High-Resolution Long-Term and Short-Term Earthquake Forecasts for California

by Maximilian J. Werner,* Agnès Helmstetter, David D. Jackson, and Yan Y. Kagan

Abstract We present two models for estimating the probabilities of future earthquakes in California, to be tested in the Collaboratory for the Study of Earthquake Predictability (CSEP). The first is a time-independent model of adaptively smoothed seismicity that we modified from Helmstetter et al. (2007). The model provides five-year forecasts for earthquakes with magnitudes $M \geq 4.95$. We show that large earthquakes tend to occur near the locations of small $M \geq 2$ events, so that a high-resolution estimate of the spatial distribution of future large quakes is obtained from the locations of the numerous small events. We further assume a universal Gutenberg–Richter magnitude distribution. In retrospective tests, we show that a Poisson distribution does not fit the observed rate variability, in contrast to assumptions in current earthquake predictability experiments. We therefore issued forecasts using a better-fitting negative binomial distribution for the number of events. The second model is a time-dependent epidemic-type aftershock sequence (ETAS) model that we modified from Helmstetter et al. (2006) and that provides next-day forecasts for $M \geq 3.95$. In this model, the forecasted rate is the sum of a background rate (proportional to the time-independent model rate) and of the expected rate of triggered events due to all prior earthquakes. Each earthquake triggers events with a rate that increases exponentially with its magnitude and decays in time according to the Omori–Utsu law. An isotropic kernel models the spatial density of aftershocks for small ($M \leq 5.5$) events, while for larger quakes, we smooth early aftershocks to forecast later events. We estimate parameter values by optimizing retrospective forecasts and find that the short-term model realizes a probability gain of about 6.0 per earthquake over the time-independent model.

Online Material: Identification of explosions and ETAS parameters.

Introduction

A wide range of ideas and hypotheses exist about how, when, and where earthquakes will occur and about how big they will be. Given the strongly stochastic nature of seismicity and the scarce data of large earthquakes, the most promising path toward weeding out the good ideas from the bad ones is to evaluate the hypotheses in a rigorous and transparent manner. Prospective earthquake forecasts make scientific hypotheses of earthquake occurrence testable, transparent, and refutable. A major step along this path was taken by the Working Group on Regional Earthquake Likelihood Models (RELM), which invited long-term (five-year) forecasts for California in a specific format to facilitate comparative testing (Field, 2007; Schorlemmer et al., 2007; Schorlemmer and Gerstenberger, 2007; Schorlemmer et al., 2010). The Collaboratory for the Study of Earthquake Predictability (CSEP) later inherited and expanded RELM’s mission to regionally and globally test prospective forecasts (Jordan, 2006; Schorlemmer et al., 2010; Zechar, Schorlemmer et al., 2010).

The development of better models to be tested in CSEP remains the highest priority. We present two models to estimate the probabilities of future earthquakes in California. The first model provides time-independent, long-term (five-year) estimates of the probabilities of future earthquakes of magnitudes $M \geq 4.95$ in California. The second model estimates short-term (one-day) probabilities of future $M \geq 3.95$ earthquakes. Both models are extended and/or modified versions of those developed by Helmstetter et al. (2007) and Helmstetter et al. (2006). We made improvements to the
long-term model and updated it using the latest available data. The short-term model was extended from southern California to all of California. We made some further modifications and reestimated its parameters using all available data up to 1 April 2010.

Both models are based on simple yet hotly contested hypotheses. Both models solely require past seismicity as data input. Whether and at which point other data such as geological or geodetic data could improve the forecasts remains an interesting open question. For the long-term model, we hypothesized that future earthquakes are more likely to occur in areas where past earthquakes have occurred, including small ones. To turn this hypothesis into a testable forecast, we smoothed the locations of past seismicity using an adaptive kernel and tested the resulting density’s predictive skill. Moreover, the magnitude of each earthquake is independently distributed according to a tapered Gutenberg–Richter distribution with corner magnitude $M_c$, 8.0 relevant for California (Bird and Kagan, 2004), irrespective of geological setting, proximity to mapped faults, or the sizes of past earthquakes. Because of the assumed magnitude independence, the model also uses small $M \geq 2$ events to forecast large $M \geq 4.95$ events. The small quakes indicate regions of active seismicity, and retrospective tests support the claim that including these small events improves forecasts of larger ones.

The time-dependent, short-term model makes similar assumptions. First, the same assumption holds regarding the distribution of magnitudes. Second, the subjective, retrospective distinction between fore-, main- and aftershocks is eliminated: every earthquake can trigger other events, and those triggered events may be larger than the triggering shock. The time-dependence enters in the form of the well-known Omori–Utsu law (Utsu et al., 1995). However, in this model the Omori–Utsu law applies to all earthquakes, not just to large earthquakes that trigger smaller ones. In addition to the contribution to the seismicity from past quakes, the model also includes a background rate, which is modeled as a spatially heterogeneous Poisson process. The background’s normalized spatial distribution is taken directly from the first, time-independent model, but its rate is estimated. The model is a particular implementation of the epidemic-type after-shock sequence (ETAS) model (Ogata, 1988) and may be viewed as an expression of the simple but powerful null hypothesis of earthquake clustering and triggering, based on empirical distributions of seismicity.

In this article (and in the electronic supplement to this paper), we describe the two models, their calibration on earthquake catalogs, results from retrospective forecasts, and what we learned about seismicity in the process. With these forecasts, we respond to CSEP’s call for testable forecasts in a specific format: the expected number of quakes in individual magnitude bins of width $0.1$ from $M_4$ to $M_9$ in each spatial cell of $0.1^\circ$ by $0.1^\circ$ latitude/longitude in a predefined testing region that includes California and extends about $1^\circ$ beyond its borders (Schorlemmer and Gerstenberger, 2007).

In this paper, we first describe the data, explain the method for estimating the spatial distribution of spontaneous seismicity, calculate the expected number of events, and apply the Gutenberg–Richter distribution to generate the time-independent five-year forecast. Then we introduce the ETAS model, describe the parameter estimation, discuss the model’s goodness of fit to the data, and compare the ETAS forecast with the time-independent model during the Baja California swarm of February 2008.

Data: The ANSS Earthquake Catalog

We used the Advanced National Seismic System (ANSS) catalog in the period from 1 January 1981 until 1 April 2010 with magnitude $M \geq 1.7$ in the RELM/CSEP collection region, defined by the polygon in table 2 by Schorlemmer and Gerstenberger (2007) (see Fig. 1a). The catalog listed 300,278 events. Underground nuclear explosions at the Nevada Test Site contaminate the earthquake catalog. In the electronic supplement to this paper, we document how we identified and removed 21 explosions from the catalog, 9 of which were large $M \geq 5$ events. We also deleted 156 events that were exact copies of other entries, leaving 300,101 earthquakes $M \geq 1.7$ in the collection region. We selected earthquakes after 1981 to ensure a homogeneous data set recorded and analyzed using the electronic Caltech/USGS Seismic Processing system (Hutton et al., 2010).

Spatial Distribution of Spontaneous Seismicity

Adaptive Kernel Smoothing of Seismicity

Following Helmstetter et al. (2007), we estimated the density of seismicity at location $r$ by smoothing the locations of earthquakes using an isotropic adaptive kernel $K_d(r)$. We compared two choices for $K_d(r)$, a power-law

$$K_d(r) = \frac{C(d)}{(|r|^2 + d^2)^{3/2}},$$

and a Gaussian

$$K_d(r) = C'(d) \exp\left[-\frac{|r|^2}{2d^2}\right],$$

where $d$ is the adaptive smoothing distance, and $C(d)$ and $C'(d)$ are normalizing factors, so that the integral of $K_d(r)$ over an infinite area equals 1. We measured the smoothing distance $d_i$ associated with an earthquake $i$ as the horizontal (epicentral) distance between event $i$ and its $k$th closest neighbor. The number of neighbors, $k$, is an adjustable parameter, which we estimated by optimizing the resulting spatial forecasts (see Optimizing the Spatial Smoothing). We imposed $d_i \geq 0.5$ km to account for location uncertainty. The kernel bandwidth $d_i$ thus decreases in regions of dense seismicity, so that we have better resolution (smaller $d_i$) where the density is higher.
The density $\mu(r)$ at any point $r$ is then estimated from $N$ earthquakes available in a learning or input earthquake catalog by

$$\mu(r) = \sum_{i=1}^{N} K_d(|r - r_i|).$$  \hspace{1cm} (3)

However, our forecasts are given as an average number of events in each 0.1° by 0.1° cell. We therefore integrated equation (3) over each cell to obtain the seismicity rate in this cell.

Correcting for Magnitude Incompleteness

Although we would like to use the entire catalog with magnitudes $M \geq 1.7$ to calibrate the models, the catalog is not complete everywhere at this magnitude level. To correct for catalog incompleteness, we made several modifications to the method proposed by Helmstetter et al. (2007), who used a slightly modified maximum curvature method. First, we used the entire (nondeclustered) catalog with magnitudes $M \geq 1.7$, rather than just the declustered catalog. Second, we decoupled the estimation of the completeness magnitude from the spatial optimization of the kernel bandwidths by estimating the spatially variable completeness magnitude $M_0$ first, and then used the same map of $M_0$ to calibrate all models. Third, we added 0.3 magnitude units to the estimate of the completeness magnitude because the maximum curvature method is slightly biased toward smaller values (Ogata and Katsura, 1993; Woessner and Wiemer, 2005), raising the minimum completeness magnitude to $M_0 1.7$. Finally, we corrected a minor problem in the computer program of Helmstetter et al. (2007) that had artificially increased the completeness magnitude near the boundaries of the study region.

The present method is fundamentally different from most of the other catalog-based methods that estimate the completeness magnitude (e.g., Wiemer and Wyss, 2000; Woessner and Wiemer, 2005; Amorèse, 2007; Mignan et al., 2011), which treat all magnitudes within a fixed radius as equally important for characterizing the local magnitude distribution. Here, in contrast, the local magnitude distribution $P_M(r, M)$ at point $r$ is constructed by calculating the local kernel density (from the smoothed seismicity in equation 3) and weighting the magnitude of each smoothed earthquake by its distance

$$P_M(r, M) = \sum_{i=1}^{N} K_d(|r - r_i|)G_h(M - M_i).$$  \hspace{1cm} (4)

where $G_h(M)$ is a Gaussian kernel function of fixed bandwidth $h$. The kernel width $h$ was fixed to 0.15 magnitude units. We integrated $P_M(r, M)$ over each cell to get the magnitude distribution in these cells. We estimated the completeness magnitude $M_0$ in each cell as the magnitude at which the

Figure 1.  (a) Original ANSS catalog $M \geq 2$ from 1 January 1981 until 1 April 2010 in the CSEP collection polygon around California. (b) Original (black) and declustered (gray) seismicity per month. (c) Original (black) and declustered (gray) cumulative seismicity.
smoothed magnitude distribution is at a maximum (an isolated maximum, see Helmstetter et al., 2007) and then added 0.3 magnitudes to counter bias.

We smoothed the large, unphysical small-scale fluctuations of $M_0$ using a Gaussian filter with a standard deviation of 0.15°. The resulting map is shown in Figure 2, where we used the Gaussian kernel (equation 2) with a smoothing distance $d_i$ in equation (4) to the sixth nearest neighbor ($k = 6$). Most of the region has $M_0 \approx 2$. The completeness magnitude is much larger, close to $M_0 3.5$, close to the boundaries of the collection region, especially in the Mendocino area and in Mexico.

Because we are interested in estimating the rate of earthquakes $M \geq M_{\text{min}}$, we corrected the observed rate for missing events by applying a Gutenberg–Richter scaling (Gutenberg and Richter, 1944) with $b = 1$ (see Magnitude Distribution for the $b$-value estimation)

$$\mu'(r) = \mu(r) \times 10^{b(M_0(r) - M_{\text{min}})}, \quad (5)$$

Declustering Seismicity

To estimate the spatial distribution of independent (non-triggered) events $M \geq 2$, we used a modified version of the Reasenberg declustering algorithm (Reasenberg, 1985), as described by Helmstetter et al. (2007). Figure 1 compares the original with the declustered catalog. The declustering algorithm found that 56% of the events $M \geq 2$ were spontaneous. We chose this method to estimate spontaneous seismicity over more sophisticated algorithms based on stochastic process theory (Kagan, 1991; Zhuang et al., 2002, 2004; Marsan and Lengliné, 2008) because of its simplicity. The declustering algorithm was not optimized for forecasting and the particular procedure may have significant effects on the forecasts. In the future, we would like to test and optimize the declustering algorithm, too.

Optimizing the Spatial Smoothing

We estimated the parameter $k$, the number of neighbors used to compute the smoothing distance $d_i$ in equation (3), by maximizing the likelihood of a target set of earthquakes given a forecast. We built a model $\mu'(i_x, i_y)$ in each cell $(i_x, i_y)$ on a training period of the catalog and tested it (evaluated the likelihood) on a separate testing period of the catalog. Following Helmstetter et al. (2007), we optimized solely the normalized spatial density estimate in each cell $(i_x, i_y)$ using

$$\mu^{\ast}(i_x, i_y) = \frac{\mu'(i_x, i_y) N_i}{\sum_{i_x} \sum_{i_y} \mu'(i_x, i_y)}, \quad (6)$$

where $N_i$ is the number of observed target events. The expected number of events for the model $\mu^{\ast}$ thus equals the observed number $N_i$ to optimize solely the spatial forecast.

We assumed that the observed earthquakes in each cell are Poisson random variables, independent of each other in space and time. We revisit this assumption in the following text. The log-likelihood of the observations given the model is thus given by

$$\text{LL} = \sum_{i_x} \sum_{i_y} \log p[\mu^{\ast}(i_x, i_y), n], \quad (7)$$

where $n$ is the number of events that occurred in cell $(i_x, i_y)$, and the probability $p$ of observing $n$ events in cell $(i_x, i_y)$ given a forecast of $\mu^{\ast}(i_x, i_y)$ in that cell is given by the Poisson distribution.

Figure 2. (a) Completeness magnitude estimated from the maxima of the local magnitude distribution. (b) Smoothed completeness magnitude. The color version of this figure is available only in the electronic edition.
We built an extensive set of background models $\mu^*$ to test whether various model choices are important, including declustering, correcting for magnitude completeness, using a Gaussian or a power-law kernel, or changing the learning and target catalogs. The training and the target catalog were chosen so that they do not overlap. We generated forecasts based on the training catalog and evaluated them on the separate target catalog to test our forecasts out-of-sample and to avoid overfitting. We evaluated the performance of each model by calculating its probability gain per earthquake relative to a model with a reference spatial density:

$$p[\mu^*(i_x, i_y), n] = [\mu^*(i_x, i_y)]\frac{\exp[-\mu^*(i_x, i_y)]}{n!}. \quad (8)$$

Results of the Spatial Smoothing Optimization

In Figure 3, we show the effect of several model choices on the probability gain per earthquake of the generated forecasts. For this figure, we used a learning catalog from 1981 until 2004 (inclusive), a (nondeclustered) target catalog from 2005 to 2009 and only used the Gaussian kernel. First, we calculated the gains without declustering or correcting for magnitude incompleteness. For target magnitudes $M \geq 3.5$, the gains monotonically decreased as we raised the threshold of the learning catalog from 2 to 5, suggesting that including small earthquakes helps forecast the locations of moderate and large earthquakes. However, for target magnitudes $M \geq 4.95$, the gains first increase with increasing magnitude threshold of the learning catalog and then decrease. As investigated further in the following text, this concave trend is probably not a robust or significant feature. We suspect that the small number (22) of $M \geq 4.95$ target earthquakes may cause large fluctuations. Next, we used a declustered catalog as input and found that the resulting forecasts achieve higher gains irrespective of target and learning catalog thresholds. This finding suggests that a long-term forecast based on a nondeclustered catalog is dominated by short-term clustering and is too localized in space, unless one accounts for the temporal Omori–Utsu decay of clustering. We then corrected the declustered catalog for magnitude incompleteness and found that the resulting forecasts are slightly improved whenever the magnitude threshold of the learning catalog is low enough for the correction to make a difference.

In Figure 4 we compare the gains achieved by the two different kernels, from now on using a declustered learning catalog that was corrected for missing earthquakes. There is a very slight tendency for the Gaussian kernel to achieve higher gains.

In Figure 5 we investigated the fluctuations of the probability gains of different target periods. For the target magnitude threshold $M \geq 3.5$, both the values and the monotonically decreasing trend are relatively robust features. In contrast, the gains for target magnitudes $M \geq 4.95$ fluctuate more strongly. The gains of two of the four target periods show monotonically decreasing trends, while the two others display convex patterns. This suggests that statistical fluctuations strongly influence the gains whenever few target earthquakes (e.g., the largest) probe the forecast. When more target earthquakes are available, such as when the target threshold is lowered to $M \geq 3.5$, including small $M \geq 2$ earthquakes yields higher scores.

Figure 3. Effect of different choices of the learning catalog on the gain per earthquake of the resulting forecasts for the target period 2005–2009 for target magnitude thresholds (a) $M \geq 3.5$ and (b) $M \geq 4.95$.

Figure 4. Effect of choosing a Gaussian or power-law smoothing kernel on the gain per earthquake of the resulting forecasts for the target period 2005–2009 for target magnitude thresholds (a) $M \geq 3.5$ and (b) $M \geq 4.95$. 

\[ G = \exp\left(\frac{LL - LL_{\text{ref}}}{N_t}\right). \quad (9) \]
Following Helmstetter et al. (2007), we assigned a modified magnitude distribution to the region near the geothermally active Geysers region in northern California. Above $M = 3.3$, the $b$-value there is $b \approx 1.72$. The empirical and fitted magnitude distributions for California and the Geysers are shown in Figure 6.

Expected Number of Events

To estimate the expected number of earthquakes, we found the average rate of $M \geq 4.95$ events in the period from 1981 to 2010 in the CSEP testing region: $N_{\text{pred}} = 6.72$ earthquakes per year. The expected number of events per year in each space-magnitude bin $(i_x, i_y, i_M)$ was then calculated from

$$E(i_x, i_y, i_M) = N_{\text{pred}} \mu(i_x, i_y) P(i_M). \quad (11)$$

where $\mu$ is the normalized spatial background density, and $P(i_M)$ is the integrated probability of an earthquake in magnitude bin $(i_M)$ defined according to equation (10), taking into account the different $b$-value for the Geysers.

Figure 7 shows the new five-year forecast for $M \geq 4.95$ for 2012 to 2016 based on optimally and adaptively smoothing the locations of declustered, small $M \geq 2$ earthquakes. On average, we expect 33.6 earthquakes $M \geq 4.95$ over five years in the entire test region. The modified magnitude distribution near the Geysers helps us avoid an unrealistically high rate of large quakes there.

Comparison of the New Five-Year Forecast to the Forecast by Helmstetter et al. (2007)

In Figure 8 we show the ratio of the new forecast over the forecast by Helmstetter et al. (2007). Overall, the new
The forecast is not very different from the old one. We expect differences because of the updated data set, the Gaussian (as opposed to power-law) kernel, and the different bandwidth. In general, the new forecast is more specific (the southeast and northwest, where little seismicity occurs, have a new rate that is 30 times lower than before). However, there are also more unexpected differences. The narrow stripes of increases or decreases along the boundaries were artifacts in the old forecast, which we fixed by correctly smoothing the completeness magnitude. Another region showing a factor 10 difference is the area near the Geysers.

Helmstetter et al. (2007) estimated $b \approx 1.99$ in this region, while we estimated $b \approx 1.71$. We also used a slightly different technique to extrapolate the observed $M \geq 2$ rate to events $M \geq 4.95$.

The forecast by Helmstetter et al. (2007) was submitted to the five-year RELM experiment (Field, 2007). Schorlemmer et al. (2010) evaluated the submitted models after the first half of the five-year period. Twelve earthquakes $M \geq 4.95$ had been observed (the black squares in Fig. 8, five of which overlap in Baja California). The model by Helmstetter et al. (2007) outperformed all other models during this period. The new forecast, which was partially built on the same data set used to evaluate the forecast by Helmstetter et al. (2007), would have outperformed the old forecast by a gain of $G = 1.33$ per event.

However, how would the new forecast compare if it were built on the same training data? We created a forecast based on the same data that Helmstetter et al. (2007) used, up to 23 August 2005. To compare performance in the RELM experiment, we needed to estimate the magnitude distribution and the number of expected events. We estimated $b \approx 1$ for California and $b \approx 1.88$ for the Geysers, closer to the estimate $b \approx 1.99$ by Helmstetter et al. (2007) than our previous one of $b \approx 1.71$, for which we used all data up to 1 April 2010. For the first half of the five-year RELM experiment period, the new forecast performed worse than the previous one ($G = 0.85$ per earthquake). Three events near the Mendocino Triple Junction were identified as culprits for the lower score. However, the differences are small, and future earthquakes can probably change the probability gain significantly.

Retrospective Consistency Tests of the Five-Year Forecast

Thus far, we optimized the model by maximizing its likelihood via retrospective forecasts, but we did not test whether the forecast can actually explain past data. To do that, we used two consistency tests (Kagan and Jackson, 1994; Schorlemmer et al., 2007): the number of events test.
(N-test) to check that the number of observed events is consistent with the expected number of events, and the likelihood test (L-test) to test whether the observed likelihood score is consistent with the model’s expected values.

Commonly, for example, in the RELM experiment (Schorlemmer et al., 2007, 2010), a Poisson distribution is assumed to calculate the probability of a given number of earthquakes. In Figure 9a, we show the results of the N-test assuming a Poisson distribution for many overlapping five-year target periods starting from 1981. The 95% confidence bounds of the Poisson distribution are denoted with black bars. Under this assumption of variability in the number of observed earthquakes, the five-year forecast was rejected by the N-test in several target periods: in those that contain the 1992 $M_w$ 7.3 Landers earthquake (too many observed events) and in a few that happen during the relatively quiet period 1999–2003 (too few observed events, despite the 1999 $M_w$ 7.1 Hector Mine sequence).

However, the assumption of Poissonian variability in the number of observed events is evidently wrong: the variance of the number of observed events per five-year period is larger than expected from a Poisson process. The model is rejected 10 times in 24 overlapping periods. Moreover, because the number of observed events sometimes exceeded the permissible maximum and sometimes dipped below the minimum, one cannot construct a Poisson forecast that would be accepted by the N-test for every target period (or 95% of the time). Thus, a Poisson forecast is inappropriate for forecasting the number of earthquakes and, as a result, forecasts rejected by the Poissonian N-test may not be rejected by more realistic assumptions of the number variability.

To find a more realistic distribution, we analyzed the empirical distribution of the number of observed events $M \geq 4.95$ in nonoverlapping five-year intervals in the ANSS catalog from 1932 to 2008 within the CSEP testing region.

Figure 9. Retrospective evaluation of the new five-year, time-independent forecast on past, overlapping five-year target periods with starting year as ordinates. (a) N-test: crosses denote the expected number of earthquakes, black bars display the 95% confidence bounds assuming a Poisson distribution, gray bars denote those of a negative binomial distribution. Circles indicate the numbers of observed earthquakes, circles without thick edges indicate passed forecasts. Symbols with thick edges indicate that the Poisson forecast was rejected but not the NBD forecast. (b) L-test: crosses with gray bars denote the expected unconditional likelihood score and associated 95% confidence bounds (the original L-test), crosses with black bars denote the expected conditional score and associated bounds (the conditional L-test). Circles denote observed likelihood scores, symbols without thick edges denoting passed forecasts. Symbols with thick edges indicate that the original L-test failed the forecast but not the conditional L-test. The NBD forecast could not be rejected by the N-test, and the conditional L-test also does not reject the forecast. The color version of this figure is available only in the electronic edition.
We compared the fit of the Poisson distribution to the empirical distribution with the fit of a negative binomial distribution (NBD), which is defined by:

\[ p(k|\tau, \nu) = \frac{\Gamma(\tau + k)}{\Gamma(\tau)k!} \nu^\tau (1 - \nu)^k, \tag{12} \]

where \( k \) is a nonnegative integer, \( \Gamma \) is the gamma function, \( 0 \leq \nu \leq 1 \), and \( \tau > 0 \). There are many discrete distributions, but the NBD is simple, fits well, and has been used before (Vere-Jones, 1970; Kagan, 1973; Jackson and Kagan, 1999). The Akaike Information Criterion (AIC) favors the NBD (\( \text{AIC} = 130.4 \)) over the Poisson distribution (\( \text{AIC} = 246.1 \)). Using maximum likelihood, we obtained parameter estimates of the NBD (\( \tau \approx 2.76, \nu \approx 0.08 \)) and of the Poisson distribution (\( \lambda \approx 29.9 \)).

Armed with a more realistic distribution of the number of events, we created an NBD forecast for the number of earthquakes. The two parameters of our NBD forecast can be calculated from the mean and variance of the distribution. For the mean, we preferred our earlier estimate (from Expected Number of Events), \( N^\text{pred} = 33.68 \), based on the more recent, higher quality ANSS data from 1981 to 2009. To estimate the variance \( \text{Var}(N^\text{obs}) \approx 368.1 \), however, we used the longer data set from 1932. From these values, we obtained the NBD parameters (\( \tau \approx 3.37, \nu \approx 0.09 \)). The 95\% confidence bounds of the NBD are shown as gray bars in Figure 9a and are much wider than the Poisson bounds. Using the more realistic NBD, the forecast cannot be rejected during any of the target periods.

Now we turn to the L-test. We applied the L-test proposed by Schorlemmer et al. (2007) to the forecast for each of the five-year target periods. The crosses with gray bars in Figure 9b show the expected likelihood score with associated 95\% confidence bounds. The L-test rejected the forecast in the target periods that contain the 1992 \( M_w 7.3 \) Landers earthquake and otherwise passed the forecast. However, we observed strong anticorrelations between the observed likelihood scores and the number of observed earthquakes, suggesting that the L-test is dominated by the number of observed events and conveys little information in addition to the N-test.

To make the test more sensitive to the space and magnitude dimension of the forecast, we compared the observed likelihood score with the distribution of scores that would be expected if the model were correct, conditional on the number of observed earthquakes. In contrast, the L-test by Schorlemmer et al. (2007) compared the observed score with the unconditional distribution of likelihood scores, which includes and is dominated by the variation in the number of forecast earthquakes. Zechar, Gerstenberger, and Rhoades (2010) proposed the S- and M-tests, which, similar to our test, also condition on the number of observed earthquakes but additionally sum over the magnitude and spatial bins, respectively, and therefore only probe the spatial or magnitude component of the forecast. In contrast, the conditional L-test evaluates the joint spatial and magnitude dimensions of the forecast. Werner et al. (2010) applied this conditional L-test to a retrospective evaluation of the time-independent forecasts submitted to the recent Italian earthquake predictability experiment.

The results of the conditional L-test are contrasted with those of the unconditional L-test in Figure 9b. The confidence bounds are indicated by black bars and are much narrower. Conditional on the number of observed earthquakes, the forecast cannot be rejected during any of the periods. Needless to say, passing retrospective tests should be a first requirement for earthquake forecasts, because the forecast is tested against observations that were used to create it. On the other hand, a failure to adequately fit past data would have indicated problems in the model or its calibration.

In this section, we dropped the Poisson distribution in favor of the NBD to better forecast the number of earthquakes. This innocent, minor modification has major implications:

1. The sole time-independent stochastic point process is the Poisson process, and conversely a non-Poissonian distribution implies a time-dependent process.
2. We used an NBD for the total number of events, but in each individual bin the variability remains specified as Poissonian, even though summing the Poissonian rates over all bins cannot result in a NBD.

How can we justify these inconsistencies? The short answer is because it is a quick and simple solution to a much deeper problem; an approximate solution that comes at the cost of a slight theoretical inconsistency. Moreover, the conditional L-test no longer tests the total number of observed events and is thus to some extent decoupled from the NBD distribution. The time-independent model thus serves as a useful and simple approximation to a time-dependent process. This is the approach we take in this article. But in the long run, we cannot justify the inconsistency; we need time- and history-dependent process with memory, even for long-term forecasts. The ETAS branching model we discuss in the following text is such a process, and in the future, we would like to use it for long-term forecasts, too.

The Epidemic-Type Aftershock Sequences (ETAS) Model

Definition of the ETAS Model

The epidemic-type aftershock sequence (ETAS) model (Ogata, 1988, 1998) is a stochastic spatiotemporal branching point-process model of seismicity. There are numerous flavors within this family of models (see Kagan and Knopoff, 1987; Kagan, 1991; Felzer et al., 2002; Helmstetter and Sornette, 2002; Console et al., 2003; Zhuang et al., 2004; Hainzl and Ogata, 2005). Here, we used the particular formulation of Helmstetter et al. (2006), albeit with some (minor) modifications. The total seismicity rate \( \lambda(t, r, m) \) at time \( t \), location \( r \), and for magnitude \( M \) is the sum of a background rate \( \mu_b(r) \), a spatially heterogeneous time-independent
Poisson process and the sum of individual contributions to the triggering potential from all prior earthquakes occurring at times \( t_i < t \) with magnitudes \( M_i \)

\[
\lambda(t, r, M) = P_M(M) \left[ \mu_b(r) + \sum_{t_i < t} \phi_M(r, t - t_i) \right],
\]

(13)

where \( P_M(M) \) is the space- and time-independent tapered Gutenberg–Richer distribution of magnitudes (equation 10). The triggering function \( \phi_M(r, t) \) describes the spatiotemporal triggering potential at a distance \( r \) and time \( t \) after an earthquake of magnitude \( M \)

\[
\phi_M(r, t) = \rho(M) \psi(t) f(r, M),
\]

(14)

where \( \rho(M) \) is the average number of earthquakes triggered by a quake of magnitude \( M \geq M_d \)

\[
\rho(M) = k 10^{\alpha(M - M_d)},
\]

(15)

the function \( \psi(t) \) is the normalized Omori–Utsu law

\[
\psi(t) = \frac{(p - 1)c^{p-1}}{(t + c)^p}.
\]

(16)

and \( f(r, M) \) is a normalized spatial aftershock density at a distance \( r \) to the parent shock of magnitude \( M \). The constant \( M_d \) refers to a chosen threshold magnitude above which the model parameters are estimated, which may, in general, be different from the completeness magnitude \( M_b \) and the target magnitude threshold \( M_{\min} \). We tested both a Gaussian and a power-law kernel, as described in Modeling the Spatial Distribution of Aftershocks. Following Helmstetter et al. (2006), we fixed the \( c \)-value in the Omori–Utsu law (equation 16) to 0.035 days (5 minutes). We used the same tapered Gutenberg–Richer distribution (equation 10) as for the long-term forecast, with a \( b \)-value of 1.0 and corner magnitude \( M_c \), 8.0. We used earthquakes \( M \geq 2 \) to forecast events \( M_{\min} \geq 3.95 \) for the CSEP experiment. Because the earthquake catalog is not complete down to \( M_d = 2 \) after large earthquakes, we corrected for this effect (see Correcting for Spatial and Temporal Magnitude Incompleteness) and time-independent spatial incompleteness. The background rate \( \mu_b \) is given by

\[
\mu_b(r) = \mu_s \mu_0(r),
\]

(17)

where \( \mu_0(r) \) is equal to the spatial time-independent model normalized to 1, which we estimated earlier. Thus, the parameter \( \mu_s \) represents the expected number of \( M \geq M_{\min} \) background events per day.

ETAS Forecasts for One-Day Bins

The ETAS model is defined via its conditional intensity (equation 13), which is the instantaneous probability of an event (Daley and Vere-Jones, 2003). To forecast the number of earthquakes over a finite period, one cannot directly use the conditional intensity, because intervening quakes may change the rate. One solution to forecast over a one-day period is to simulate earthquakes according to the conditional intensity at the beginning of the one-day bin, and then to obtain a mean forecast by averaging over all simulations. Such simulations are computationally intensive, and it is not clear whether the 10,000 simulations commonly used to produce spatiotemporal forecasts are adequately sampling the possibilities. In this study, we followed the solution proposed by Helmstetter and Sornette (2003) and Helmstetter et al. (2006) to forecast the number of events of the next day, but with effective parameters \( k, \alpha, \mu_s \) that differ from the original parameters of the ETAS model because a Poissonian likelihood function of successive next-day forecasts is maximized.

Correcting for Spatial and Temporal Magnitude Incompleteness

To correct for the spatial incompleteness, we used the same completeness magnitude estimate as for the time-independent forecast (Fig. 2). However, for the time-dependent forecasts, the temporary incompleteness of catalogs after large earthquakes is also important. Kagan (2004), Helmstetter et al. (2006), Peng et al. (2007), and Lennartz et al. (2008) showed that the completeness magnitude after large earthquakes can temporarily increase by several units. One effect is that the forecast overestimates the observed rate because of missing events. Another effect is that secondary triggering is underestimated because early undetected aftershocks contribute to the rate. We followed the solutions proposed by Helmstetter et al. (2006) to correct both effects. We used their equation (15) to estimate the temporal magnitude incompleteness \( M_b(t, M) \) to predict the detection threshold at the time of each earthquake. We only selected earthquakes with \( M > M_b \) to estimate the seismicity rate (equation 13) and to calculate the likelihood of the forecasts. The second effect, the missing contribution from undetected aftershocks \( M_d < M < M_b \) to the observed seismicity rate, is estimated by an additive term \( p^s \) to the productivity defined by equation (16) of Helmstetter et al. (2006).

Modeling the Spatial Distribution of Aftershocks

Following Helmstetter et al. (2006), we tested different choices for the spatial kernel \( K_{d(m)}(r, m) \), which models the aftershock density at a distance \( r \) from a parent shock of magnitude \( M \). As before, we compared a power-law function (equation 1) with a Gaussian kernel (equation 2). The spatial regularization distance \( d(M) \), which replaces the adaptive bandwidth \( d_i \) in the kernels (equation 1) and (equation 2), accounts for the finite rupture size and for location errors. We assumed that \( d(M) \) is given by

\[
d(M) = 0.5 + f_d \cdot 0.01 \times 10^{0.5M} \text{ km},
\]

(18)
where the parameter \( f_d \) is estimated by optimizing the forecasts.

Large \( M > 5.5 \) earthquakes with a rupture length larger than the grid cell size of \( 0.1^\circ \) are rarely followed by isotropic aftershock distributions. We therefore used a more complex, anisotropic kernel for these events, as done previously by Wiemer and Katsumata (1999), Wiemer (2000), and Helmstetter et al. (2006). We smoothed the locations of early, nearby aftershocks to estimate the mainshock fault plane and other active faults in the immediate vicinity. We computed the distribution of later aftershocks of large \( M \geq 5.5 \) quakes by smoothing the locations of early aftershocks using

\[
K_d(r, M) = \sum_{i=1}^{N} K_d(|r - r_i|, M_i), \tag{19}
\]

where the sum is over the mainshock and all earthquakes that occur within a distance \( D_{\text{aft}}(m) \) before the issue time \( t_p \) of the forecast and not after some time \( T_{\text{aft}} \) after the mainshock. We usually took \( D_{\text{aft}} = 0.02 \times 10^{0.5 m} \text{ km} \) (approximately two rupture lengths) and \( T_{\text{aft}} = 2 \text{ days} \) but also tested other values. The kernel \( K_d(r, m) \) used to smooth the locations of early aftershocks is either a power-law (equation 1) or a Gaussian distribution (equation 2), with an aftershock zone length given by equation (18) for the mainshocks, but by \( d = 2 \text{ km} \) for the aftershocks.

Smoothing the locations of early aftershocks to forecast the spatial distribution of later aftershocks is a fast and completely automatic method to estimate the mainshock rupture plane along which aftershocks tend to cluster. In Estimated Parameter Values we tested several values of the method’s parameters.

Definition of the Likelihood and Estimation of the ETAS Model Parameters

We inverted for the values of five parameters: \( p \), the exponent in Omori’s law (equation 16); \( k \) and \( \alpha \), characterizing the productivity (equation 15); \( \mu_s \), the number of background events per day (equation 17), and \( f_d \), the size of the aftershock zone (equation 18). We estimated effective parameters by maximizing the cumulative likelihood of the next-day forecasts using a Poisson distribution (equation 8). The joint log-likelihood of the observations given a model is the sum of likelihoods over each space-time-magnitude bin \((i_x, i_y, i_M)\):

\[
\text{LL} = \sum_{i_x=1}^{N_x} \sum_{i_y=1}^{N_y} \sum_{i_M=1}^{N_M} \log p(N_p(i_x, i_y, i_M), n), \tag{20}
\]

where \( n \) is the number of observed events in each bin and the probability \( p(\cdot, n) \) is a Poisson distribution with a rate given by the expected number of events \( N_p(i_x, i_y, i_M)\):

\[
N_p(i_x, i_y, i_M) = \int_{i_x} \int_{i_y} \int_{i_M} \lambda(r, t, M) dM dy dx dt. \tag{21}
\]

The rules of the CSEP one-day forecast group determine the bin sizes: 1 day in time, 0.1° longitude and latitude in space, and 0.1 units of magnitude (Schorlemmer and Gerstenberger, 2007).

We maximized the log-likelihood (equation 20) using earthquakes above various thresholds \( M_{\text{min}} \) from 1 January 1986 until 1 April 2009 in the CSEP testing region to test the forecasts. To calculate the seismicity rate (equation 13), however, we took into account earthquakes \( M \geq 2 \) since 1 January 1981 that occurred within CSEP collection region. We tested the effect of variations of the spatial aftershock kernel and different target magnitude thresholds.

To quantify the performance of the short-term forecasts with respect to the time-independent forecast, we used the probability gain per earthquake equation (9) of the ETAS model likelihood with respect to the time-independent but spatially varying background density that we estimated earlier. We normalized \( \mu(r) \) so that the total number of expected target events equalled the observed number.

By maximizing the likelihood of the Poisson forecasts, we assumed that the Poisson distribution in each space-time-magnitude bin is a first-order approximation to the actual, model-dependent distribution. The actual likelihood function of the ETAS model is known, but the predictive next-day likelihood function must be simulated and can deviate substantially from the Poisson distribution. However, rather than performing computationally intensive simulations, we used the current model to estimate the mean rate in each bin. Extensions beyond the Poisson distribution will be left for the future.

Estimated Parameter Values

The parameter estimation program was modified from the one written by Helmstetter et al. (2006): we extended the region to the CSEP collection region of California, and we fixed some issues in the code. We tested different versions of the ETAS model (various spatial kernels, different learning and target magnitude thresholds, different parameter values of the early aftershock-kernel smoothing procedure, etc.). Table 1 presents the results.

The exponent \( \alpha \) in the productivity law (equation 15) measures the relative importance of small versus large earthquakes for the triggering budget (Helmstetter, 2003; Felzer et al., 2004; Helmstetter et al., 2005; Christophersen and Smith, 2008; Hainzl et al., 2008). Felzer et al. (2004) and Helmstetter et al. (2005) found that \( \alpha = 1 \) by fitting the number of aftershocks as a function of mainshock magnitude. Because the number of small earthquakes increases exponentially with decreasing magnitude, their collective ability to trigger earthquakes equals that of the large earthquakes (Helmstetter, 2003). As a consequence, small, undetected earthquakes have a significant, time-dependent impact on the observed seismicity budget and the failure to model their effect causes parameter bias (Sornette and Werner, 2005a, 2005b; Saichev and Sornette, 2005, 2006; Werner, 2007).
The latter has much farther reach than the Gaussian kernel. Larger than of models with the power-law kernel, because the ground rate of models with the Gaussian kernel was slightly magnitude threshold (Sornette and Werner, 2005b). The back- mate of the fraction of triggered events depends on the because events below 56%. However, this is a lower bound on the actual proportion 2007; Zhuang et al., 2006) estimated \( \alpha = 0.8 \pm 0.1 \) for all models for targets \( M \geq 3.95 \). Because of the strong negative correlation between \( \alpha \) and \( K \), it is possible that \( \alpha = 1 \) is within the uncertainties of the estimates. The estimate of \( \alpha \) decreases when the aftershock kernel is purely isotropic and does not include the smoothing of early aftershocks, consistent with the simulations by Hainzl et al. (2008). Helmstetter et al. (2006) estimated \( \alpha \approx 0.43 \) using essentially the same optimization procedure that we used, but for the target magnitude range \( M \geq 2 \). In the electronic supplement to this paper, we list the estimated parameters for the \( M \geq 2 \) range: the estimates are almost identical to those of Helmstetter et al. (2006).

The exponent \( p \) in Omori’s law was relatively high, \( p \approx 1.27 \). However, typical estimates are based on smaller magnitudes, they maximize a non-Poissonian likelihood function, and the entire cascade of triggered earthquakes is usually fit, while we obtained an estimate of the direct or local exponent (Helmstetter and Sornette, 2002).

The background rate was also fairly stable across the different models, at \( \mu_5 \approx 0.08 \), which was about 44% of the average daily number of earthquakes \( M \geq 3.95 \) over the entire period. The estimated proportion of triggered events is thus 56%. However, this is a lower bound on the actual proportion because events below \( M_0(t) \) were not counted, and any estimate of the fraction of triggered events depends on the magnitude threshold (Sornette and Werner, 2005b). The background rate of models with the Gaussian kernel was slightly larger than of models with the power-law kernel, because the latter has much farther reach than the Gaussian kernel.

The parameter \( f_d \) in Table 1 is a measure of the size of the aftershock zone. For the Gaussian spatial kernel (equation 2), the length \( d(m) \) (equation 18) equals the standard deviation of the Gaussian, so that \( f_d \) should be about 1. This parameter can be used as a sanity check of the optimization. We found reasonable values for \( f_d \) between 0.5 and 0.9 in the case of the Gaussian kernel.

Using the power-law kernel, \( f_d \) was very small: between 0.08 and 0.25. When we did not smooth the locations of early aftershocks to forecast later ones, the value increased to a more reasonable 0.25. Nevertheless, these values are difficult to interpret. It suggests using the Gaussian kernel because the fit is better understood, despite the marginally better performance of the power-law kernel.

A lower probability gain per earthquake was obtained for both spatial kernels when we did not smooth the locations of early aftershocks to forecast the locations of later after- shocks: For example, model 1, which used the anisotropic smoothing method for events \( M > 5.5 \), scored \( G = 6.19 \), while model 7, which used the isotropic kernel, obtained \( G = 6.04 \). The gains of the smoothing method are particularly strong during the early days in an aftershock sequence, as expected. We calculated the average daily log-likelihood ratio for the first 10 days after all 18 \( M > 6 \) events in the target period and found that model 1 outperformed model 7 by an average daily log-likelihood ratio of 0.28. During the aftershock sequences, the smoothing method thus works better than the isotropic kernel. However, the gains are somewhat diluted because the parameters that use the isotropic kernel tend to give better forecasts on the days of the actual

### Table 1

<table>
<thead>
<tr>
<th>Model</th>
<th>Kernel</th>
<th>( \alpha )</th>
<th>( p )</th>
<th>( k )</th>
<th>( \mu_5 )</th>
<th>( f_d )</th>
<th>( L_{ETAS} )</th>
<th>( L_{TT} )</th>
<th>( N_{obs} )</th>
<th>( N_{pred} )</th>
<th>( G )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gaussian</td>
<td>0.84</td>
<td>1.28</td>
<td>0.34</td>
<td>0.083</td>
<td>0.89</td>
<td>-15.977</td>
<td>-18.749</td>
<td>1521</td>
<td>1529.0</td>
<td>6.19</td>
</tr>
<tr>
<td>2</td>
<td>Power law</td>
<td>0.82</td>
<td>1.26</td>
<td>0.39</td>
<td>0.072</td>
<td>0.08</td>
<td>-15.944</td>
<td>-18.729</td>
<td>1521</td>
<td>1528.8</td>
<td>6.32</td>
</tr>
<tr>
<td>3*</td>
<td>Gaussian</td>
<td>0.81</td>
<td>1.30</td>
<td>0.33</td>
<td>0.084</td>
<td>0.78</td>
<td>-15.991</td>
<td>-18.729</td>
<td>1521</td>
<td>1518.9</td>
<td>6.13</td>
</tr>
<tr>
<td>4*</td>
<td>Power law</td>
<td>0.81</td>
<td>1.27</td>
<td>0.32</td>
<td>0.075</td>
<td>0.17</td>
<td>-15.957</td>
<td>-18.729</td>
<td>1521</td>
<td>1483.5</td>
<td>6.27</td>
</tr>
<tr>
<td>5*</td>
<td>Gaussian</td>
<td>0.84</td>
<td>1.27</td>
<td>0.33</td>
<td>0.081</td>
<td>0.94</td>
<td>-15.976</td>
<td>-18.729</td>
<td>1521</td>
<td>1511.3</td>
<td>6.19</td>
</tr>
<tr>
<td>6*</td>
<td>Power law</td>
<td>0.82</td>
<td>1.27</td>
<td>0.42</td>
<td>0.074</td>
<td>0.09</td>
<td>-15.945</td>
<td>-18.729</td>
<td>1521</td>
<td>1533.6</td>
<td>6.31</td>
</tr>
<tr>
<td>7*</td>
<td>Gaussian</td>
<td>0.84</td>
<td>1.27</td>
<td>0.35</td>
<td>0.081</td>
<td>0.59</td>
<td>-15.972</td>
<td>-18.729</td>
<td>1521</td>
<td>1529.8</td>
<td>6.20</td>
</tr>
<tr>
<td>8*</td>
<td>Power law</td>
<td>0.80</td>
<td>1.28</td>
<td>0.46</td>
<td>0.073</td>
<td>0.16</td>
<td>-15.948</td>
<td>-18.729</td>
<td>1521</td>
<td>1518.0</td>
<td>6.31</td>
</tr>
<tr>
<td>9*</td>
<td>Gaussian</td>
<td>0.77</td>
<td>1.28</td>
<td>0.49</td>
<td>0.078</td>
<td>0.51</td>
<td>-16.013</td>
<td>-18.729</td>
<td>1521</td>
<td>1507.8</td>
<td>6.04</td>
</tr>
<tr>
<td>10*</td>
<td>Power law</td>
<td>0.71</td>
<td>1.20</td>
<td>0.60</td>
<td>0.057</td>
<td>0.25</td>
<td>-16.008</td>
<td>-18.729</td>
<td>1521</td>
<td>1586.9</td>
<td>6.06</td>
</tr>
<tr>
<td>11*</td>
<td>Gaussian</td>
<td>0.83</td>
<td>1.28</td>
<td>0.36</td>
<td>0.079</td>
<td>0.74</td>
<td>-15.980</td>
<td>-18.729</td>
<td>1521</td>
<td>1496.9</td>
<td>6.17</td>
</tr>
<tr>
<td>12*</td>
<td>Power law</td>
<td>0.82</td>
<td>1.27</td>
<td>0.41</td>
<td>0.074</td>
<td>0.11</td>
<td>-15.945</td>
<td>-18.729</td>
<td>1521</td>
<td>1522.1</td>
<td>6.32</td>
</tr>
<tr>
<td>13**</td>
<td>Gaussian</td>
<td>0.84</td>
<td>1.27</td>
<td>0.34</td>
<td>0.081</td>
<td>0.75</td>
<td>-15.975</td>
<td>-18.729</td>
<td>1521</td>
<td>1520.7</td>
<td>6.19</td>
</tr>
<tr>
<td>14**</td>
<td>Power law</td>
<td>0.81</td>
<td>1.28</td>
<td>0.42</td>
<td>0.074</td>
<td>0.10</td>
<td>-15.943</td>
<td>-18.729</td>
<td>1521</td>
<td>1513.6</td>
<td>6.34</td>
</tr>
</tbody>
</table>

*Input catalog: ANSS catalog 1/1/1981 to 4/1/2009 in the CSEP collection region (167,528 earthquakes \( M \geq 2 \)). Target catalog: ANSS catalog 1/1/1986 to 4/1/2009 in the CSEP testing region (1521 earthquakes for \( M_{min} = 3.95 \)). Spatial background model is the long-term density estimated for the time-independent forecast. Reference model: time-independent forecast with average number \( \mu_{T1} = 0.18 \) of daily \( M \geq 3.95 \).

1Using the corrective term \( \rho^* \).

2Early aftershock-smoothing kernel for quakes above \( M \geq 5 \) (default \( M \geq 5.5 \)).

3Early aftershock-smoothing kernel for quakes above \( M \geq 6 \) (default \( M \geq 5.5 \)).

4Early aftershock-smoothing kernel for quakes above \( M \geq \infty \) (default \( M \geq 5.5 \)).

5Early aftershock-smoothing kernel for aftershocks occurring up to \( T = 1 \) day after a large event (default \( T = 2 \) days).

6Early aftershock-smoothing kernel for aftershocks occurring up to \( T = 3 \) days after a large event (default \( T = 2 \) days).
M > 6 events. There seems to be a subtle trade-off effect here. The benefits of the smoothing technique did not change much when we varied the temporal window for smoothing from 1 to 3 days nor when we lowered the magnitude threshold for selecting events from M 6 to M 5 (see models 5 through 9).

We tested whether adding the contribution ρ* (see Correcting for Spatial and Temporal Magnitude Incompleteness) of undetected events Md < M < M0(t) increases the likelihood values: This was not the case over the entire period. However, certain individual sequences were much better fit using this correction, for example, the 1992 Mw 7.3 Landers earthquake sequence (not shown). There seems to be a subtle trade-off effect between modeling sequences and performing well on the days of the actual mainshocks.

Observed and Predicted Number of Events

In Figure 10, we compare the observed (target) earthquakes and the predicted number of events. In Figure 10a, we compare the daily forecasts against the daily observations. Three groups of days can be identified. First, there are days on which the forecasts are low (close to the background rate), and few events occur that match the forecast. The second group corresponds to low forecasts near the ETAS background rate, but to many observed events that are inconsistent with the model. On those days, large earthquakes and triggered sequences occurred without foreshocks on previous days that could have helped the forecast. But the model’s performance was worse than it could be: The CSEP rules for the one-day forecast group allow an update of the forecast only once every 24 hours. The full potential of the ETAS model can only be realized if updates are allowed after every earthquake. The third group close to the ascending solid line (marking the perfect match) shows the potential: these are days of active sequences that started the day before, the information of which allowed the model to reasonably forecast seismicity.

The cumulative forecasts (Fig. 10b) followed the observed number of events relatively well, including the aftershock sequences. But the forecast systematically underestimated individual sequences, for example, the period after the 1992 Mw 7.3 Landers earthquake. We expected this underprediction as the model cannot update after each event. At the same time, the total cumulative number of events was matched well at the end of the entire period. This is because the likelihood penalizes for mismatches in the total number of events. But since the days during which large quakes and their aftershocks occur cannot be accurately forecast because of the updating rules, the parameters are biased: The background rate is overestimated to match the total number of events. These complications are a result of the one-day forecasts that are unnecessary for many model families. At the same time, the ETAS model will remain a poor predictor of strong events if no foreshocks raise the forecast.

How did the ETAS model perform against the time-independent forecast over the course of the entire period? In Figure 11, the daily log-likelihood ratios between the ETAS model and the time-independent forecast are shown. The largest ratios occurred on or just after large M > 6 earthquakes, which we marked by vertical dashed lines. The daily log-likelihood ratios are also indexed by the number of observed events on each day. The performance of the two models was very similar for days on which no earthquakes occur. The differences increased with the number of observed events. For instance, on the day of the 1999 Hector Mine earthquake, many earthquakes occurred, and the ETAS model strongly outperformed the Poisson model forecast. Several small earthquakes occurred before this large event, thereby locally increasing the ETAS forecast with respect to the Poisson model (see the following discussion of Fig. 13).

On days when the ETAS rate decayed to its background level, the time-independent forecast had a higher rate, so that

![Figure 10](image-url)

**Figure 10.** (a) Comparison of the expected daily number of earthquakes M ≥ 3.95 with the number of observed earthquakes (circles). The solid horizontal line denotes the background rate of the ETAS model, and the solid ascending line would mark a perfect match. (b) Comparison of the cumulative number of observed events and the ETAS forecast.
if an earthquake occurred, the Poisson model beat the ETAS forecast. But the likelihood ratio on those days was never very large, because the background rate of the ETAS model is not much smaller than the Poisson rate. Therefore, the ETAS model is to some extent guarded against surprise events by its background rate, while making vastly better forecasts during active sequences.

Observed and Modeled Temporal Distribution of Aftershocks

In Figure 12, we compare the observed number of events with ETAS forecasts from January 1992 until August 1994, along with the daily probability gain per quake (equation 9), which normalizes the likelihood ratio by the number of observed events per day. The figure shows how the ETAS forecast tracked the observed seismicity during the aftershock sequences of this particularly active period. The largest gains correspond to the most active days, during which individual gains were as large as $G \sim 10^4$ per day per earthquake.

To get an even more detailed picture of the ETAS model’s performance in individual aftershock sequences, we compared the forecast and the observations just before and after the four major earthquakes in Figure 13. The four earthquakes illustrate two successes and two failures of the ETAS model compared with the time-independent forecast. The 1986 $M_w$ 6.4 Chalfant earthquake occurred when the ETAS forecast was higher than the Poisson forecast, as even on the day before the event the ETAS model outperformed the time-independent forecast. On the day of the event, the model underestimated the number of events because when the forecast was made, the large mainshock had not yet occurred. In the following days, the ETAS model forecast the number of events better than the Poisson forecast, as measured by the large gains. The 1992 $M_w$ 7.3 Landers earthquake sequence shows similar properties, except that the ETAS model rate was slightly below the Poisson forecast on the day of the shock so that the gain was slightly below 1. The 1994 $M_w$ 6.7 Northridge earthquake was also not well forecast by the ETAS model, its rate having sunk below the Poisson forecast. Nevertheless, the gains during the aftershock sequence are substantial. The 1999 $M_w$ 7.1 Hector Mine earthquake, on the other hand, is a success story for the ETAS model. Although its total rate, summed over the region, is smaller than the Poisson forecast (as can be seen in Fig. 13), a few small quakes locally increased the ETAS rate above the Poisson forecast, so that the ETAS model outperformed the Poisson model on the day of the Hector Mine earthquake.

Figure 13 can also be used to judge the fit of the ETAS model aftershock forecasts with the actual observed events. The fluctuations in the number of observed aftershocks are clearly larger than the (mean) ETAS model forecast. Even 95% confidence bounds of the Poisson distribution around the mean forecast cannot enclose the observations. So while the likelihood ratio and the probability gain registered an immense improvement of the ETAS model over the time-independent model, we did not test whether the observations are consistent with the ETAS model, as we did for the long-term forecast discussed previously. If we were to apply daily the current CSEP number test (Schorlemmer et al., 2007), which assumes a Poisson uncertainty in the number of events, the model would be rejected on many days. However, at this point it is clear that the original RELM tests are inappropriate for one-day forecasts. Rather, a first step in the right direction would be a model-dependent, simulated likelihood distribution against which the observations are counted (Werner and Sornette, 2008). Because of the computational complexity, however, we need to leave this extension to future work.
Observed and Modeled Spatiotemporal Distribution of Aftershocks

In Figure 14, we compare the observed events with the spatiotemporal forecast around the 1992 $M_w 7.3$ Landers earthquake. We show forecasts for the day before the mainshock (27 June 1992), for the day of the mainshock (28 June 1992), and for the two subsequent days (29–30 June 1992). Also shown are all small events $M \geq 2$ that occurred in the previous week and contributed substantially to the ETAS model forecast, along with the target $M \geq 3$: 95 earthquakes of each day. We observed large differences between the Gaussian and power-law kernels on the days after the mainshock: the power-law kernel has a larger and smoother spatial extent than the Gaussian kernel (see fig. 7 of Helmstetter et al., 2007). The power-law kernel forecasts a high rate in many bins in which no events occur. But on the other hand, it better predicted remote targets in bins unaffected by a Gaussian forecast. We preferred the Gaussian kernel because the estimates for the parameter $f_d$ seemed more reasonable.

The present model was included in a study by Woessner et al. (2011) of the predictive skill of a large suite of statistical and physics-based models on the aftershock sequence of the 1992 $M_w 7.3$ Landers, California, earthquake. The authors found that the statistical models substantially outperformed the models based on static-stress transfer and rate-state friction, and that the present ETAS model provided the best spatial forecasts both among the statistical models and overall.

Time-Dependent Next-Day $M \geq 3.95$ Forecasts Based on the ETAS Model

An example of an ETAS model forecast for 12 February 2008 (using model 1 in Table 1) is shown in Figure 15, along with a comparison to a time-independent $M \geq 3.95$ forecast based on the long-term forecast described previously. On that day, four earthquakes $M \geq 3.95$ occurred in roughly the same location in Baja California, Mexico, as part of a swarm of about 10 earthquakes during a two-week period. Because on previous days several earthquakes had occurred, the ETAS model rate is locally higher than the time-independent forecast by a factor of almost 1000. Other differences between the ETAS model and the background model are less pronounced and mainly due to small, recent events that increased the ETAS model rate.
For the time-dependent forecasts, we used the same magnitude distribution as for the time-independent forecast, including the same corner magnitude $M_c = 8.0$.

**Discussion and Conclusions**

We presented two models that were modified from Helmstetter et al. (2006) and Helmstetter et al. (2007) for estimating the probabilities of future earthquakes in California. The time-independent model uses an optimized adaptive kernel to smooth the locations of small $M \geq 2$ declustered earthquakes. In retrospective tests, we found that including the locations of small earthquakes always help forecast $M \geq 3.5$ earthquakes. The evidence is less clear for larger $M \geq 4.95$ target earthquakes, although during two out of four target periods, small earthquakes do increase the skill of the forecast. It is possible that the statistical fluctuations due to the smaller (target) sample size mask the benefit of smoothing small quakes. Larger, preferably global data sets might help solidify the evidence. Nonetheless, the numerous small earthquakes help map out active fault structures that may or may not be present in available geologic maps. The resulting forecasts are of much higher resolution (specificity) than similar forecasts, a characteristic that may be important for the model’s encouraging performance in prospective tests.

We made several improvements to the long-term model by Helmstetter et al. (2007), including testing several model choices in more detail, fixing minor mistakes in the associated code, recalibrating the model on up-to-date data, and improving the procedure for estimating the completeness magnitude. However, there remains much room for improvements. For example, we are still unsatisfied with the completeness magnitude estimation. In the future, we would like to use a more robust method. Another area for improvement involves the use of anisotropic kernels to smooth seismicity, although the method to smooth the locations of early aftershocks to forecast those of later ones seems to perform quite well. Finally, the earthquake catalog we used is relatively short (30 years), compared with average recurrence times of very large earthquakes and the crust’s memory. In the future, we would like to find a trade-off between high-quality recent data and lower quality older data.

The long-term forecast was first cast as a time-independent Poisson process, but in retrospective tests we found that the number of earthquakes in five-year intervals is not well-described a Poisson process. We therefore also used a negative binomial distribution to forecast the number of earthquakes, and this forecast could not be rejected by retrospective consistency tests. For these tests, we modified the L-test by conditioning the simulated likelihood values on the number of observed events, thereby increasing the test’s sensitivity to the space and magnitude dimension of the forecast and less dependent on the number of observed events. The conditional L-test could replace the current L-test in RELM and CSEP experiments.

But while we modified the variance of the total number distribution of the long-term model, we did not change the variances in the rates of each individual space-magnitude bin. This inconsistency can only be solved completely by replacing the time-independent Poisson process by a time-dependent process of earthquake clustering. Calculating improved five-year forecasts will likely involve simulations of (or approximations to the distributions of) branching
processes. The time-dependent ETAS model we presented here is such a branching model, but thus far we solely used it for average next-day forecasts in order to compete in CSEP’s one-day forecast competition. Longer-term simulations of the present ETAS model are an interesting avenue of future work.

For the one-day forecasts, we similarly assumed that the Poisson distribution adequately represents the variability of the number of earthquakes during one day. Werner and Sornette (2008) have shown that the variability is likely to be larger than that. Moreover, the rates in adjacent bins are not independent (see also Lombardi and Marzocchi, 2010). However, the extension to a model-dependent number and likelihood distribution for each bin is left for the future.

Several features distinguish the present implementation from other ETAS flavors, including (1) a method to correct for spatial and temporal magnitude incompleteness, (2) smoothing the locations of early aftershocks to forecast later ones, and (3) using small \( M \geq 2 \) earthquakes to forecast larger \( M \geq 3.95 \) events. A detailed comparison with other short-term models within the CSEP framework will shed light on the importance of these features in prospective tests. The ETAS model forecasts outperformed the time-independent forecast with a probability gain per earthquake of about 6.

We expect the models to perform well, based on the report by Schorlemmer et al. (2010) on the success of the time-independent model by Helmstetter et al. (2007) in the RELM experiment. However, with increasing time span and the occurrence of a very large event, smoother models based on tectonic, geological, and geodetic data might work better. Determining the reasons for the timing of such a change will help us build better earthquake models.

Data and Resources

We used the Advanced National Seismic System (ANSS) earthquake catalog made publicly available by the Northern California Earthquake Data Center at http://www.ncedc.org/anss/catalog-search.html (last accessed April 2010) in the period from 1 January 1981 until 1 April 2010 with magnitude \( M \geq 2 \) and in the spatial region defined by the CSEP collection region, defined in table 2 by Schorlemmer and Gerstenberger (2007).

For the calculations in Retrospective Consistency Tests of the Five-Year Forecast, we used the ANSS catalog in the CSEP testing region from 1 January 1932 until 1 April 2010 with magnitude \( M \geq 4.95 \).

Acknowledgments

We thank A. Christophersen and J. Woessner for stimulating discussions. MJW is supported by the EXTREMES project of ETH’s Competence Center Environment and Sustainability (CCES). AH is supported by the European Commission under grant TRIGS-043251, and by the French National Research Agency under grant ASEISMIC. DDJ and YYK appreciate support from the National Science Foundation through grant...

Werner, M. J. (2007). On the fluctuations of seismicity and uncertainties in 
earthquake catalogs: Implications and methods for hypothesis testing, 
*Ph.D. Thesis*, University of California, Los Angeles, California.

seismic rate estimates, forecasts, and predictability experiments, 

Retrospective evaluation of the five-year and ten-year CSEP-Italy 

Wiemer, S. (2000). Introducing probabilistic aftershock hazard mapping, 


earthquake catalogs: examples from Alaska, the western United States, 

catalogues: Estimating the magnitude of completeness and its 

Woessner, J., S. Hainzl, W. Marzocchi, M. J. Werner, A. M. Lombardi, 

Zechar, J. D., M. C. Gerstenberger, and D. Rhoades (2010). Likelihood-
based tests for evaluating space–rate–magnitude earthquake forecasts, 

Zechar, J. D., D. Schorlemmer, M. Liukis, J. Yu, F. Euchner, P. J. Maechling, 
and T. H. Jordan (2010). The collaboratory for the study of earthquake 
predictability perspective on computational earthquake science, 

Zhuang, J., A. Christophersen, M. K. Savage, D. Vere-Jones, Y. Ogata, and 
earthquakes: Their influences on foreshock probabilities, *J. Geophys. 

Zhuang, J., Y. Ogata, and D. Vere-Jones (2002). Stochastic declustering of 

clustering features by using stochastic reconstruction, *J. Geophys. 