag{23}$$

where LL_{LT} is the log-likelihood of the long-term model, LL is the likelihood of the ETES model and N is the total number of events. The gain defined by (23) is related to the information per earthquake I defined by Kagan and Knopoff [1977] (see also Daley and Vere-Jones [2004]) by $G = 2^I$.

A certain caution is needed in interpreting the probability gain for the ETES model. Earthquake temporal occurrence is controlled by Omori's law which diverges to infinity for time approaching zero. Calculating the likelihood function for aftershock sequences illustrates this point: the rate of aftershock

occurrence after a strong earthquake increases by a factor of thousands. Since $\log(1000) = 6.9$, one early aftershock yields a contribution to the likelihood function analogous to about 7 additional free parameters. This means that the likelihood optimization procedure as well as the probability gain value is strongly dependent on early aftershocks. As Figure 6 demonstrates, many early aftershocks are missing from earthquake catalogs [Kagan, 2004], therefore the likelihood function value substantially depends on poor quality data in the beginning of aftershock sequence.

Similarly, earthquake hypocenters are concentrated on a fractal set with a correlation dimension slightly above 2.0 [Helmstetter et al., 2005]. Due to random location errors for small inter-earthquake distances the dimension increases close to 3.0. This signifies that the likelihood function would substantially depend on location uncertainty, since kernel width $K_d(\vec{r})$ (Equations 1 and 2) can be made smaller if a catalog with higher location accuracy is used.

Results and discussion

Model parameters and likelihood

The model parameters are obtained by maximizing the LL. The optimization converges after about 100 iterations. We have tried different values of the initial parameters and we have checked that the final values do not depend on the initial values. The results are given in Table 1 and in Figure 8. We have tested different versions of the model (spatial kernel, unconstrained or fixed α value, and different values of the minimum magnitude).

An example of our daily forecasts (using model 3 in Table 1) is shown in Figure 9, for the day of 2004/10/23. All 6 earthquakes which occurred during that day are located in areas of high predicted seismicity rate (large values of N_p). All except one occurred close enough in time and space from a recent earthquake, so that the short-term predicted number $N_p(\vec{r})$ is larger than the long-term average rate $\mu_{LT}(\vec{r})$. The probability gain (23) per earthquake for this day is 26.

Figure 8 shows the LL of the daily forecasts, for the time period from 1985/1/1 to 2004/3/10, and for each iteration of the optimization as a function of the model parameters. The variation of the LL with each model parameter gives an idea of the resolution of this parameter. The unconstrained inversion gives a probability gain $G = 11.7$, and an exponent $\alpha = 0.43$, much smaller than the direct estimation $\alpha = 0.8 \pm 0.1$ [Helmstetter, 2003] or $\alpha = 1$ [Felzer et al., 2004; Helmstetter et al., 2005] using a fit of the number of aftershocks as a function of the mainshock magnitude. The optimization with α fixed to 0.8, closer to the observed value, provides a probability gain $G = 11.1$ slightly smaller than the best model. Note that there is a negative correlation between the parameters K and α in Table 1: K is larger for a smaller α in order to keep the number of forecasted earthquakes constant.

Comparison of predicted and observed aftershock rate

Figure 10 compares the predicted number of events following Landers mainshock, for the unconstrained model 2 (see Table 1), and for models 3 and 5 with α fixed to 0.8. Model 3 underestimates the number of aftershocks, but predicts the correct variation of the seismicity rate with time. In contrast, model 2 (with $\alpha = 0.43$) highly underestimates the number of aftershocks until 10 days after Landers, because the low value of α yields a relatively small increase of seismicity at the time of the mainshock. Model 2 then provides a good fit to the end of the aftershock sequence, when enough aftershocks have occurred so that the predicted seismicity rate increases due to the importance of secondary aftershocks. The saturation of the number of aftershocks at early times in Figure 10 (for both the model and the data) is due to the increase of the threshold magnitude m_c (see equation (12)), which recovers the usual value $m_c = m_0 = 2$ about 10 days after Landers. Adding the corrective term $\rho^*(m)$ defined by (13), in order to account for the contribution of undetected early aftershocks in the rate of triggered seismicity, better predicts the rate of aftershocks just after Landers, but gives on average a smaller probability gain than without including this corrective term (see models 3 and 5 in Table 1 and Figure 10).

Figure 11 shows the predicted number of earthquakes and the probability gain (see equation (23)) in the time window $1992-1995$, for model 3. The model underestimates the rate of aftershocks for Joshua-Tree ($m = 6.1$) and Landers ($m = 7.3$) mainshocks, slightly over-estimates for Northridge ($m = 6.6$) and provides a good fit (not shown) for Hector-Mine ($m = 7.1$) and for San-Simeon ($m = 6.5$). All models overestimate by a factor larger than two the aftershock productivity of the 1987 $m = 6.6$ Superstition Hills earthquake. This shows that there is a variability of aftershock productivity not taken into account by

the model, which may partly be due to error in magnitudes. This implies that a model which estimates the parameters of each aftershock sequences (aftershocks productivity, Omori p exponent and GR b-value), such as the STEP model [Gerstenberger et al., 2004] may perform better than the ETES model which uses the same parameters for all earthquakes (except for the increase in productivity $\rho(m)$ with magnitude).

Figure 12 shows the predicted number of $m \geq 2$ earthquakes per day for model 3 (see Table 1) as a function of the observed number. Most points in this plot are close to the diagonal, i.e., the model generally gives a good prediction of the number of events per day. A few points however have a large observed number of earthquakes but a small value of the predicted number. These points correspond to days for which a large earthquake and its first aftershocks occurred while the seismicity was close to its background level, when the predicted seismicity rate was small.

We can complexify the model to take into account fluctuations of aftershock productivity, as done in the STEP model, by using early aftershocks to estimate the productivity $\rho(m)$ of large earthquakes. However, it needs to be investigated whether magnitude errors, biases and systematic effects significantly contribute to prediction efficiency. A method which adjust parameters to available data may seemingly perform better, especially in retrospective testing when various adjustments are possible. But if aftershock rate fluctuations are being caused by various technical factors and biases, this forecast advantage can be spurious.

Proportion of aftershocks in seismicity

The background seismicity is estimated to be $\mu_s = 2.81$ $m \geq 2$ earthquakes per day for model 3, compared to the long-term average rate $\mu_{LT} = 9.4$ earthquakes, i.e., the proportion of triggered earthquakes is 70%. This number under-estimates the actual fraction of triggered earthquakes, because it does not count the early aftershocks that occur a few hours after a mainshock, between the present time t_p and the end of the prediction window $t_p + T$ (see Figure 6). We have also removed from the catalog aftershocks smaller than the threshold magnitude $m_c(t, m)$ given by (12).

Scaling of aftershock productivity with mainshock magnitude

There may be several reasons for the small value $\alpha = 0.43$ selected by the optimization, compared to the value $\alpha = 1$ estimated by Felzer et al. [2004] and Helmstetter et al. [2005]. A smaller α value corresponds to a weaker influence of large earthquakes. A model with a small α has thus a shorter memory in time, and can adapt faster to fluctuations of the observed seismicity. A smaller α predicts a larger proportion of secondary aftershocks after a large mainshock. Therefore, it can better account for fluctuations of aftershock productivity. Indeed, if the rate of early aftershocks is low, a model with a small α will predict a small number of future aftershocks (less secondary aftershocks). In contrast, in case of a high rate of early aftershocks, it will predict a high aftershock activity due to the effect of secondary aftershocks. A model with a smaller α is also less sensitive to error in magnitudes. An error on the mainshock magnitude of 0.3 gives an error for the rate of direct aftershocks of a factor two for $\alpha = 0.8$, and a factor 1.3 for $\alpha = 0.4$. Finally, a model with a smaller α may provide a better forecast for the spatial distribution of aftershocks. Because the real spatial distribution of aftershocks is significantly different from the isotropic model (used for $m \leq 5.5$ earthquakes), a model with a smaller α may perform better than the model with the true α. A small α gives more importance to secondary aftershocks, and can thus better model the heterogeneity of the spatial distribution of aftershocks. In contrast, a larger α value produces a quasi-isotropic distribution at short times, dominated by the mainshock contribution.

The corrective contribution $\rho^*(m) > \rho(m)$ (13), introduced to take into account the contribution of missing aftershocks, can also bias the value of α. Using this term $\rho^*(m)$ with a value of α smaller than the true value over-estimates the contribution of small earthquakes just after a large earthquake when $m_c > m_0$. For this reason we did not used this contribution (except for model 5 in Table 1). We also tried to use this method (13) to take into account the fact that the minimum magnitude of triggering earthquakes m_0 is probably below the detection threshold. We found that this parameter m_0 is not constrained by maximizing the LL. The best fitting value of m_0 is equal to the detection threshold $m = 2$, and the LL is almost independent of m_0. Therefore, we took m_0 equal to the detection threshold $m_0 = 2$.

The main interest of short-term forecasts is to pre-

dict the rate of seismicity after a large mainshock, when the best model with $\alpha = 0.4$ clearly underestimates the observations. Therefore, we choose to constrain the value of $\alpha = 0.8$ (models 3 and 4 in Table 1). This model gives a slightly smaller likelihood than the best model, but provides a best fit just after a large mainshock. Note that this value $\alpha = 0.8$ is probably still smaller than the true value; a direct estimation of this parameter from a fit of the number of aftershocks as a function of mainshock magnitude gives indeed $\alpha = 1.05 \pm 0.1$ [Helmstetter et al., 2005]. Using for the forecasts a value of α smaller than the physical value allow us to improve the performance of the model because a smaller α best accounts for fluctuations in aftershock productivity as well as problems in the earthquake catalogs (incompleteness, location and magnitude errors).

Spatial distribution of aftershocks.

The power-law kernel (14) gives a slightly better LL than the Gaussian kernel (15) (see Table 1) for the unconstrained models 1 and 2 (α is adjustable parameter), but the Gaussian kernel works a little better when α is fixed to 0.8 (see models 3 and 4 in Table 1). The parameter f_d defined in (16) is the ratio of the typical aftershock zone $d(m)$ (16) and of the mainshock rupture length $L(m) = 0.01 \times 10^{0.5m}$ km. For the Gaussian kernel (15) $f_d \approx 1$, i.e., the average distance between a mainshock and its (direct) aftershocks is close to the mainshock rupture length.

For the power-law kernel (14), the average distance is not defined. In this case, $d(m)$ represents the scaling distance at which $f_{pl}(r)$ starts to decrease with r. The inversion of f_d using a power-law kernel gives an unrealistically small value of $f_d \leq 0.06$ for model 2 (see Table 1), so that $d(m) \approx 0.5$ km (fixed minimum value of $d(m)$ equal to the location error) independently of the magnitude of the triggering earthquake for $m \leq 5$. It gives short-range interactions, with most of the predicted rate concentrated in the cell of the triggering earthquake. Using a complex spatial distribution of aftershocks for $m \geq 5.5$ earthquakes (obtained by smoothing the location of early aftershocks, see section) slightly improves the LL compared to the simple isotropic kernel (see models 3 and 6 in Table 1).

Probability gain as a function of magnitude

Table 1 shows the variation of the probability gain G as a function of the lower cut-off magnitude $m_{\min}$. We used $m \geq 2$ earthquakes in models 7-11 to estimate the forecasted rate of $m \geq m_{\min}$ earthquakes, with $m_{\min}$ ranging between 3 and 6, and using the same parameters as in model 3 (but multiplying the background rate μ_s by $10^{-(m_{\min}-2.0)}$ to estimate the background rate for $m \geq m_{\min}$ earthquakes). The probability gain is slightly larger for $m_{\min} = 3$ than for $m_{\min} = 2$, but then G decreases with $m_{\min}$ for $m_{\min} \geq 4$. For $m_{\min} = 6$ (only 8 earthquakes), the long-term model (with a rate adjusted so that it predicts the exact number of observed events) performs even better than the ETES model ($G < 1$) for model #10 in Table 1.

We think that this variation with $m_{\min}$ does not mean that our model predicts only small earthquakes ("aftershocks"), or that larger earthquakes have a different distribution in space and time than smaller ones, but that these results simply reflects the large fluctuations of the probability gain from one earthquake to another one: the difference in likelihood between ETES and the long-term model is mainly due to a few large aftershock sequences. We thus need a large number of earthquakes and aftershock sequences to compare different forecasts.

Table 2 compares the predicted seismicity rate at the time and location of each $m \geq 6$ earthquake, estimated for the ETES model and for the long-term model. For each earthquake, we give two values of the predicted number of earthquakes, using the same parameters of the ETES model, but changing the time at which we update the forecasts, either midnight (universal time) for model #10 (see line 10 in Table 1) or at 13:00 for model #11. The large differences in the predicted seismicity rate between these two models show that the forecasts are very sensitive to short-term clustering, which has a large influence on the predicted seismicity rate. This suggests that the number of $m \geq 6$ earthquakes in the catalog (8 earthquakes from 1985 to 2004) is too small to compare our short-term and long-term models for this magnitude range.

While some of these $m \geq 6$ earthquakes are preceded by a short-term (hours) increase of seismicity (Superstition-Hill, Joshua-Tree, Landers, Big-Bear, Hector-Mine), the long-term model performs better than the ETES model if the forecasts are not updated between the foreshock activity and the mainshock (e.g., with model #10, between Elmore-Ranch and Superstition-Hill, and between Landers and Big-Bear). Landers occurred about two months after Joshua-Tree, and its hypocenter was just outside

Joshua-Tree aftershock zone, so that the predicted seismicity rate at the location of Landers hypocenter, and before the precursory foreshock activity (which started 6 hours before Landers) was slightly lower than the long-term rate. Joshua-Tree had foreshocks, which started two hours before the mainshock, and thus were not included in the daily forecasted rate for both ETES models. Hector-Mine was also preceded by foreshocks, with $m \leq 3.6$, which started about 20 hours before the mainshock. Therefore, the predicted seismicity rate (using ETES model #10) is 120 larger than the long-term rate for Hector-Mine. Other large $m \geq 6$ earthquakes (Elmore-Ranch, Northridge, San-Simeon), were not preceded by any significant foreshock activity. Therefore the forecasted seismicity was smaller than the long-term average.

Updating the forecasts more often (each hour, or after each earthquake) would of course increase the performance of our short-term forecasts. But optimizing and testing the forecasts would then be much more difficult and time consuming if the duration of the forecasts (one day) is different from the time interval between two forecasts. Moreover, preliminary earthquake catalogs are much less accurate in the first few hours, especially after a strong earthquake.

Conclusion

We have proposed a model for daily earthquake forecasts in southern California. Our model is based on empirical laws of seismicity: the Gutenberg-Richter magnitude distribution, Omori's law, and the exponential increase of triggered seismicity with the mainshock magnitude. Our model includes only data from earthquake catalogs (time, magnitude and locations).

Including small $m \geq 2$ earthquakes in long-term forecasts significantly improves the predictions. Our model performs better than a more complex one which incorporates geological data, when tested on $m \geq 5$ earthquakes since 1996. Our model also forecasts well the spatial distribution of future aftershocks by smoothing the locations of early aftershocks. We can obtain a good forecast of the aftershocks within a few hours of a large $m \geq 5.5$ earthquake, based on plentiful early aftershocks. Even if the spatial distribution of aftershocks is controlled by Coulomb stress changes, our empirical method may be more accurate, and faster, than direct calculations of the Coulomb stress change. Our method is accurate because the distribution of early aftershocks represents well the mainshock rupture surface, and because our method accounts for secondary aftershocks.

Retrospective tests for $m \geq 2$ earthquakes in the time period 1985/1/1 to 2004/3/10 show that our short-term model realizes a probability gain of 11.5 over a long-term stationary Poissonian forecast. Several features of our model could be improved. First, geologic slip rate and geodetic strain rate data could be used to better constrain the long-term seismicity. Second, a better estimate of the magnitude distribution, resulting from statistical studies of the relationship between fault geometry and earthquakes, could improve the forecasting of large quakes. Third, other research [e.g. Gerstenberger et al. 2004] suggests that aftershock productivity and magnitude distribution may vary considerably from one sequence to another. Comparing our model with others proposed to the Regional Earthquake Likelihood Models (RELM) working group [Kagan et al. 2003a; Jackson et al., 2004; Schorlemmer et al. 2005] should help to improve all available models. For example, the STEP model of Gerstenberger et al. [2004] can be compared directly with our model.

Acknowledgments. We acknowledge the Advanced National Seismic System for the earthquake catalog. This work is partially supported by NSF-EAR02-30429, by the Southern California Earthquake Center (SCEC) and by the James S. McDonnell Foundation 21st century scientist award/studying complex system. SCEC is funded by NSF Cooperative Agreement EAR-0106924 and USGS Cooperative Agreement 02HQAG0008. The SCEC contribution number for this paper is 895.

References

Bird, P., and Y. Y. Kagan, 2004. Plate-tectonic analysis of shallow seismicity: apparent boundary width, beta, corner magnitude, coupled lithosphere thickness, and coupling in seven tectonic settings, *Bull. Seismol. Soc. Amer.*, **94**(6), 2380-2399.

Console, R., M. Murru and A. M. Lombardi, Refining earthquake clustering models, J. Geophys. Res. 108, 2468, doi:10.1029/2002JB002130 (2003)

Daley, D. J., and Vere-Jones, D., 2004. Scoring probability forecasts for point processes: The entropy score and information gain, *J. Applied Probability*, **41A**, 297-312, (Sp. Iss.).

Felzer, K. R., R. E. Abercrombie and G. Ekström, Secondary aftershocks and their importance for aftershock forecasting, Bull. Seism. Soc. Am., 93, 1433-1448 (2003).

Felzer, K. R., R. E. Abercrombie, and G. Ekström (2004), A common origin for aftershocks, foreshocks, and multiplets, Bull. Seis. Soc. Am., 94, 88-99.

Frankel, A. et al., Seismic hazard maps for California, Nevada, and Western Arizona/Utah, United States Geological Survey Open-File Report 97-130 (1997).

Frankel, A. D., Petersen, M. D., Muller, C. S., Haller, K. M., Wheeler, R. L., Leyendecker, E. V., Wesson, R. L., Harmsen, S. C., Cramer, C. H., Perkins, D. M. and Rukstales, K. S. (2002) Documentation for the 2002 Update of the National Seismic Hazard Maps: U. S. Geological Survey, Open-file Report 02-420, 33 pp.

Gerstenberger, M., S. Wiemer, and L. Jones (2004) Real-time forecasts of tomorrow's earthquakes in California: a new mapping tool, U.S. Geological survey, Open file report, submitted.

Harris, R. A. and R. W. Simpson (1998) Suppression of large earthquakes by stress shadows: A comparison of Coulomb and rate-and-state failure, J. Geophys. Res., 103, 24,439-24,451.

Helmstetter A. (2003) Is earthquake triggering driven by small earthquakes?, Phys. Rev. Let. 91, 058501.

Helmstetter, A. and D. Sornette (2002) Sub-critical and super-critical regimes in epidemic models of earthquake aftershocks, J. Geophys. Res., 107, 2237, doi:10.1029/2001JB001580.

Helmstetter A. and D. Sornette (2003a) Predictability in the ETAS model of interacting triggered seismicity, J. Geophys. Res. 108, 2482, doi:1029/2003JB002485.

Helmstetter, A., and D. Sornette (2003b) Importance of direct and indirect triggered seismicity in the ETAS model of seismicity, Geophys. Res. Lett., 30, 1576, doi:1029/2003GL017670.

Helmstetter, A., and D. Sornette (2003c) Foreshocks explained by cascades of triggered seismicity, J. Geophys. Res., 108, 2457, 10.1029/2003JB002409.

Helmstetter A., D. D. Jackson, and Y. Y. Kagan (2005), Importance of small earthquakes for stress transfers and earthquake triggering, J. Geophys. Res. 110, B05S08, 10.1029/2004JB003286.

Izenman, A. J. (1991), Recent developments in nonparametric density estimation, J. Am. Stat. Assoc., 86, 205-224.

Jackson, D. D., and Y. Y. Kagan (1999) Testable earthquake forecasts for 1999, *Seism. Res. Lett.*, **70**(4), 393-403.

Jackson, D. D., D. Schorlemmer, M. Gerstenberger, Y. Y. Kagan, A. Helmstetter, S. Wiemer, and N. Field (2004), Prospective Tests of Southern California Earthquake Forecasts, *Eos Trans. AGU*, **85**(47), Fall Meet. Suppl., Abstract S21C-08.

Kafka, A. L. and S. Z. Levin (2000), Does the spatial distribution of smaller earthquakes delineate areas where larger earthquakes are likely to occur? Bull. Seism. Soc. Am., 90, 724-73.

Kagan, Y. Y. (1991) Likelihood analysis of earthquake catalogues, Geophys. J. Int., 106, 135-148.

Kagan, Y. Y. (1999) Universality of the seismic moment-frequency relation, Pure Appl. Geoph., 155, 537-573.

Kagan, Y. Y., 2004. Short-term properties of earthquake catalogs and models of earthquake source, Bull. Seismol. Soc. Amer., 94(4), 1207-1228.

Kagan, Y., and Knopoff, L. (1976) Statistical search for non-random features of the seismicity of strong earthquakes, Phys. Earth Planet. Inter., 12, 291-318.

Kagan, Y. Y. and L. Knopoff (1977) Earthquake risk prediction as a stochastic process, *Phys. Earth Planet. Inter.*, **14**(2), 97-108.

Kagan, Y. Y. and L. Knopoff (1987) Statistical short-term earthquake prediction, Science 236, 1563-1467.

Kagan, Y. Y. and D. D. Jackson (1994) Long-term probabilistic forecasting of earthquakes, J. Geophys. Res. 99, 13,685-13,700.

Kagan, Y. Y. and D. D. Jackson, (1998) Spatial aftershock distribution: Effect of normal stress, J. Geophys. Res. 103, 24,453-24,467.

Kagan, Y. Y. and Jackson, D. D. (2000) Probabilistic forecasting of earthquakes, Geophys. J. Int. 143, 438-453.

Kagan, Y. Y., D. D. Jackson, D. Schorlemmer, and M. Gerstenberger (2003a). Testing hypotheses of earthquake occurrence, *Eos Trans. AGU*, **84**(47), Fall Meet. Suppl., Abstract S31G-01.

Kagan, Y. Y., Y. F. Rong, and D. D. Jackson (2003b). Probabilistic forecasting of seismicity, Chapter 5.2 in *"EARTHQUAKE SCIENCE AND SEISMIC RISK REDUCTION"*, eds. F. Mulargia and R. J. Geller, pp. 185-200, Kluwer, Dordrecht.

Ogata, Y. (1988) Statistical models for earthquake occurrence and residual analysis for point processes, J. Amer. Statist. Assoc. 83, 9-27.

Ogata, Y., 2004. Space-time model for regional seismicity and detection of crustal stress changes, *J.*

Geophys. Res., **109**(B3), Art. No. B03308; Correction *J. Geophys. Res.*, **109**(B6), Art. No. B06308.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, Numerical Recipes in Fortran: The Art of Scientific Computing, 2nd ed., pp. 992, Cambridge Univ. Press, New York (1992).

Reasenberg, P. A., Second-order moment of central California seismicity, 1969-82, J. Geophys. Res. 90, 5479-95 (1985)

Reasenberg, P. A. (1999) Foreshock occurrence before large earthquakes, J. Geophys. Res. 104, 4755-4768.

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

Rhoades, D. A., and F. F. Evison, 2004. Long-range earthquake forecasting with every earthquake a precursor according to scale, *Pure Appl. Geophys.*, **161**(1), 47-72.

Schorlemmer D., M. Gerstenberger S. Wiemer, and D. D. Jackson (2005) Earthquake likelihood model testing, in preparation, preliminary draft available at http://moho.ess.ucla.edu/∼kagan/sjg.pdf.

Scotti, O., S. Steacy, M. Cocco, J. Zahradnik and J. McCloskey, Coulomb stress modelling as a practical tool in real-time aftershock hazard assessment: the example of the PRESAP blind test, EOS Trans. AGU, 84(46), Fall Meet. Suppl., Abstract S31A-06 (2003).

Steacy, S., D. Marsan, S. S. Nalbant and J. McCloskey, Sensitivity of static stress calculations to the earthquake slip distribution, J. Geophys. Res. 109, B04303, doi:10.1029/2002JB002365, 2004.

Stein, R. S. (1999) The role of stress transfer in earthquake occurrence, Nature 402, 605-609.

Wells, D. L. and K. J. Coppersmith (1994) New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement, Bull. Seism. Soc. Am. 84, 974-1002.

Wiemer, S. (2000) Introducing probabilistic aftershock hazard mapping, Geophys. Res. Lett., 27, 3405-3408.

Wiemer, S. and K. Katsumata (1999), Spatial variability of seismicity parameters in aftershock zones, J. Geophys. Res., 104, 13,135-13,151.

Zhuang J., Y. Ogata and D. Vere-Jones, Analyzing earthquake clustering features by using stochastic reconstruction, J. Geophys. Res. 109, B05301, doi:10.1029/2003JB002879, 2004.

Agnès Helmstetter, Lamont-Doherty Earth Observatory, 61 Rte 9W, Palisades, NY 10964 (e-mail: agnes@ldeo.columbia.edu)

David D. Jackson, Department of Earth and Space Sciences, University of California, Los Angeles, California 90095-1567. (e-mail: djackson@ucla.edu)

Yan Y. Kagan, Department of Earth and Space Sciences, University of California, Los Angeles, California 90095-1567. (e-mail: ykagan@ucla.edu)

This preprint was prepared with AGU's LaTeX macros v5.01. File predpap033005 formatted April 17, 2005.

Table 1. Model parameters, log likelihood LL, and probability gain G. The spatial kernel $f(r)$ is either a Gaussian ('gs', see Eq. (15)) or a power-law ('pl', see Eq. (14)).

model	$f(r)$	$m_{\min}$	α	p	K	μ_s	f_d	LL	G
1	gs	2.	0.44	1.18	1.18	2.42	1.03	-430928	11.66
2	pl	2.	0.43	1.19	1.29	2.09	0.06	-430754	11.69
3	gs	2.	0.80[a]	1.18	0.45	2.81	0.41	-434301.	11.07
4	pl	2.	0.80[a]	1.18	0.45	2.83	0.006	-435180.	10.93
5[b]	gs	2.	0.80[a]	1.20	0.38	2.98	0.44	-436103.	10.77
6[c]	gs	2.	0.80[a]	1.18[a]	0.45[a]	2.81[a]	0.41[a]	-436684.	10.68
7	gs	3.	0.80[a]	1.18[a]	0.45[a]	2.81 [a]	0.41[a]	-55077.	11.94
8	gs	4.	0.80[a]	1.18[a]	0.45[a]	2.81 [a]	0.41[a]	-7463.	8.46
9	gs	5.	0.80[a]	1.18[a]	0.45[a]	2.81[a]	0.41[a]	-986.	4.41
10	gs	6.	0.80[a]	1.18[a]	0.45[a]	2.81[a]	0.41[a]	-149.6	0.75
11[d]	gs	6.	0.80[a]	1.18[a]	0.45[a]	2.81[a]	0.41[a]	-135.9	4.16

[a]Fixed parameter

[b]Using the corrective term $\rho^*(m)$ defined by (13).

[c]Using the same isotropic kernel for all earthquakes (including $m \geq 5.5$ earthquakes).

[d]Updating the forecasts at 13:00 instead of 0:00 (universal time).

Table 2. Comparison of the predicted number of $m \geq 6$ events per day, for the days when a $m \geq 6$ earthquake occurred, at the location of the earthquake (i.e., within the cell of $0.05^o \times 0.05^o$), for the ETES model (N_{ETES}, using models 10 and 11 in Table 1) and for the long-term model (N_{LT}).

earthquake	date	time	m	N_{ETES}	N_{LT}	N_{ETES}/N_{LT}
Elmore-Ranch[a]	1987/11/24	01:54	6.2	4.3×10^{-8}	1.7×10^{-7}	0.25
Elmore-Ranch[b]				4.5×10^{-8}		0.26
Superstition-Hill[a]	1987/11/24	13:15	6.6	1.0×10^{-7}	4.0×10^{-7}	0.25
Superstition-Hill[b]				7.8×10^{-4}		1950.
Joshua-Tree[a]	1992/04/23	04:50	6.1	3.0×10^{-7}	9.6×10^{-7}	0.31
Joshua-Tree[b]				2.9×10^{-7}		0.30
Landers[a]	1992/06/28	11:57	7.3	1.2×10^{-6}	1.6×10^{-6}	0.75
Landers[b]				1.4×10^{-6}		0.87
Big-Bear[a]	1992/06/28	15:05	6.5	1.7×10^{-7}	6.2×10^{-7}	0.27
Big-Bear[b]				5.5×10^{-3}		8871.
Northridge[a]	1994/01/17	12:30	6.6	7.8×10^{-8}	2.8×10^{-7}	0.28
Northridge[b]				7.9×10^{-8}		0.28
Hector-Mine[a]	1999/10/16	09:46	7.1	6.5×10^{-5}	5.4×10^{-7}	120.
Hector-Mine[b]				3.1×10^{-7}		0.57
San-Simeon[a]	2003/12/22	19:15	6.5	4.2×10^{-8}	1.6×10^{-7}	0.26
San-Simeon[b]				4.2×10^{-8}		0.26

[a]Forecasts updated each day at 00:00 universal time (model 10 in Table 1)

[b]Forecasts updated each day at 13:00 universal time (model 11 in Table 1)

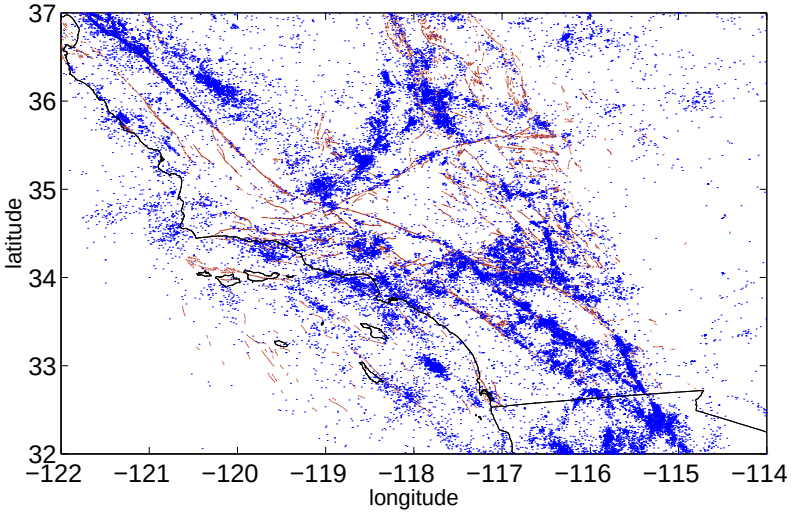

Figure 1. Declustered catalog, obtained with Reasenberg's [1985] algorithm, including 6861 $m \geq 3$ earthquakes in the time window $1932 - 1979$, and 46,937 earthquakes with $m \geq 2$ for $1980 - 2003$.

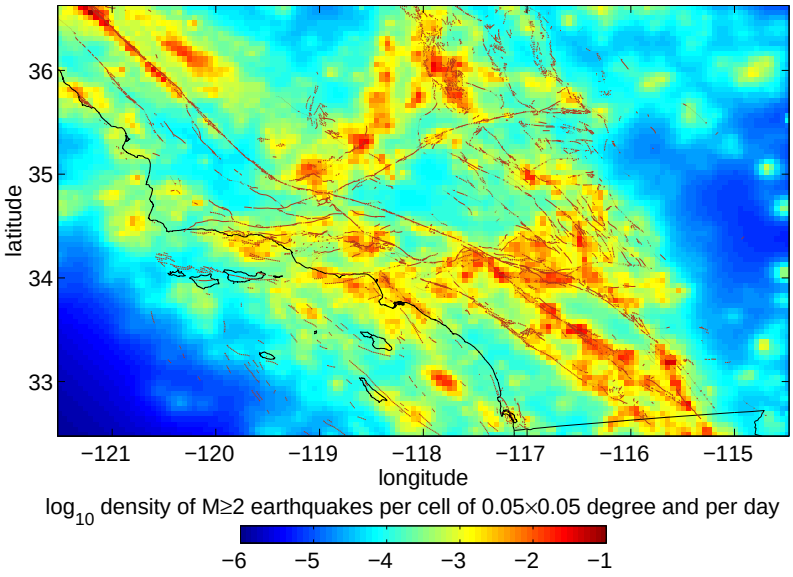

Figure 2. Long-term density of seismicity $\mu_{\mathrm{LT}}(\vec{r})$ obtained by declustering and smoothing the ANSS catalog.

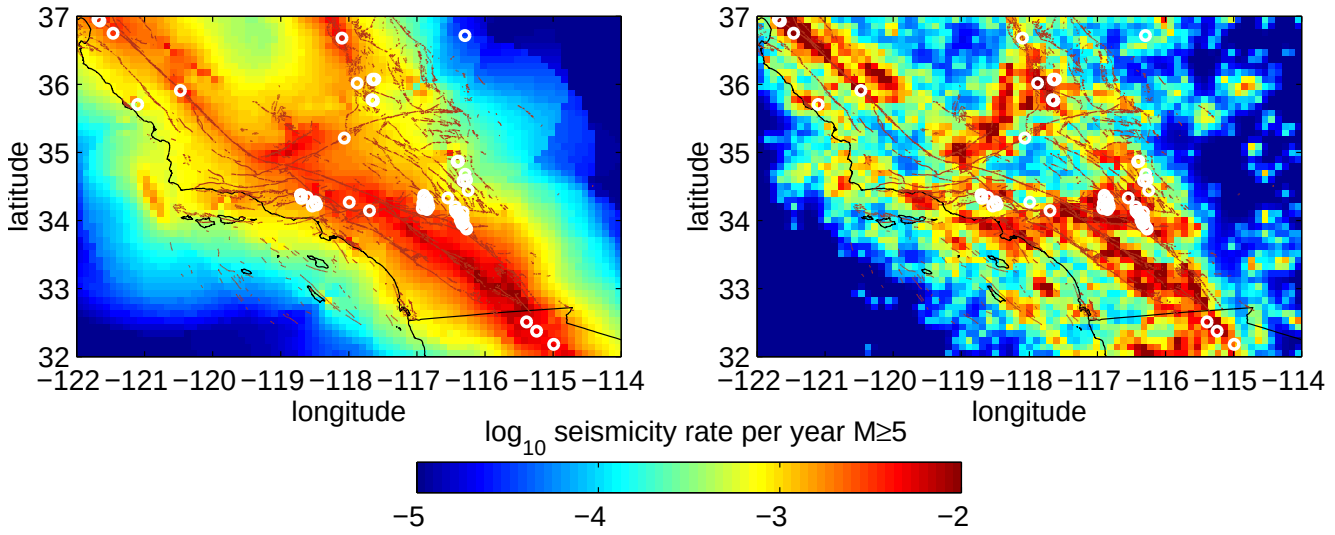

Figure 3. Long-term density of seismicity $\mu_{\mathrm{LT}}(\vec{r})$ for the model by Kagan and Jackson [1994; 2000] (KJ94) (left) and for the present work (right). White circles represent $m \geq 5$ earthquakes that occurred between 1990 and 2004.

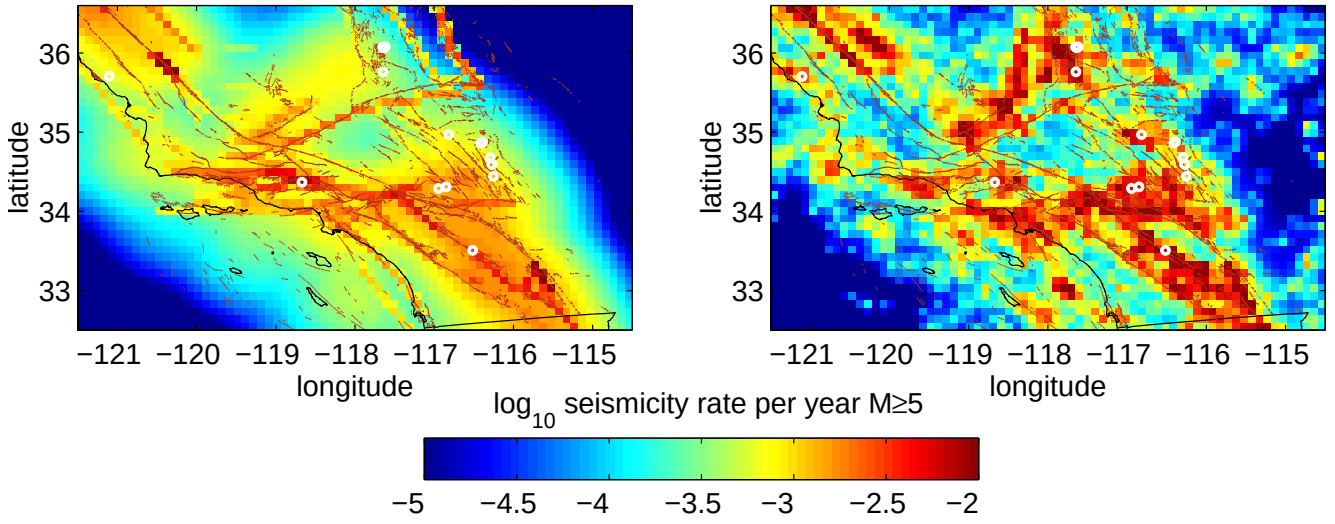

Figure 4. Long-term density of $m \geq 5$ earthquakes $\mu_{\mathrm{LT}}(\vec{r})$ for the model by Frankel et al. [1997] (left) and for our long-term model (right). White circles represent $m \geq 5$ earthquakes that occurred between 1996 and 2004.

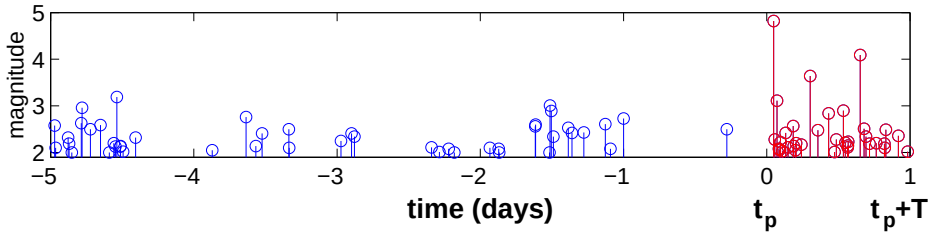

Figure 5. Plot of the magnitude versus time for a few days in the ANSS catalog, which illustrates the fact that a significant fraction of earthquakes that will occur in the next day (between the present time t_p and $t_p + T$) may be triggered by earthquakes that will occur in the next day ($t_p < t < t_p + T$).

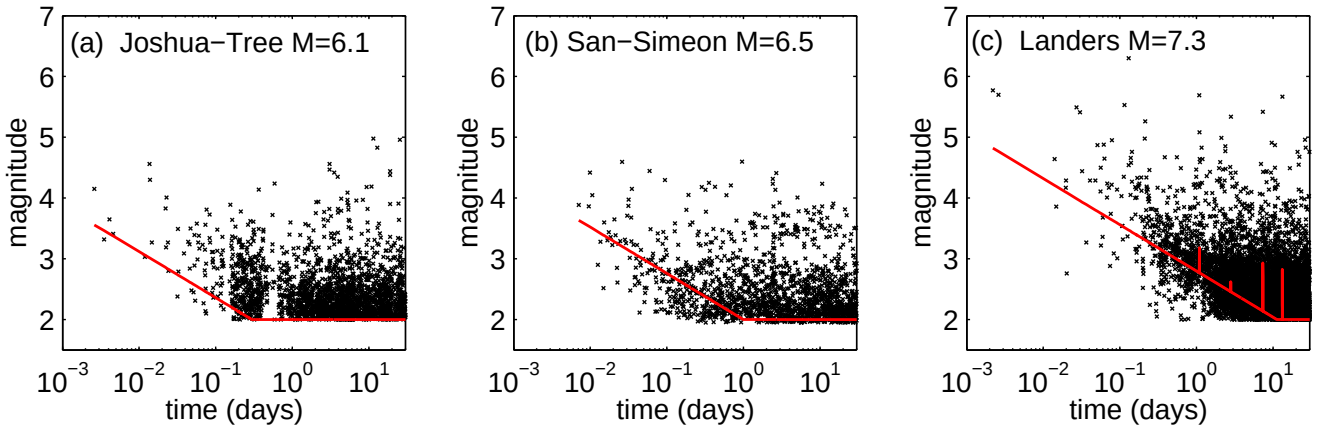

Figure 6. Magnitude versus time since mainshock for aftershocks (a) of Joshua-Tree $m = 6.1$, (b) San-Simeon $m = 6.5$, and (c) Landers $m = 7.3$ earthquakes. The continuous line represents the threshold magnitude estimated from (12) and includes the effect of all $m \geq 5$ earthquakes. The vertical lines in (c) are due to the increase of $m_c(t)$ after large $m \geq 5$ aftershocks. Dates of these earthquakes are shown in Table 2.

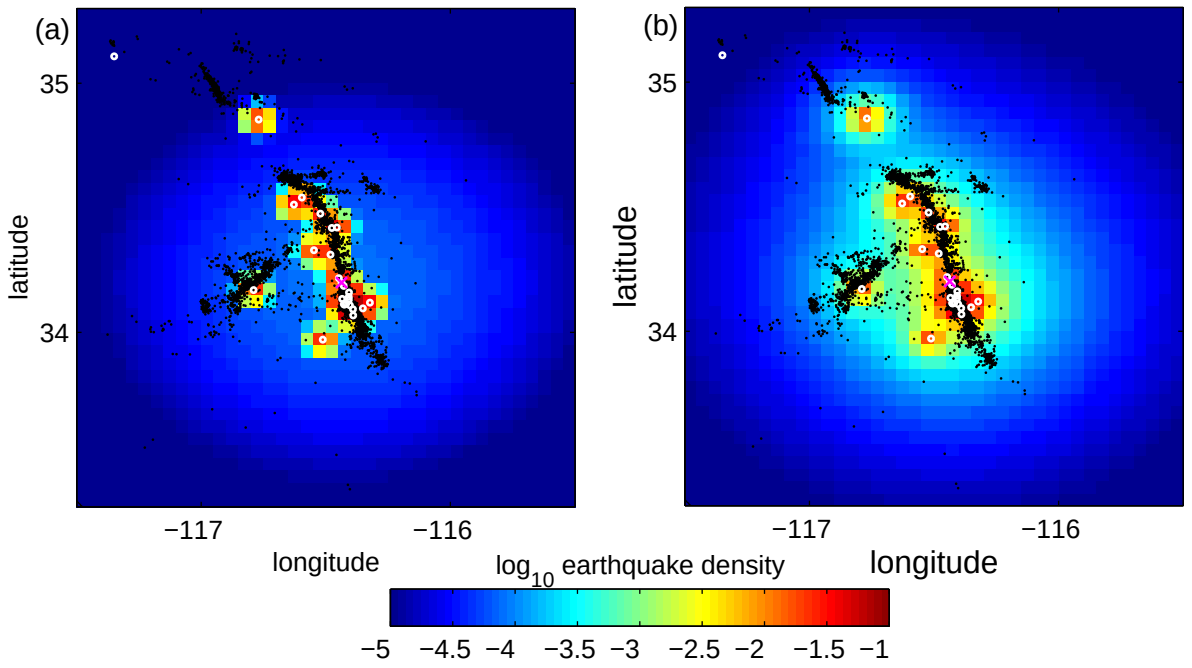

Figure 7. Density of aftershocks estimated by smoothing the location of early aftershocks (white circles) that occurred less than two hours after Landers mainshock ($m = 7.3$, 1992/06/28), using either (a) a Gaussian kernel (15) or (b) a power-law kernel (14).

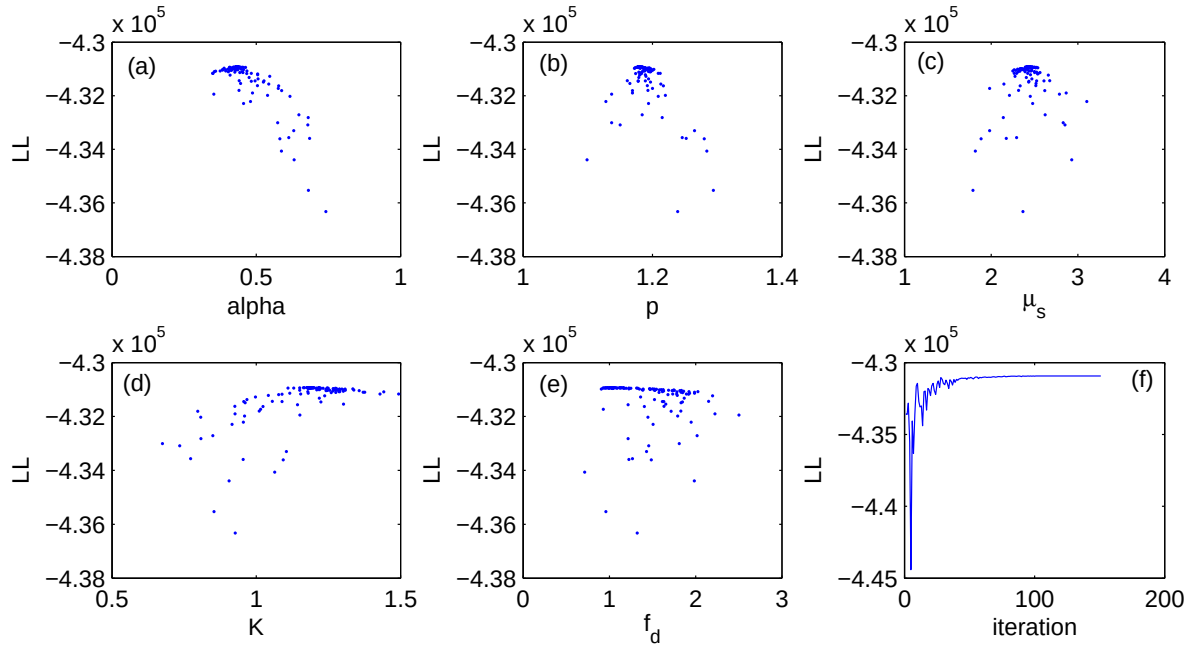

Figure 8. Results of the optimization of the log-likelihood LL for the unconstrained model 1, using a Gaussian kernel. Value of LL as a function of each model parameter (a-e) and as a function of the number of iterations (f).

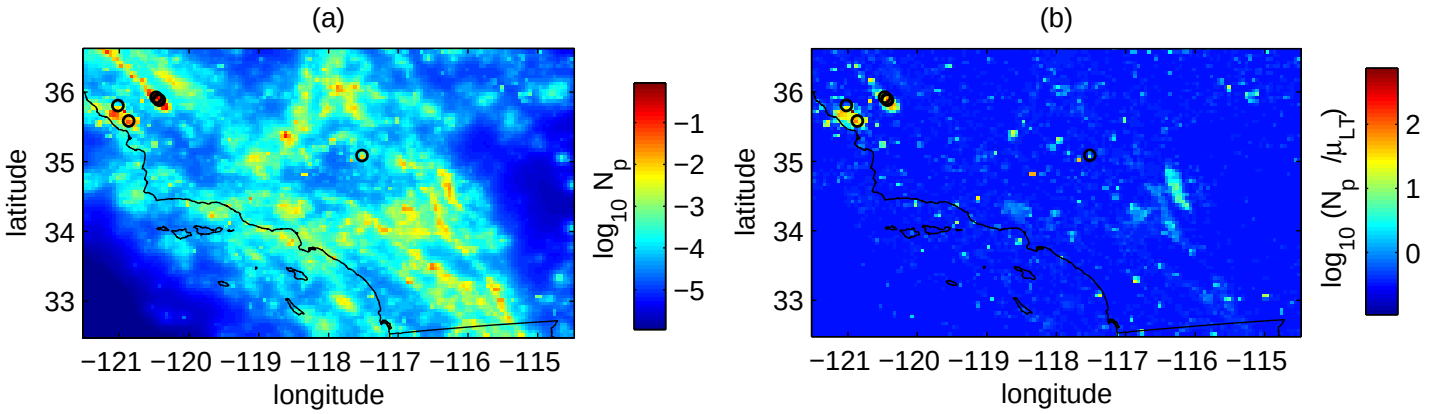

Figure 9. (a) Forecasted number of events with $m \geq 2$ per cell for the day of October 23^{rd} 2004 (logarithmic scale). Black circles represent observed earthquakes with $m \geq 2$ which occurred during this day. Two of these events are aftershocks of the $m = 6.5$ 2003/12/22 San-Simeon earthquake (located at lat=35.7° and lon=−121.1°), three are associated with the $m = 6$ 2004/09/28 Parkfield mainshock (lat=35.81° and lon=−120.37°), and one is an aftershock of a $m = 3.7$ 2004/09/09 earthquake (lat=35.09° and lon=−117.52°). The predicted number of events for this day was 8.39 and the observed number was 6. Most of these earthquakes are better predicted by the time-dependent ETES forecast than by the long-term model (i.e., $N_p(\vec{r}) > \mu_{\rm LT}(\vec{r})$ in the cells within which these earthquakes occurred). Only one event, which occurred at 11 km away from the San-Simeon earthquake (lat=35.81° and lon=−121.02°), was better predicted by the long-term model, because it occurred just outside of the main aftershock zone. (b) Ratio of the forecasted number of events estimated using the time dependent and long-term model (logarithmic scale). High values of $N_p/\mu_{\rm LT}$ (up to 800) are associated with recent large earthquakes, such as Parkfield, San-Simeon, Landers, Hector-Mine and Northridge.

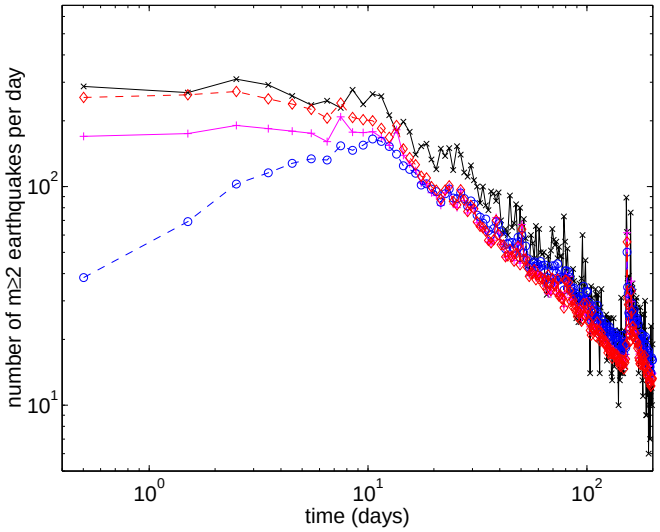

Figure 10. Observed (crosses), and predicted number of $m \geq 2$ earthquakes per day as a function of the time since Landers mainshock, for model 2 (circles), model 3 (pluses), and model 6 (diamonds). The saturation at $t \leq 10$ days is due to the incompleteness of the catalog for small magnitudes.

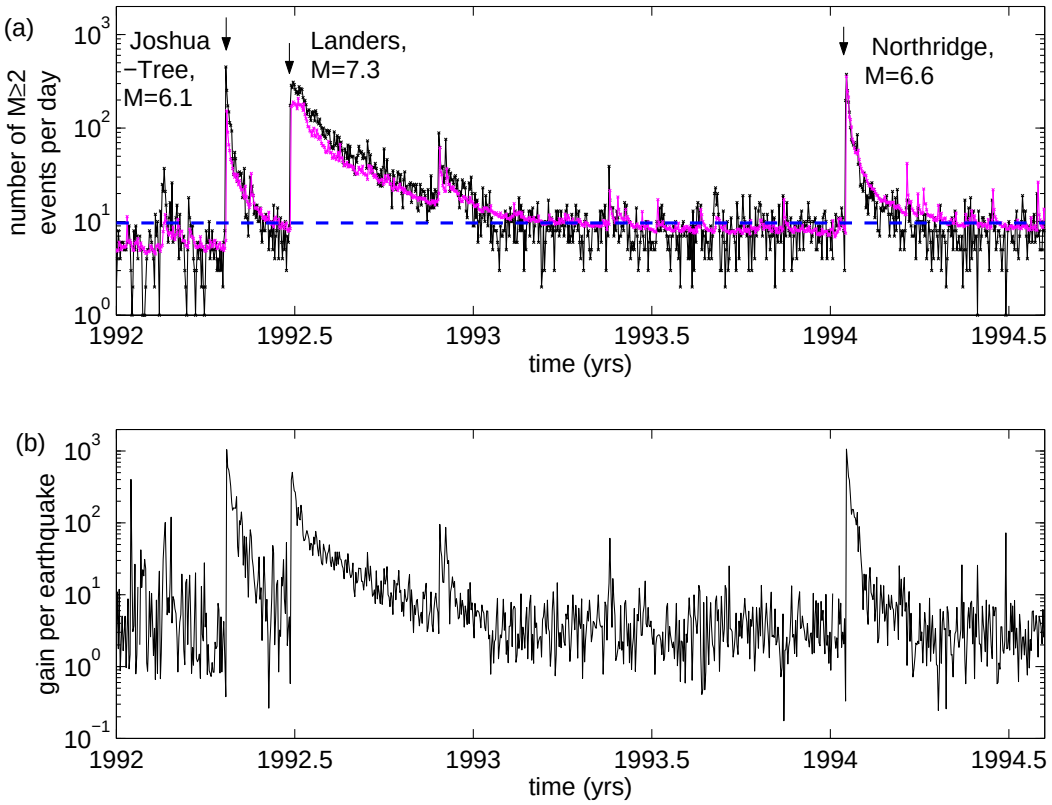

Figure 11. (a) Observed (black) and predicted (gray) number of $m \geq 2$ earthquakes per day in southern California for model 3 (see Table 1). Dashed line is the background rate $\mu_s = 2.81$/day. (b) Probability gain per earthquake defined in (23).

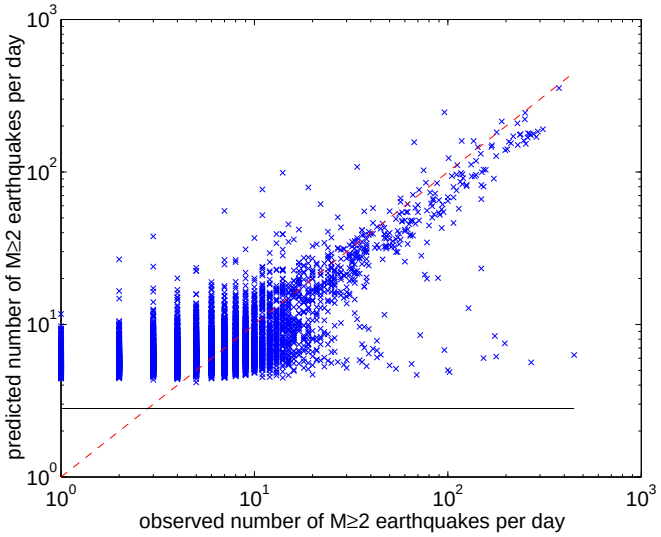

Figure 12. Predicted number of $m \geq 2$ earthquakes per day for model 3 (see Table 1) as a function of the observed number, for the time period 1985-2003. The dashed line represents the perfect fit. The horizontal line is the background rate $\mu_s = 2.81$/day.