Appendix S—Constraining Epidemic Type Aftershock Sequence (ETAS) Parameters from the Uniform California Earthquake Rupture Forecast, Version 3 Catalog and Validating the ETAS Model for Magnitude 6.5 or Greater Earthquakes

By Jeanne L. Hardebeck[1]

Abstract

Operational earthquake forecasting in the Uniform California Earthquake Rupture Forecast, version 3 (UCERF3) model will be implemented using the Epidemic Type Aftershock Sequence (ETAS) model. Parameter values for the ETAS model are determined by fitting that model to the recent instrumental earthquake catalog. A grid search is done, and the log-likelihood is used as a measure of fit to estimate the best ETAS parameter values and their uncertainty, as well as to investigate trade-offs between parameters. For operational earthquake forecasts based on the ETAS model to be useful for emergency response and hazard mitigation purposes, they must accurately forecast the probability of large, potentially damaging earthquakes. Because of the relatively low rate of large earthquakes, the parameters of the ETAS model are from a catalog dominated by small earthquakes. It is, therefore, important to investigate whether the large earthquakes in California are explained adequately by the ETAS model derived from small earthquakes. First, I tested the null hypothesis that the transformed times of the magnitude greater than or equal to 6.5 ($M \geq 6.5$) events are Poissonian, which should be the case if the ETAS model successfully accounts for all earthquake clustering. Next, I studied the distribution of the ETAS intensity at the locations and times of $M \geq 6.5$ earthquakes, and test the null hypothesis that this distribution is the same as for small earthquakes, which would imply that the ETAS model forecasts the large earthquakes as well as it does for the small earthquakes. Finally, I used the ETAS model to decluster the catalog, and to test the null hypothesis that there is no residual clustering of $M \geq 6.5$ earthquakes in the declustered catalog. I cannot reject these null hypotheses, so I conclude that the ETAS model adequately models the occurrence of $M \geq 6.5$ earthquakes. Additionally, I showed that including small earthquakes in the ETAS modeling greatly improves the ability of the model to forecast large earthquakes.

Introduction

Parameter values for the Epidemic Type Aftershock Sequence (ETAS) model are determined by fitting the model to the recent instrumental earthquake catalog. The strategy is to

[1]U.S. Geological Survey.

determine representative parameter values by fitting the entire catalog, rather than attempting to fit exactly any particular aftershock sequences. A grid search over parameter space is used, and the goodness-of-fit function can be used to estimate the uncertainty of the parameters, as well as to investigate any trade-offs between parameters. Statistical tests are used to validate the best-fitting model parameters and to compare them to other ETAS parameters for California that appear in the literature.

If operational earthquake forecasts based on the ETAS model are to be useful for emergency-response and hazard-mitigation purposes, they must accurately forecast the probability of large, potentially damaging earthquakes. Because of the small number of recorded large earthquakes in California, the ETAS model parameters are based primarily on small earthquakes. Therefore, a key question for the usefulness of operational earthquake forecasting in the Uniform California Earthquake Rupture Forecast, version 3 (UCERF3) model, is whether the ETAS model will successfully forecast the large (magnitude greater than or equal to 6.5 [$M \geq 6.5$]) potentially damaging earthquakes, or whether there is additional clustering of the large earthquakes that is not captured by the ETAS model. I tested the null hypothesis that the clustering of larger ($M \geq 6.5$) earthquakes in California is well described by the ETAS model with the parameters derived from the whole catalog. The small number of $M \geq 6.5$ events in California means that the null hypothesis is difficult to disprove, even if false. I, therefore, tested this null hypothesis in several ways.

Data

The California earthquake catalogs used here are based on the UCERF3 earthquake catalog from appendix K (this report). The recent instrumental catalog, beginning in 1993, and complete to $M \geq 2.5$ earthquakes except during the early parts of aftershock sequences, is used to estimate the ETAS parameters. The full modern instrumental catalog, beginning in 1984 and also generally complete to $M \geq 2.5$, also will be used to validate the ETAS parameters. Finite source models for large earthquakes are used when available. The earthquake finite rupture planes are represented as sets of points with 1-km spacing.

Test catalogs of large earthquakes, complete to various lower-threshold magnitudes, are selected using the regional completeness thresholds determined for the UCERF, version 2 (UCERF2) catalog (Felzer, 2008). The first two catalogs are instrumental catalogs, beginning in 1932 and 1957, and spanning the full California spatial region. The rest of the catalogs contain a combination of historical and instrumental events, with minimum magnitudes ranging from $M6.5$ to $M7.6$, and with spatial coverage and beginning time chosen to create the largest possible catalog complete to the given minimum magnitude (table S1).

The b-value of each instrumental catalog is estimated using maximum likelihood, following Utsu (1965) and Aki (1965):

$$b = \frac{N}{\ln(10)\sum_{i=1}^{N}\left(M_i - M_{\min} + 0.5\Delta M\right)} \quad (1)$$

where

N is the number of events,
M_i are the event magnitudes,
M_{min} is the minimum magnitude, and
ΔM is the discretization of the magnitude scale.

The *b*-values for the instrumental catalogs are all approximately 1.0, consistent with the rigorous *b*-value estimates of Felzer (2008).

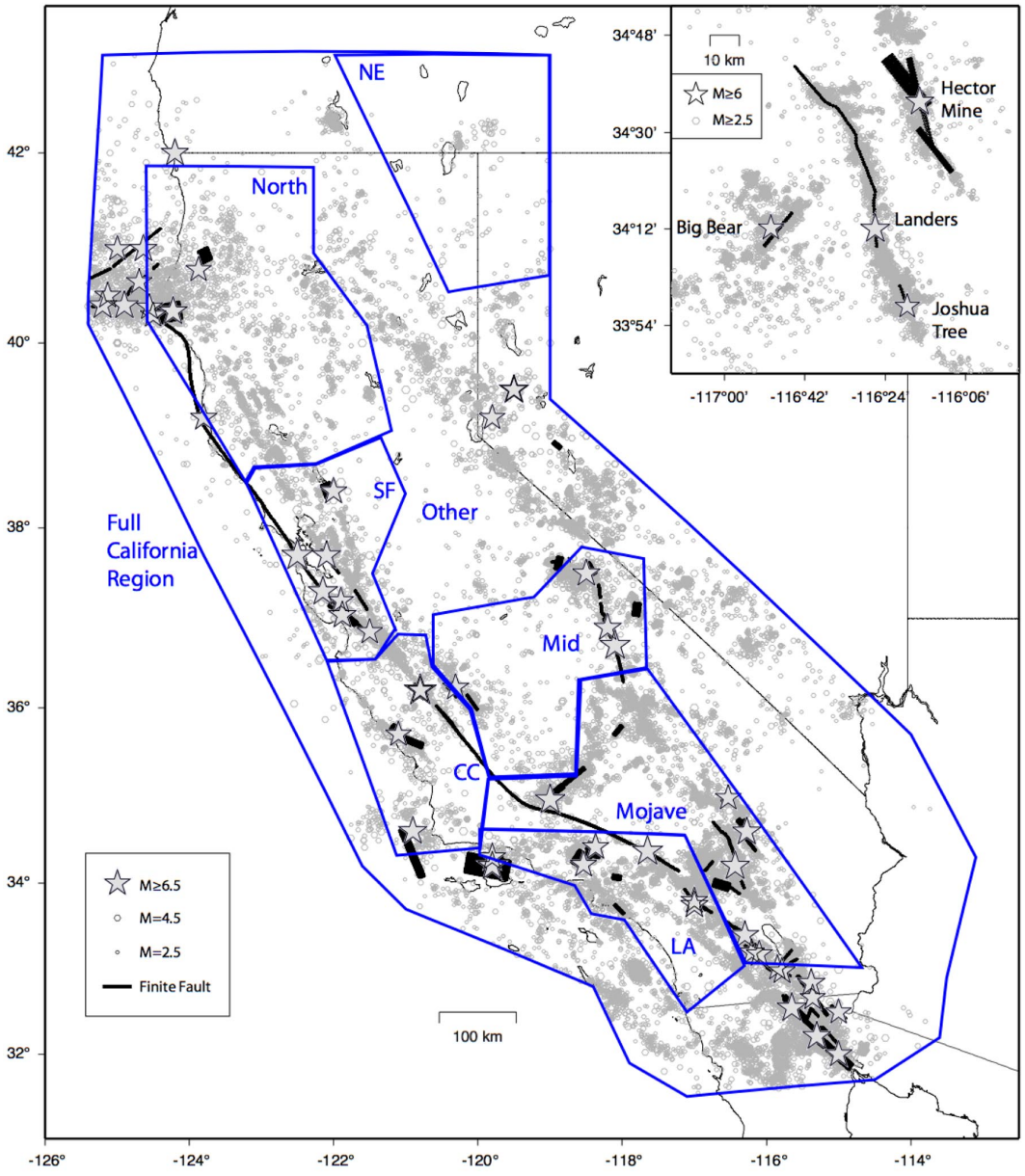

Figure S1. Map showing California with the earthquakes from the Uniform California Earthquake Rupture Forecast, version 3 (UCERF3) catalog. All earthquakes with magnitudes greater than or equal to (*M*≥2.5) are shown as circles, all earthquakes *M*≥6.5 are shown as stars, and black lines indicate finite fault models. The full UCERF3 California region and the UCERF, version 2 (UCERF2) spatial subregions given by Felzer (2008) are outlined and labeled in blue. The inset shows all *M*≥2.5 earthquakes occurring in the Eastern California Shear Zone (ECSZ), from 1992–2011, including four *M*≥6.0 events shown as stars and labeled. CC, central coast region; LA, Los Angeles region; NE, northeast region; SF, San Francisco region. Map base from generic mapping tools (Wessel and Smith, 1998).

Table S1. California earthquake catologs.

[The full Uniform California Earthquake Rupture Forecast, version 3 (UCERF3) catalog includes earthquakes with magnitudes greater than or equal to 2.5, but is not complete to that magnitude. The instrumental catalogs contain instrumentally recorded events only, and the beginning times of the catalogs are selected to produce different levels of completeness. The historical plus instrumental catalogs also contain earthquakes based on historical accounts. The beginning time and spatial coverage of each catalog was chosen to produce the largest catalog complete to the given minimum magnitude M_{min}. Spatial regions (fig. S1) are those given in Felzer (2008). CC, central coast region; LA, Los Angeles region; NE, northeast Region; SF, San Francisco region]

Catalog name	Time (yrs)	M_{min}	M_{max}	N	*b*	Spatial coverage (UCERF2 regions)
Full UCERF3 catalog						
UCERF3	1769-2011	2.5	7.9	58111	---	Full California region
Recent instrumental catalog						
UCERF3_1993	1993-2011	2.5	7.2	31392	0.99	Full California region
Instrumental catalogs						
UCERF3_1984	1984-2011	2.5	7.3	55044	1.02	Full California region
UCERF3_1957	1957-2011	5.3	7.3	174	0.96	Full California region
UCERF3_1932	1932-2011	6.0	7.5	58	1.02	Full California region
Historical + instrumental catalog						
UCERF3_M65	1870-2011	6.5	7.8	19	---	North, SF, CC, LA, Mid
UCERF3_M66	1865-2011	6.6	7.8	17	---	North, SF, CC, LA, Mid
UCERF3_M67	1865-2011	6.7	7.8	14	---	North, SF, CC, LA, Mid
UCERF3_M68	1865-2011	6.8	7.8	10	---	North, SF, CC, LA, Mid
UCERF3_M69	1910-2011	6.9	7.5	10	---	North, SF, CC, LA, Mid, Mojave, Other
UCERF3_M70	1910-2011	7.0	7.5	7	---	North, SF, CC, LA, Mid, Mojave, Other
UCERF3_M71	1885-2011	7.1	7.8	9	---	North, SF, CC, LA, Mid, Mojave, Other
UCERF3_M72	1870-2011	7.2	7.8	8	---	North, SF, CC, LA, Mid, Mojave, Other
UCERF3_M73	1870-2011	7.3	7.8	5	---	North, SF, CC, LA, Mid, Mojave, Other
UCERF3_M74	1855-2011	7.4	7.9	3	---	North, SF, CC, LA, Mojave
UCERF3_M75	1855-2011	7.5	7.9	4	---	North, SF, CC, LA, Mid, Mojave
UCERF3_M76	1855-2011	7.6	7.9	3	---	North, SF, CC, LA, Mid, Mojave

The ETAS Model

The ETAS model assumes that the earthquake rate as a function of space and time is described by the intensity function (Ogata, 1988):

$$\lambda(t,\mathbf{x}) = \lambda_o \mu(\mathbf{x}) + \sum_{\{i:t_i<t\}} \lambda_T\left(t,\mathbf{x} \mid M_i,t_i,\mathbf{x}_i\right), \quad (2)$$

where $\lambda(t,\mathbf{x})$ is the earthquake rate density (that is, number of earthquakes per unit time per unit spatial volume) of $M \geq M_{min}$ earthquakes at any given time, t, and location in space, $\mathbf{x}$; λ_0 is the total background rate; and $\mu(\mathbf{x})$ is the probability density function (PDF) of the spatial density of background earthquakes. The summation is over the rate density of triggered earthquakes owing

to all earthquakes prior to the time of interest, *t*. The rate density of triggered earthquakes owing to a mainshock *i* at time t_i and location $\mathbf{x}_i$ and with magnitude M_i is given by (Ogata, 1988):

$$\lambda_T(t, \mathbf{x} \mid M_i, t_i, \mathbf{x}_i) = k10^{a(M_i - M_{\min})}\left(t - t_i + c\right)^{-p} \mathrm{D}\left(\mathbf{x}, \mathbf{x}_i\right), \quad (3)$$

where D(**x,** $\mathbf{x}_i$) gives the spatial distribution of triggered events. Here the ETAS *a*-value is fixed to the *b*-value already in equation 1, which ensures that Båth's law, that the largest aftershock is on average 1.2-magnitude units less than the mainshock, holds independent of mainshock magnitude (Felzer and others, 2002).

The spatial distribution of triggered earthquakes, in keeping with the UCERF3 ETAS implementation as well as recent observations (Felzer and Brodsky, 2006; Felzer and Kilb, 2009; Marsan and Lengliné, 2010), is designed such that the linear density is of the form:

$$D_L(r) = c_S\left(r + d\right)^{-q}, \quad (4)$$

where $D_L(r)$ is the rate of all events at distance *r* from the mainshock, and c_S is a spatial normalization constant such that the integral of $D_L(r)$ from 0 to the assumed maximum triggering distance, r_{max}, is equal to 1. For $q \neq 1$:

$$c_s = \frac{1-q}{\left(r_{\max} + d\right)^{1-q} - d^{1-q}} \quad (5)$$

and for q=1:

$$c_s = \frac{1}{\ln\left(r_{\max} + d\right) - \ln d}. \quad (6)$$

The spatial density D(**x,**$\mathbf{x}_i$), at small distances where the seismogenic thickness is not important, is, therefore, $D_L(r)$ distributed over the surface of a sphere of radius *r*, so:

$$\mathrm{D}(\mathbf{x}, \mathbf{x}_i) = c_S r^{-2}\left(r + d\right)^{-q} / 4\pi. \quad (7)$$

At distances greater than the seismogenic thickness, *H*, the spatial density D(**x,** $\mathbf{x}_i$) is approximated by distributing the linear density of events $D_L(r)$ over the surface of the sides of a cylinder with radius *r* and height *H*, so:

$$\mathrm{D}(\mathbf{x}, \mathbf{x}_i) = c_S r^{-1}\left(r + d\right)^{-q} / 2\pi H. \quad (8)$$

These functions are equal at $r=H/2$, so this is used as the transition point between these two definitions of D(**x,**$\mathbf{x}_i$). Here, I assume that H=12 km, and r_{max}=1,000 km. If the mainshock, event *i*, has a finite source model in the database, it is represented as N_i point sources. Each point source produces its own contribution to the intensity function, following equation 3, except replacing *k* with k/N_i to retain the total contribution of the earthquake. This conforms to the assumptions made in the UCERF3 ETAS implementation.

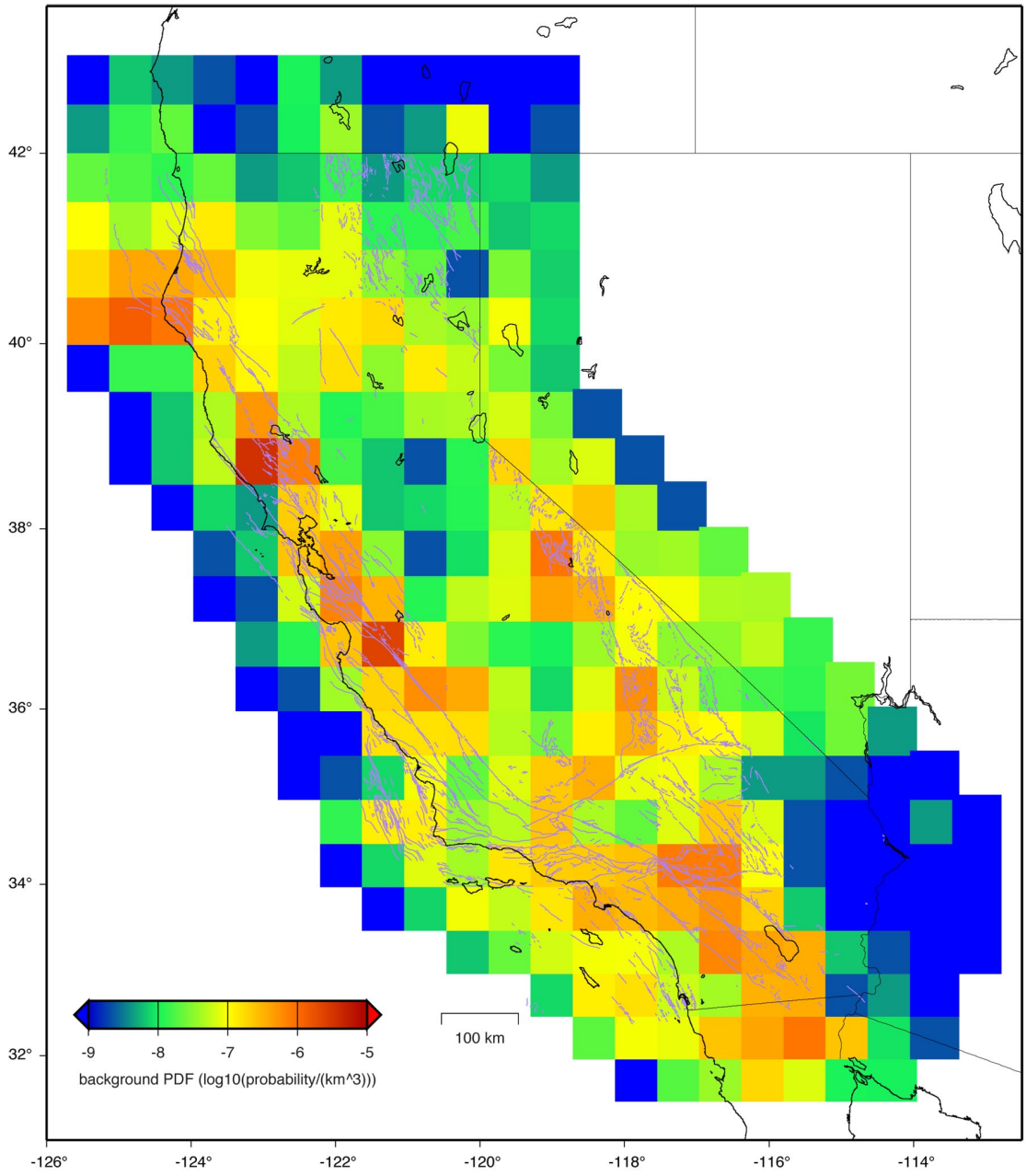

Figure S2. Map showing the probability distribution function for the spatial density of background earthquakes, $\mu(\mathbf{x})$. The California region was divided into 50× 50-km grid cells, and the background rate in each cell was estimated from the instrumental catalog 1984–2011 using the technique of Hainzl and others (2006). The rates were then normalized so that $\mu(\mathbf{x})$ sums to 1 over the whole region. Rates in each cell are assumed to be uniform over a 12-km seismogenic thickness.

This ETAS formulation, therefore, has six parameters: λ_0, k, c, p, d, and q. Additionally, the spatial distribution of background events, $\mu(\mathbf{x})$, must be estimated. A simplification is made by specifying λ_0 and k in terms of the branching ratio, $n=N_{aft}/N$ (for example, Helmstetter and Sornette, 2002), where N_{aft} is the total number of triggered events. Then $\lambda_0=(1-n)*N/T$, where T

is the duration of the catalog. The productivity parameter k is related to the branching ratio because it controls the number of aftershocks triggered by each earthquake. It can be given as a function of n and the other ETAS parameters, as shown in section, "ETAS Productivity Parameter k." The spatial distribution $\mu(\mathbf{x})$ of background events is estimated by dividing the California region into 50×50-km grid cells, and estimating the background rate in each cell using the modern instrumental catalog 1984–2011 and the technique of Hainzl and others (2006). Rates are assumed to be uniform with depth. If a cell contains fewer than four events, the full event rate is used as the background rate. For cells containing no events, the background rate is assumed to be 0.5 event for the duration of the catalog over all these cells. The rates of all cells were then normalized so that the integration of $\mu(\mathbf{x})$ over the whole region sums to 1 (fig. S2). Tests using a spatially invariant $\mu(\mathbf{x})$ indicate that the preferred ETAS parameter values are not very sensitive to the choice of $\mu(\mathbf{x})$.

ETAS Productivity Parameter *k*

The number of aftershocks triggered by all earthquakes of magnitude M is given by the product of the integrated aftershock rate, $\lambda(t|M)$, with the number of mainshocks of magnitude M, $N(M)$. The total number of aftershocks, N_{aft}, can be determined by integrating $\lambda(t|M)N(M)$ over the full time, T, and magnitude range, M_{min}-M_{max}, of the catalog:

$$\begin{aligned} N_{aft} &= \int_{t=0}^{T} \int_{M=M_{\min}}^{M_{\max}} \lambda(t \mid M) N(M) dM dt \\ &= \int_{t=0}^{T} \int_{M=M_{\min}}^{M_{\max}} k\left(t+c\right)^{-p} 10^{b(M-M_{\min})} 10^{a-b(M-M_{\min})} dM dt \\ &= k10^{a} \int_{t=0}^{T} \left(t+c\right)^{-p} dt \int_{M=M_{\min}}^{M_{\max}} dM \end{aligned} \tag{9}$$

The constant a in the Gutenberg-Richter distribution (Gutenberg and Richter, 1944) can be calculated by integrating to find an expression for the total number of events, N:

$$N = \int_{M=M_{\min}}^{M_{\max}} N(M) dM = 10^{a} \int_{M=M_{\min}}^{M_{\max}} 10^{-b(M-M_{\min})} dM = -\frac{10^{a}}{b\ln(10)} \left[10^{-b(M_{\max}-M_{\min})} - 1\right]. \tag{10}$$

It follows that:

$$10^{a} = \frac{Nb\ln(10)}{\left[1-10^{-b(M_{\max}-M_{\min})}\right]}. \tag{11}$$

Substituting into equation (9):

$$N_{aft} = \frac{kNb\ln(10)}{\left[1-10^{-b(M_{\max}-M_{\min})}\right]} \int_{t=0}^{T} (t+c)^{-p} dt \int_{M=M_{\min}}^{M_{\max}} dM \tag{12}$$

So, if $p \neq 1$:

$$N_{aft} = \frac{knb\ln(10)\left(M_{\max} - M_{\min}\right)\left[\left(T+c\right)^{(1-p)} - c^{(1-p)}\right]}{\left(1-p\right)\left[1-10^{-b\left(M_{\max}-M_{\min}\right)}\right]}, \quad (13)$$

so, then, with the branching ratio defined as $n=N_{aft}/N$:

$$k = \frac{n*\left(1-p\right)\left[1-10^{-b\left(M_{\max}-M_{\min}\right)}\right]}{b\ln(10)\left(M_{\max} - M_{\min}\right)\left[\left(T+c\right)^{(1-p)} - c^{(1-p)}\right]}. \quad (14)$$

Or, if p=1:

$$N_{aft} = \frac{knb\ln(10)\left(M_{\max} - M_{\min}\right)\left[\ln\left(T+c\right) - \ln c\right]}{\left[1-10^{-b\left(M_{\max}-M_{\min}\right)}\right]}, \quad (15)$$

so, then:

$$k = \frac{n*\left[1-10^{-b\left(M_{\max}-M_{\min}\right)}\right]}{b\ln(10)\left(M_{\max} - M_{\min}\right)\left[\ln\left(T+c\right) - \ln c\right]}. \quad (16)$$

ETAS Parameters from the Recent Instrumental Catalog

Parameter Estimation

A grid search is performed over the ETAS parameter space to determine the best-fitting set of parameters for the recent instrumental catalog, as well as to explore the parameter space and investigate trade-offs between parameters. The goodness of fit is described by the log likelihood of the model:

$$\ln L(n,c,p,d,q) = \sum_{i=1}^{N} \ln \lambda\left(t_i, \mathbf{x}_i\right) - \int_V \int_0^T \lambda\left(t, \mathbf{x}\right) dtdV, \quad (17)$$

where V indicates the spatial volume of the catalog, and t_i and $\mathbf{x}_i$ represent the time and hypocenter of the ith earthquake in the catalog.

The grid search is performed in two steps: (1) a coarse search over a wide range of parameter space, and then (2) a finer search in the vicinity of the parameter values producing the maximum log-likelihood from the coarse search. The coarse search covers the parameter ranges shown in figure S3, with steps of 0.1 in p and n, steps of 0.5 in $\log_{10}(c)$, steps of 0.4 in q, and steps of 0.2 in $\log_{10}(d)$, for 141,570 parameter combinations. The fine grid covers 0.1 of the range of the coarse search for each parameter, with 0.1 of the spacing, giving a similar number of parameter combinations. The substantial computation time was reduced by running a pre-processing step to identify which earthquakes (j=1…i-1) do not contribute significantly to the ETAS intensity at the time and location of event i. All events j that contribute less than the background intensity for event i, using a representative set of ETAS parameter values, are flagged and not used in the computation of the intensity for event i. Several subsets of the data were fit both with and without the flagged events, and the results verify that this simplification does not have a significant effect on the preferred ETAS parameters.

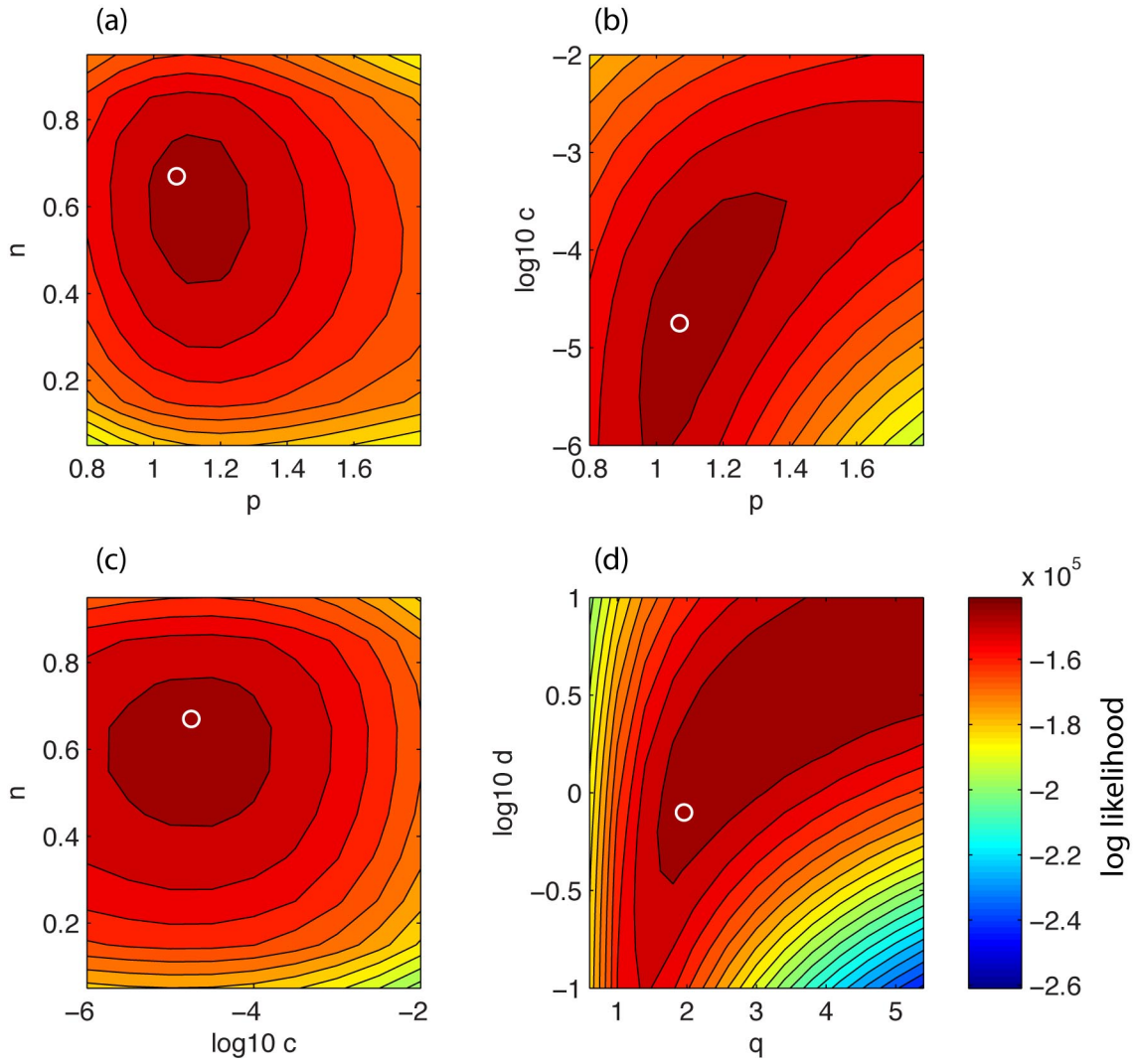

Figure S3. The log-likelihood for Epidemic Type Aftershock Sequence models with the given parameter values, for the recent instrumental catalog beginning in 1993, with all other parameters fixed to the maximum-likelihood values (table S2). White circle indicates maximum likelihood parameter values. (a) The branching ratio, *n*, versus the direct *p*-value; (b) The *c*-value versus the direct p-value; (c) The branching ratio, *n*, versus the c-value; (d) Distance decay *d*-value versus distance decay *q*-value.

The log-likelihood for a range of ETAS parameter values for the recent instrumental catalog (1993–2011, with $M \geq 2.5$) is shown in figure S3, and the preferred parameter values are given in table S2. The log-likelihood function is flat in the region of the best-fit parameter values, indicating substantial uncertainty. Two parameter trade-offs also are shown in the figures: one between the Omori law *p*-value and *c*-value, and one between the *q*-value and *d*-value of the spatial kernel.

Table S2. Preferred Epidemic Type Aftershock Sequence parameters from the recent instrumental catalog, 1993–2011, for earthquakes with magnitudes greater than or equal to 2.5.

[Uncertainty range gives parameter values with log-likelihoods within 5 percent of the maximum log-likelihood, over a reasonable range of values of the other parameters]

Parameter	Value	Range
b	0.99	
n	0.67	0.45 – 0.75
p	1.07	1.0 – 1.4
c (years)	1.78E-05	1.00E-6 – 3.16E-4
q	1.96	>1.8
d (km)	0.79	>0.63
k ((years)^(p-1))	2.84E-03	3.79E-4 – 4.97E-3

The best-fit p-value is 1.07, with values in the range 1.0–1.4 having log-likelihoods within 5 percent of the maximum log-likelihood (over a reasonable range of values for the other parameters). The best-fit p-value is lower than the preferred p=1.37 of Felzer and others (2003) for California, but the range encompasses this value. (This direct p-value describes only the direct aftershocks, as distinct from the apparent p-value for a full aftershock sequence including secondary triggering.) The preferred c-value is about 10 minutes, and values ranging from about30 seconds to about 3 hours have log-likelihoods within 5 percent, with large c-values associated with large p-values. This range of c-values is reasonable if we assume that c represents detection limitations during the early part of an aftershock sequence (for example, Enescu and others, 2009).

The spatial parameters q and d are more poorly constrained and trade off with each other, as shown in figure S3*d*. The maximum likelihood values are q=1.96 and d=0.79 km. Minimum values of q=1.8 and d=0.63 km have log-likelihoods within 5 percent of the maximum log-likelihood. However, the upper limits of both parameters are not bounded by the grid search, which considered values as high as d=10 km and q=5.4, which are unreasonably large values. The best-fit value of q is consistent with the previously observed range of q between 1.3 and 2.1 (Felzer and Brodsky, 2006; Felzer and Kilb, 2009; Marsan and Lengliné, 2010), and the best-fit value of d is on the order of the earthquake location error. The best-fit values for q and d, therefore, appear reasonable, despite the strong parameter trade-offs and the lack of upper bounds.

The best-fit branching ratio is n=0.67, ranging from 0.45 to 0.75 with log-likelihoods within 5 percent. This is similar to previous estimates that the southern California catalog, with M_{min}=2.8–3.8, contains about two-thirds triggered events (for example, Gardner and Knopoff, 1974). However, this branching ratio cannot be used for catalogs with different minimum magnitudes. Sornette and Werner (2005) have shown that catalogs with a larger minimum magnitude will exhibit an apparent smaller fraction of triggered earthquakes. Sornette and Werner (2005) show that the apparent aftershock fraction, n_{app}, is related to the true aftershock fraction, n_{true}, by:

$$n_{app} = n_{true}\left(\frac{M_{\max} - M_{\min}}{M_{\max} - M_0}\right), \tag{18}$$

where M_{min} is the lower magnitude bound of the catalog, M_{max} is the largest earthquake, and M_0 is the minimum magnitude earthquake capable of triggering. Without knowing n_{true} or M_0, but

assuming that they and M_{max} are constant, we can show that the ratio of n_{app} to the magnitude range (M_{max}-M_{min}) is a constant:

$$\frac{n_{app}}{M_{\max} - M_{\min}} = \frac{n_{true}}{M_{\max} - M_0} = c_n , \tag{19}$$

Equation 19 can be used to estimate the apparent n for a catalog with any M_{min}, assuming that the constant c_n has been constrained. The instrumental catalogs give c_n=n/(M_{max}-M_{min} =0.67/(7.9-2.5)= 0.12, assuming that M_{max}=7.9 for California. So, to estimate the expected branching parameter n for catalogs with M_{min}≠2.5, I used:

$$n = 0.12\left(7.9 - M_{\min}\right). \tag{20}$$

Parameter Validation

The ETAS parameters are validated on the full modern instrumental catalog, 1984–2011, M_{min}=2.5. I checked for any unmodeled temporal clustering by comparing the observed and predicted cumulative number of earthquakes through time. The cumulative number of observed M≥2.5 earthquakes is similar to the cumulative number of M≥2.5 earthquakes predicted by the ETAS model using the best-fit parameters for the catalog starting in 1993 (fig. S4). Between 1984 and 1993, the fit also is reasonable, except that the catalog appears to be missing a large number of aftershocks of the April 25, 1992 M7.2 Cape Mendocino earthquake. This is not likely to be a problem with the productivity parameters, given that the ETAS model matches the aftershock sequences of similar-sized events, including the June 28, 1992 M7.3 Landers earthquake that was not used in the parameter fitting.

Super-thinning (Clements and others, 2011) is used to check for any unmodeled spatial clustering in the 1984–2011 catalog. The catalog is first thinned by retaining each earthquake with a probability P=min{1,r_0/λ(t,**x**)}, where r_0 is the earthquake rate (number of earthquakes per unit time per unit spatial volume) such that a constant earthquake rate r_0 over the California region would produce on average the same total number of earthquakes as the number in the observed catalog. Next, superposition is used to add "earthquakes" in low-probability regions with probability P=max{0,r_0-λ(t,**x**)}. The resulting catalog should be spatially homogeneous if the ETAS model is correct. This is verified by computing the L-function, L(r)=sqrt(K(r)/π), where K(r) is the average number of events within 2D (epicentral) distance r of a given event, divided by the 2D average earthquake rate, r_0*H*T. The function L(r)-r should be 0 to within its confidence bounds, as given by Clements and others (2011), and generally is for r<40 km (fig. S5). For r≤6, L(r)-r is slightly greater than the 95-percent confidence bounds. The small amount of unmodeled clustering on these small-length scales may be attributable to earthquake location uncertainty, approximations in the finite-source models of large earthquakes, and the lack of finite-source models for intermediate-sized events. The deficiency of earthquakes at r>40 km persists to at least 1,000 km, suggesting an edge effect rather than a real deficiency over some particular distance range.

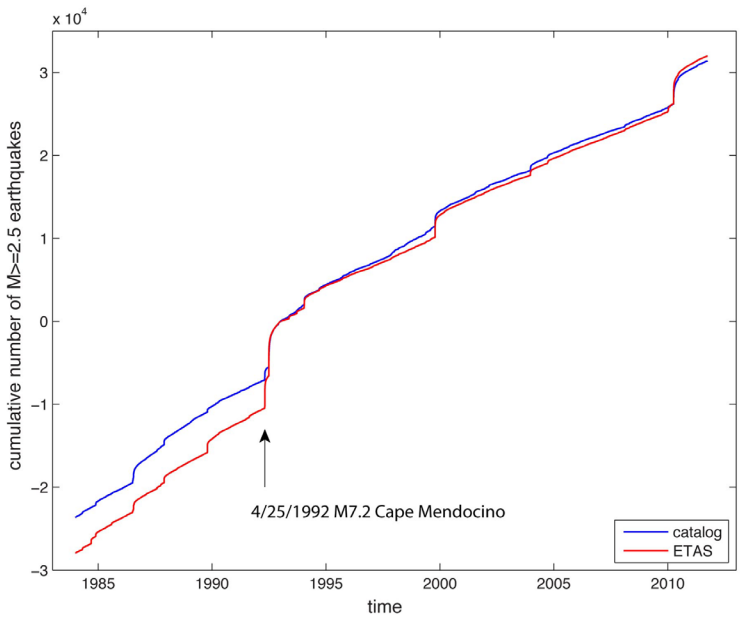

Figure S4. Graph showing cumulative number of earthquakes with magnitudes greater than or equal to 2.5 ($M \geq 2.5$) (blue), compared to the cumulative number predicted by the Epidemic Type Aftershock Sequence model (red) using the preferred parameters (table S2.) The cumulative number was set to 0 on January 1, 1993.

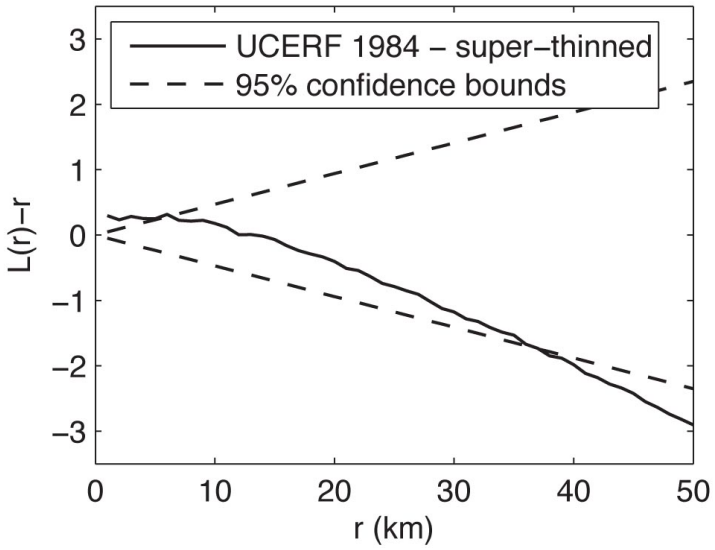

Figure S5. Graph showing L-function for one realization of super-thinning (Clements and others, 2011) of the 1984–2011 catalog. The catalog is thinned by retaining each earthquake with a probability $P=\min\{1,r_0/\lambda(t,\mathbf{x})\}$, and then using superposition to add "earthquakes" with probability $P=\max\{0,r_0-\lambda(t,\mathbf{x})\}$. The L-function as a function of distance, r, is $L(r)=\mathrm{sqrt}(K(r)/\pi)$, where $K(r)$ is the average number of events within 2D (epicentral) distance r of a given event, divided by the 2D average earthquake rate, r_0*H*T. $L(r)-r$ (solid line) and the 95-percent confidence bounds (dashed lines) are shown. The standard deviation of $K(r)$ is $2\pi r^2 A/(r_0 AHT)^2$, following Clements and others (2011), where A is the area of the California region.

Next, the spatial goodness-of-fit of the preferred ETAS parameters was compared with those from two other ETAS models for California (Hardebeck and others, 2008; Werner and others, 2011). The ETAS parameters from those models and a uniform spatial distribution of background events were used in order to compare only the ETAS clustering parameters. The Hardebeck and others (2008) ETAS model is parameterized similarly to the parameterization used here, with a=b=1.0, p=1.34, c=0.095 days, q=1.3, q=0.01 km, n=0.6, and k=0.008 (where n and k are specified separately, with n determining the background rate and k the productivity.) Werner and others (2011) allow a and b to differ, and calculated that a=0.8, b=1.0, p=1.27, and c=0.035 days. They determined that n=0.56 for M≥3.95 earthquakes, which translates to n=0.76 for M≥2.5 earthquakes, using equation 19. Their preferred spatial kernel is a Gaussian distribution:

$$\mathrm{D}(\mathbf{x},\mathbf{x}_i) = c_S \exp\left(-r^2 \Big/ 2d^2\right). \tag{21}$$

Werner and others (2011) allowed d to depend on magnitude in order to account for finite ruptures, which is not required here because the finite ruptures of large earthquakes are included explicitly. Instead, I use a value of d representative of a M=3 earthquake, d=0.71 km. The values of k reported by Werner and others (2011) are calculated to optimize 1-day forecasts, and, therefore, include the effects of secondary triggering during that day, so they cannot be used directly here. Instead, k is calculated from the branching ratio n and the other ETAS parameters, as described in section "ETAS Productivity Parameter k."

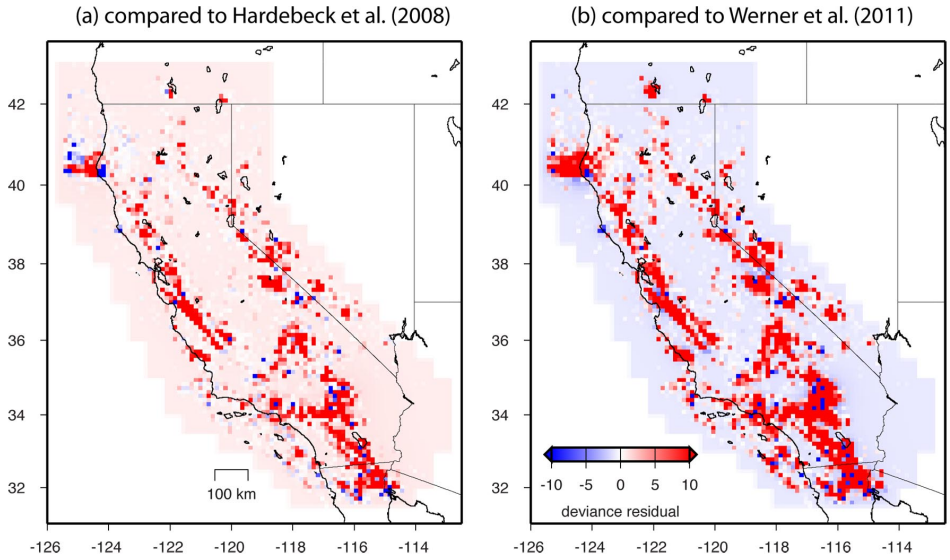

Figure S6. Maps showing deviance residuals for the preferred Epidemic Type Aftershock Sequence (ETAS) parameters, compared to the ETAS parameters (a) from Hardebeck and others (2008), and (b) modified from Werner and others (2011). The region was divided into 12×12-km bins, equivalent to the seismogenic thickness, and the difference between the log-likelihoods of the two models is found for each bin, using the 1984–2011 catalog. Red indicates that the preferred parameters from this study have a greater log-likelihood than the other parameters, while blue indicates that the other parameters have a greater log-likelihood than the preferred parameters. The sum of the deviance residuals is 55,107 when compared to the Hardebeck and others (2008) parameters, and 80,671 when compared to the modified Werner and others (2011) parameters.

The ETAS parameters calculated here are compared with the parameters of Hardebeck and others (2008) and the modified parameters from Werner and others (2011) using the deviance residuals (Wong and Schoenberg, 2009; Clements and others, 2011). The California region is divided into 12×12-km bins, and the difference between the log-likelihoods of the two models being compared is determined for each bin, using the 1984–2011 catalog. The best-fitting parameters of this study outperform both the Hardebeck and others (2008) parameters (fig. S6a), and the modified Werner and others (2011) parameters (fig. S6b). The bins where the Hardebeck and others (2008) or the modified Werner and others (2011) parameters greatly outperform the best-fit parameters are scattered throughout California, and generally are not spatially coherent over more than a few bins. This suggests that there are no particular tectonic regions for which the preferred ETAS parameter values are inappropriate, at least as compared to other published parameter values.

Large Earthquake Clustering

Next I address whether the ETAS model explains the clustering of large ($M \geq 6.5$) earthquakes, or whether additional large-earthquake clustering is present. The most important goal of operational earthquake forecasting is to correctly forecast the probability of large, potentially damaging earthquakes. Therefore, it must be established whether or not the ETAS model adequately explains the clustering of past $M \geq 6.5$ events.

I used three approaches to test for unmodeled clustering of $M \geq 6.5$ earthquakes. First, I tested the null hypothesis that the transformed times of the $M \geq 6.5$ events are Poissonian, which should be the case if the ETAS model successfully accounts for all earthquake clustering. Next, I studied the distribution of the ETAS intensity, $\lambda(t,\mathbf{x})$, at the locations and times of $M \geq 6.5$ earthquakes, and tested the null hypothesis that this distribution is the same as for small earthquakes, which would imply that the ETAS model forecasts the large earthquakes as well as it does for small earthquakes. Finally, I used the ETAS model to decluster the catalog, and tested the null hypothesis that there is no residual clustering of $M \geq 6.5$ earthquakes in the declustered catalog.

Temporal Clustering: Transformed Time

To test the success of the ETAS model in explaining the temporal earthquake clustering, I computed the transformed time (Ogata, 1988) of each earthquake, tt_i, which is the expected number of earthquakes to have occurred by time t_i, given the ETAS model:

$$tt_i = \int_V \int_0^{t_i} \lambda\left(t, \mathbf{x}\right) dt dV \, . \tag{22}$$

If all of the temporal clustering is explained by the ETAS model, the earthquakes should occur with uniform probability in transformed time, so the transformed times should be a stationary Poisson process.

Three tests are performed on the normalized interevent transformed times to assess whether the transformed times are consistent with a stationary Poisson process. The normalized interevent transformed times are:

$$\tau_i = (tt_i - tt_{i-1}) N / tt_{\max} \, , \tag{23}$$

where N is the number of events and tt_{max} is the maximum transformed time. The first test uses a Kolomogorov-Smirnov (K-S) test to determine if the cumulative distribution of interevent times comes from the cumulative distribution expected for a Poisson process:

$$\mathrm{cdf}(\tau) = 1 - \exp(-\tau). \tag{24}$$

The second test is a runs test, which checks whether there is a temporal trend in the interevent times by testing the number of sequences of consecutive below-median or above-median interevent times against the expected number of sequences for random data. The third test is an autocorrelation test, which determines whether there is a tendency for sequential interevent times to be correlated, indicating clustering. The sequence of interevent times, with the mean removed, is correlated with the same sequence shifted one place. The significance is found by comparing the autocorrelation with the results of a suite of simulations where the interevent times are randomly reordered.

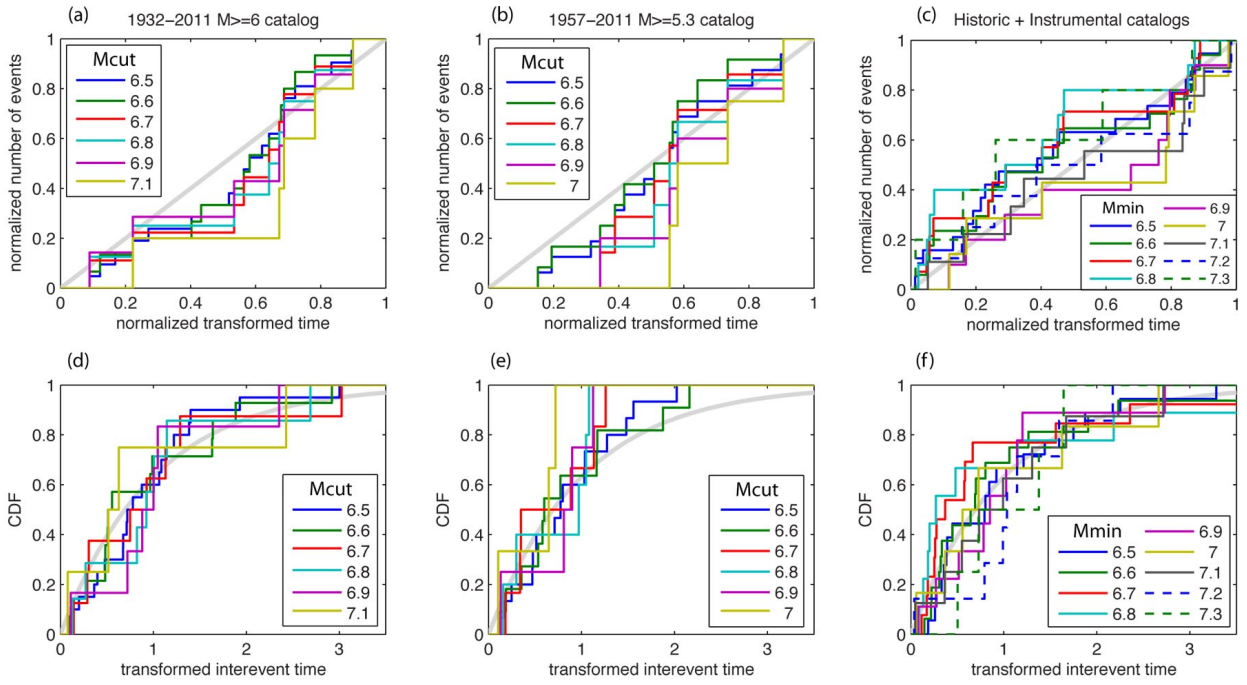

Figure S7. Graphs showing number of earthquakes as a function of transformed time for various catalogs and comparison of cumulative distribution of transformed interevent times for these catalogs. (a), (b), Number of earthquakes as a function of transformed time for subcatalogs of the 1932 and 1957 instrumental catalogs including all earthquakes with magnitudes greater than or equal to cut-off magnitude ($M \geq M_{cut}$) for a range of M_{cut}. Because the subcatalogs have different numbers of events, the numbers of events and transform times have been normalized. One-to-one line shown in gray. (c) Number of earthquakes as a function of transformed time for the historical plus instrumental catalogs with a range of the minimum magnitude (M_{min}). One-to-one line shown in gray. (d), (e), and (f), Comparison of the cumulative distribution of transformed interevent times, for the catalogs in the panels above, with the expected distribution for a Poisson process (gray line.)

First, I considered the $M \geq 6.5$ events in the 1932 and 1957 instrumental catalogs (there are not enough large events in the 1984 or 1993 catalogs for meaningful statistics), and tested whether their transformed times within those catalogs are consistent with a stationary Poisson process (fig. S7). For the parameters b, c, p, d, and q, I used the best-fit values from the recent instrumental catalog (table S2). I estimated the branching ratio n for each catalog using the M_{min} of the catalog and equation 20. For all cut-off magnitudes of M_{cut}=6.5–7.1, the K-S, runs, and autocorrelation tests do not reject, at the 95-percent confidence level, the null hypothesis that the $M \geq M_{cut}$ earthquakes come from a stationary Poisson process in transformed time. This, therefore, fails to reject the null hypothesis that the clustering of these larger events can be explained by the same ETAS model that fits the modern instrumental catalog. However, there are very few $M \geq 6.5$ events in these instrumental catalogs, so it is plausible that the apparent poor fit for the 1957 catalog (fig. S7b), for example, is a real signal that cannot be distinguished from the null hypothesis solely because of the small sample size.

Next, I considered the combined historical and instrumental earthquake catalogs, which provide larger sample sizes of $M \geq 6.5$ events (table S1). I computed the ETAS transformed time, assuming these catalogs are governed by the same spatial-temporal clustering parameters as the instrumental catalogs. For the parameters b, c, p, d, and q, I used the median values from the recent instrumental catalog (table S2). I estimated the branching ratio n for each catalog from equation 20, using the M_{min} of the catalog. For these catalogs, the K-S, runs, and autocorrelation tests do not reject the null hypothesis at 95-percent confidence that the earthquakes come from a stationary Poisson process in transformed time (fig. S7). There is, therefore, no evidence for statistically significant temporal clustering of large earthquakes that is not explained by the ETAS model.

Spatial-Temporal Clustering: ETAS Intensity

A simple test of whether the ETAS model explains the spatial-temporal patterns of large earthquakes as well as it explains the spatial-temporal patterns of small earthquakes is to compare the ETAS intensity function, $\lambda(t,\mathbf{x})$ from equation 2, at the times and locations of large events compared to small events. A high intensity at the time and location of an earthquake implies that its occurrence is well explained by the ETAS spatial-temporal clustering model. Additionally, the intensity at the time and location of an event is an indication of how well the ETAS model could have performed in forecasting that event, as it is the intensity that is used to create forecasts.

I computed intensities for all earthquakes using the full UCERF3 catalog, even though this catalog is incomplete. I computed the intensity from the ETAS model using all earthquakes $M \geq 2.5$, and the preferred ETAS parameters (table S2). The cumulative distribution of intensity for all $M \geq 6.5$ earthquakes is shown in figure S8. These results cannot be compared directly to the intensity distribution for all $M < 6.5$ events, however, because of the incompleteness of the catalog. The detection of $M < 6.5$ events increases with time through the catalog, so these events preferentially sample later times with greater apparent seismicity rates and, therefore, larger overall intensities. To create resampled $M < 6.5$ catalogs that match the size and temporal sampling of the larger event catalogs, I randomly chose one $M < 6.5$ event for each $M \geq 6.5$ event, occurring within 10 years before or after the large earthquake. The observed intensity distribution for the $M \geq 6.5$ events is inside the range of 3,000 realizations of the resampled $M < 6.5$ events. A K-S test shows that the distribution of intensities for the large earthquakes is not statistically distinguishable from the mean of the 3,000 realizations of the small events, at the 95-

percent confidence level. Therefore, we cannot reject the null hypothesis that the distribution of ETAS intensity at the times and locations of the $M \geq 6.5$ earthquakes is the same as for small events. This implies that the ETAS model would have worked to forecast the $M \geq 6.5$ earthquakes during the catalog time period just as well as it would have worked to forecast small events.

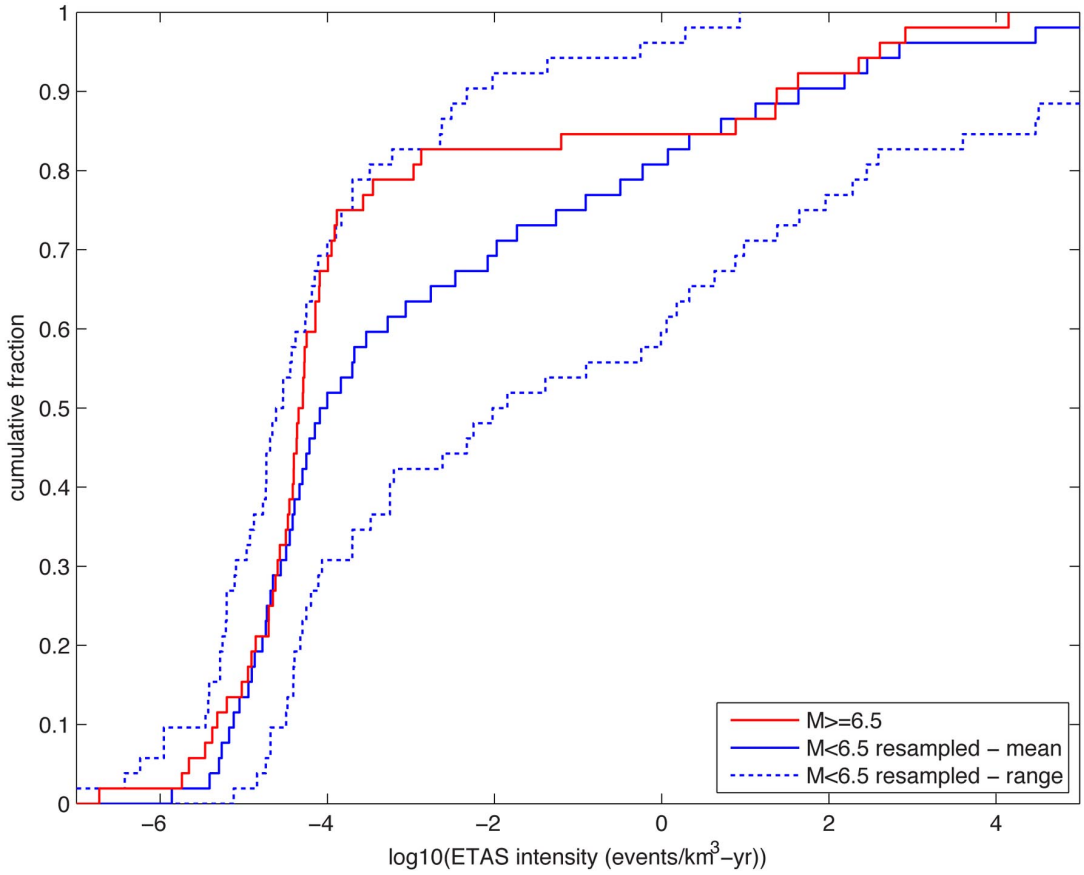

Figure S8. Graph showing cumulative distribution of the Epidemic Type Aftershock Sequence intensity at the locations and times of all earthquakes with magnitudes greater than or equal to 6.5 ($M \geq 6.5$) in the full Uniform California Earthquake Rupture Forecast, version 3 catalog (red line). For comparison, the mean (blue solid line) and range (blue dashed lines) of 3,000 resampled catalogs of earthquakes with magnitudes less than 6.5 ($M < 6.5$) that match the size and temporal sampling of the $M \geq 6.5$ event catalogs are shown. The resampled catalogs were created by randomly choosing one $M < 6.5$ event for each $M \geq 6.5$ event, occurring within 10 years before or after the larger earthquake.

It is illustrative to consider the ETAS intensity as a function of the M_{min} of the catalog for a few recent large earthquakes. The computed ETAS intensity for nine significant earthquakes since 1984 is shown in figure S9. The intensity is computed for catalogs starting in 1984, and with minimum magnitudes ranging from $M_{min}=2.5$ to $M_{min}=6.0$. All intensities are adjusted from the probability of a $M \geq M_{min}$ earthquake to the probability of a $M \geq 6.5$ earthquake, assuming $b=1.0$.

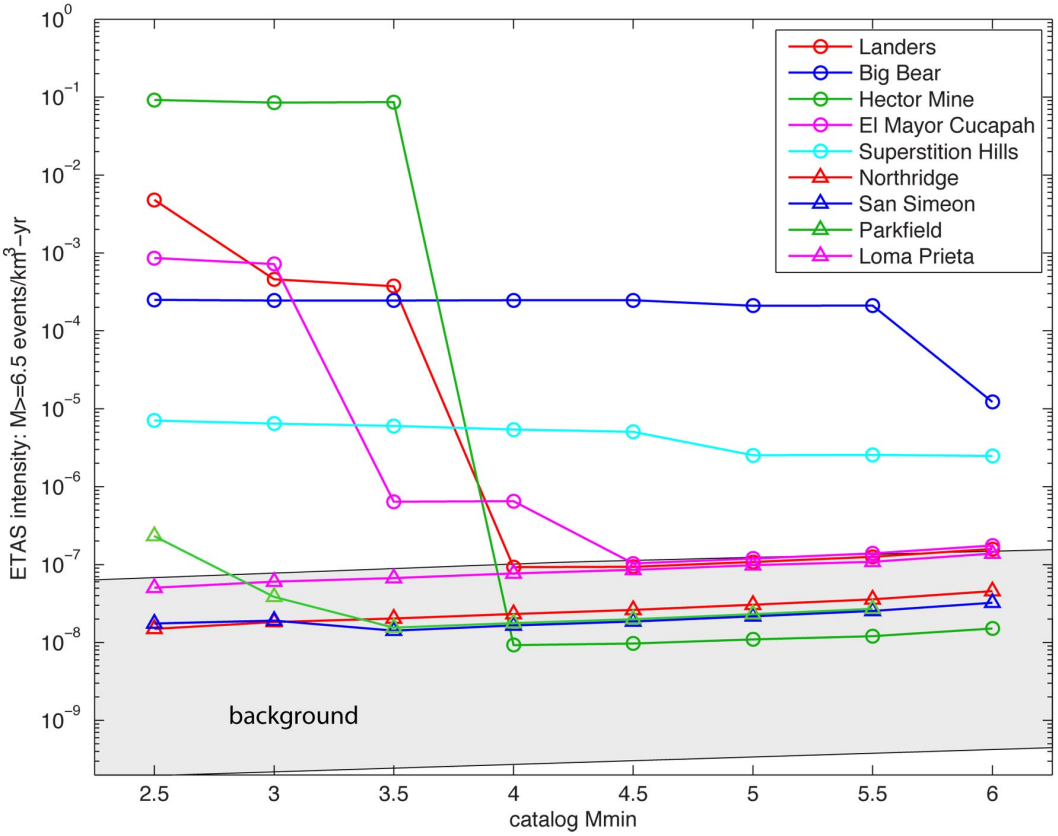

Figure S9. Graph showing computed Epidemic Type Aftershock Sequence intensity for nine significant earthquakes in California since 1984. The intensity is computed for catalogs starting in 1984, and with minimum magnitudes ranging from minimum magnitude (M_{min})=2.5 to M_{min}=6.0. All intensities are adjusted from the probability of a magnitude greater than or equal to ($M\geq$) M_{min} ($M\geq M_{min}$) earthquake to the probability of a $M\geq$6.5 earthquake, assuming b=1.0. Gray shading shows the range of background intensities over the California region, for the given M_{min}.

An example of large earthquake clustering is a series of four M>6 earthquakes in the Eastern California Shear Zone (ECSZ) in the 1990s (Hauksson and others, 1993, 2002; fig. S1). This sequence began on April 23, 1992, with the M6.2 Joshua Tree earthquake. On June 28, 1992, the M7.3 Landers earthquake occurred, the largest event of the sequence, followed about 3 hours later by the M6.5 Big Bear aftershock. Seven yr later, the October 16, 1999 M7.1 Hector Mine earthquake occurred within the extended Landers aftershock zone. The intensity at the time and location of the Big Bear aftershock is high for all M_{min}, consistent with the clear triggering of this event by the Landers mainshock. The ETAS intensities at the locations and times of the Landers and Hector Mine events increase dramatically when the minimum magnitude is decreased. If only $M\geq$4 events are considered, the intensity at the location and time of either earthquake is not significantly different from the background of the model. If all $M\geq$3.5 events are included in the ETAS model, the expected rate of $M\geq$6.5 earthquakes increases by many orders of magnitude. Secondary triggering through the smaller events links together the larger earthquakes of this sequence. A cluster of at least 25 aftershocks of the Joshua Tree earthquake near the eventual Landers hypocenter, including 3 $M\geq$3 events on the day of Landers, contributed to the very high ETAS intensities at Landers (Hauksson and others, 1993). A cluster of Landers

aftershocks near the eventual Hector Mine hypocenter played a similarly important role in the triggering of the Hector Mine earthquake. The high triggering intensity at the Hector Mine hypocenter was owing to a sequence of 16 $M \geq 3$ earthquakes in 1992, 1996, and 1999, including a $M3.7$ event about 7 hours prior to the Hector Mine mainshock. Felzer and others (2002), in a detailed study, also concluded that Hector Mine probably was triggered by this cluster of aftershocks, rather than directly by the Landers mainshock.

Two other events show a high ETAS intensity. The November 24, 1987 $M6.7$ Superstition Hills earthquake was preceded by the $M6.5$ Elmore Ranch earthquake about 11 hours earlier (Hudnut and others, 1989). These two earthquakes are close enough in space and time that the ETAS intensity owing to the Elmore Ranch earthquake is high at the location and time of the Superstition Hills earthquake. So, like the Big Bear earthquake, which also was closely preceded by a $M \geq 6.5$ earthquake, the intensity for the Superstition Hills earthquake is high for any M_{min}. In contrast, the April 4, 2010 $M7.2$ El Mayor-Cucapah earthquake was preceded by several weeks of foreshocks, including a cluster with a $M4.4$ event during the 24 hours preceding the mainshock (Hauksson and others, 2010). Including $M \geq 4.0$ earthquakes in the ETAS model increases the intensity somewhat above background intensity owing to the $M4.4$ foreshock, and including all $M \geq 2.5$ earthquakes further increases the intensity by including more of the foreshock sequence.

Four other earthquakes show ETAS intensities similar to background intensity for all M_{min} earthquakes. The intensity at the time and location of the October 18, 1989 $M6.9$ Loma Prieta earthquake was not raised significantly above background by the preceding June 27, 1988 $M5.3$ and August 8, 1989 $M5.4$ Lake Elsman earthquakes, which are considered to be foreshocks to the Loma Prieta earthquake (for example, Perfettini and others, 1999). The 1994 $M6.7$ Northridge earthquake (Hauksson and others, 1995) occurred on a conjugate thrust fault to the 1971 $M6.6$ San Fernando earthquake rupture (Mori and others, 1995). There were two clusters of possible foreshocks during the weeks prior to the Northridge earthquake, each including a $M \geq 3.5$ event (Hauksson and others, 1995), but the approximate 30-km distance of these clusters from the Northridge hypocenter greatly diminishes their triggering potential in the ETAS model. The 2003 $M6.5$ San Simeon earthquake (McLaren and others, 2008) was not preceded by any foreshocks or other large events, so it is not surprising that the intensity is low for all M_{min} earthquakes. The 2004 $M6.0$ Parkfield earthquake (Bakun and others, 2005) occurred just 9 months after the San Simeon earthquake. There was no cataloged foreshock activity before the Parkfield event, although Meng and others (2010) detected a sequence of small earthquakes on the San Andreas Fault at Parkfield immediately after the San Simeon earthquake that escaped catalog detection.

These examples show that small earthquakes, when present in the catalog, can significantly affect the probabilities of large events. Including small earthquakes in the ETAS model allows for the identification of regions with active clusters of triggered earthquakes, which may be foreshocks to later large earthquakes, and to assign earthquake probabilities to these regions that are significantly higher than background probabilities. Had operational earthquake forecasting been done at the time of the ECSZ sequence, for example, forecasts based only on large earthquakes would have had limited usefulness, while forecasts based on small earthquakes would have been quite successful. Including smaller events is, therefore, important for operational earthquake forecasting to successfully assess the probability of large, potentially damaging earthquakes.

Spatial-Temporal Clustering: Declustered Catalogs

Another way to investigate the spatial-temporal clustering of large earthquakes is to consider the spatial and temporal separations of all pairs of $M \geq 6.5$ earthquakes, after declustering the catalog, and to search for patterns such as an excess number of earthquake pairs with certain temporal and spatial separations. I declustered each of the historical and instrumental earthquake catalogs by removing all earthquakes with a ≥0.5 probability of being triggered, where this probability is defined as the fraction of the total ETAS intensity that comes from triggering. Because the distribution of triggering probabilities is strongly bimodal with peaks near 0 and 1, sophisticated stochastic declustering methods (for example, Zhuang and others, 2002) are not necessary. I considered every pair of events that appeared together in at least one declustered catalog (but did not count pairs more than once if they appeared together in more than one catalog) because these pairs of events do not trigger one another according to the ETAS model. I compared the distribution of spatial and temporal separations to the predicted distribution for random earthquake occurrence. The predicted spatial separations are derived from the spatial probability distribution for the background earthquakes (fig. S2), and the predicted temporal separations assume that earthquakes are random and stationary in time. The difference between the observed and predicted number of pairs in bins of 5 years by 50 km is shown in figure S10a.

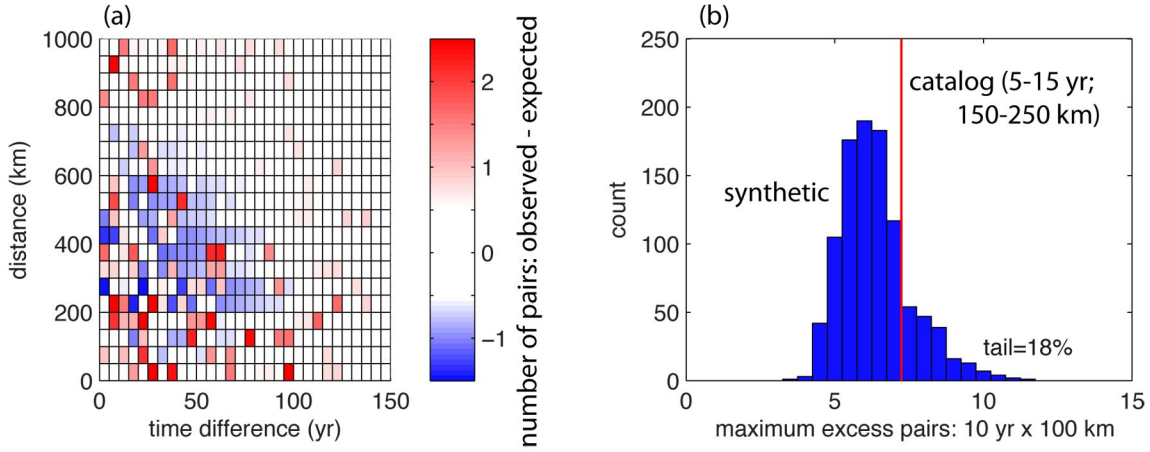

Figure S10. Graphs showing (a) number of pairs of magnitude greater than or equal to 6.5 earthquakes from the declustered instrumental and historical catalogs with the given spatial and temporal separation, minus the predicted number of pairs. The predicted spatial separations are derived from the spatial probability distribution for the background earthquakes (fig. S2), and the predicted temporal separations assume that earthquakes are random and stationary in time. (b) Red line shows the observed excess number of earthquake pairs within 5–15 years and 150–250 km of each other. The histogram shows the maximum number of excess pairs in any 10 yearrs × 100 km window for 3,000 synthetic catalogs. The synthetic event locations are drawn randomly from the spatial probability distribution for the background earthquakes (fig. S2), and the synthetic earthquake times are random over a period of 150 years.

The distribution of spatial-temporal separations shows a number of bins with an excess of one to two pairs, or a deficit of about one pair, which is not surprising given the small sample size. There is not a substantial excess of pairs at relatively short spatial and temporal distances (<20 years and <150 km), indicating that the ETAS model does a good job of accounting for

large earthquake clustering on these scales. To test for significant large earthquake clustering at larger scales, I chose spatial-temporal regions with an excess of pairs, and tested the null hypothesis that these excess pairs could have occurred by random chance. One spatial-temporal test region includes earthquake pairs with 5-15 year temporal separation and 150-250 km spatial separation, with an excess of 7.2 pairs (non-integer because the expected value can be non-integer.) To test the significance of this apparent pattern, I created 3,000 synthetic datasets, each with the same number of event pairs as the real dataset, with the event locations drawn randomly from the spatial probability distribution for the background earthquakes (fig. S2). Eighteen percent of the synthetic datasets showed at least one 10-year × 100 km spatial-temporal region with an excess of ≥7.2 pairs. Similar tests for other regions with an excess of event pairs had similar results. I. therefore, do not reject the null hypothesis that the pattern of spatial and temporal separations occurred at random, and that there is no true clustering of large earthquakes in the declustered catalogs.

Conclusions

Parameter values for the Epidemic Type Aftershock Sequence (ETAS) model are determined by fitting the model to the recent instrumental earthquake catalog. The ETAS parameters are validated on the modern instrumental catalog, by comparing the observed with the predicted cumulative number of earthquakes through time. The cumulative number of observed magnitude greater than or equal to ($M\geq$)2.5 earthquakes is similar to the cumulative number predicted by the ETAS model using the best-fit parameters, except that the catalog appears to be missing a large number of aftershocks of the April 25, 1992 M7.2 Cape Mendocino earthquake. This is not likely to be a problem with the productivity parameters, given that the ETAS model matches the aftershock sequences of similar-sized events. The ETAS parameters were further validated on the 1984–2011 catalog using super-thinning. A small excess of earthquakes at short distances (≤6 km) was observed, which may be attributable to location uncertainty, the approximate nature of the large-earthquake finite rupture models, and the lack of finite rupture models for intermediate-sized earthquakes. Deviance residuals showed that the parameter values found here out-performed published alternatives.

I then addressed whether the ETAS model explains the clustering of large ($M\geq$6.5) earthquakes, or whether additional large-earthquake clustering is present. I used three approaches to test for unmodeled clustering of $M\geq$6.5 earthquakes. First, I tested the null hypothesis that the transformed times of the $M\geq$6.5 events are Poissonian, which should be the case if the ETAS model successfully accounts for all earthquake clustering. Next, I studied the distribution of the ETAS intensity, $\lambda(t,\mathbf{x})$, at the locations and times of $M\geq$6.5 earthquakes, and tested the null hypothesis that this distribution is the same as for small earthquakes, which would imply that the ETAS model forecasts large events as well as it does for small events. Finally, I used the ETAS model to decluster the catalog, and to test the null hypothesis that there is no residual clustering of $M\geq$6.5 earthquakes in the declustered catalog. I cannot reject these null hypotheses, so I concluded that the ETAS model adequately models the occurrence of $M\geq$6.5 earthquakes.

I also showed that small earthquakes, when present in the catalog, can significantly affect the probabilities of large events. Including smaller earthquakes in the ETAS model allows for the identification of regions with active clusters of triggered earthquakes, which may be foreshocks to later large earthquakes, and for the assignment of earthquake probabilities to these regions that are significantly higher than background probabilities. An example of large earthquake clustering is a series of four magnitude greater than 6 earthquakes in the Eastern California Shear Zone

(ECSZ) in the 1990s. An ETAS model including only large events results in a high modeled earthquake rate only at the location and time of the Big Bear earthquake. Including all earthquakes $M \geq 2.5$ increases the probabilities at the times and locations of the Landers and Hector Mine earthquakes as well. Had operational earthquake forecasting been done at the time of the ECSZ sequence, forecasts based only on larger earthquakes would have had limited usefulness, while forecasts based on smaller earthquakes would have been quite successful.

Acknowledgments

I am grateful to Agnès Helmstetter (University of Grenoble) and Frederic Schoenberg (UCLA) for constructive reviews of this appendix.

References Cited

Aki, K., 1965, Maximum likelihood estimate of *b* in the formula *log N = a - b M* and its confidence limits: Bulletin of the Earthquake Research Institute, v. 43, p. 237–239.

Bakun, W.H., Aagaard, B., Dost, B., Ellsworth, W.L., Hardebeck, J.L., Harris, R.A., Ji, C., Johnston, M.J.S., Langbein, J., Lienkaemper, J.J., Michael, A.J., Murray, J.R., Nadeau, R.M., Reasenberg, P.A., Reichle, M.S., Roeloffs, E.A., Shakal, A., Simpson R.W., and Waldhauser, F., 2005, Implications for prediction and hazard assessment from the 2004 Parkfield earthquake: Nature, v. 437, p. 969–974.

Clements, R.A., Schoenberg, F.P., and Schorlemmer, D., 2011, Residual analysis for space-time point processes with applications to earthquake forecast models in California: Annals of Applied Statistics, v. 5, no. 4, p. 2549–2571.

Enescu, B., Mori, J., Miyazawa, M., and Kano, Y., 2009, Omori-Utsu law c-values associated with recent moderate earthquakes in Japan: Bulletin of the Seismological Society of America, v. 99, p. 884–891.

Felzer, K.R., 2008, Appendix I—Calculating California seismicity rates*, in* the Uniform California Earthquake Rupture Forecast, Version 2 (UCERF 2): U.S. Geological Survey Open-File Report 2007-1437. (Also available at *http://pubs.usgs.gov/of/2007/1437/i/*.)

Felzer, K.R., Abercrombie, R.E., and Ekström, G., 2003, Secondary aftershocks and their importance for aftershock forecasting: Bulletin of the Seismological Society of America, v. 93, p. 1433–1448.

Felzer, K.T., Becker, T.W., Abercrombie, R.E., Ekström, G., and Rice, J.R., 2002, Triggering of the 1999 M_w 7.1 Hector Mine earthquake by aftershocks of the 1992 M_w 7.3 Landers earthquake: Journal of Geophysical Research, v. 107, no. B9, p. 2190, doi:10.1029/2001JB000911.

Felzer, K.R., and Brodsky, E.E., 2006, Decay of aftershock density with distance indicates triggering by dynamic stress: Nature, v. 441, p. 735–738.

Felzer, K.R., and Kilb, D., 2009, A case study of two *M* ~5 mainshocks in Anza, California—is the footprint of an aftershock sequence larger than we think?: Bulletin of the Seismological Society of America, v. 99, p. 2721–2735.

Gardner, J.K., and Knopoff, L.,1974, Is the sequence of earthquakes in southern California, with aftershocks removed, Poissonian?: Bulletin of the Seismological Society of America, v. 64, p. 1363–1367.

Gutenberg, B., and Richter, C.F., 1944, Frequency of earthquakes in California: Bulletin of the Seismological Society of America, v. 34, p. 185–188.

Hainzl, S., Scherbaum, F., and Beauval, C., 2006, Estimating background activity based on interevent-time distribution: Bulletin of the Seismological Society of America, v. 96, p. 313-320.

Hardebeck, J.L., Felzer, K.R., and Michael, A J., 2008, Improved tests reveal that the accelerating moment release hypothesis is statistically insignificant: Journal of Geophysical Research, v. 113, B08310, doi:10.1029/2007JB005410.

Hauksson, E., Jones, L.M., Hutton, K., and Eberhart-Phillips, D., 1993, The 1992 Landers earthquake sequence—seismological observations: Journal of Geophysical Research, v. 98, no. B11, p. 19,835–19,858.

Hauksson, E., Jones, L.M., and Hutton, K., 1995, The 1994 Northridge earthquake sequence in California—seismological and tectonic aspects: Journal of Geophysical Research, v. 100, no. B7, p. 12,335–12,355, doi:10.1029/95JB00865.

Hauksson, E., Jones, L.M., and Hutton, K., 2002, The 1999 M_w 7.1 Hector Mine, California, earthquake sequence—complex conjugate strike-slip faulting: Bulletin of the Seismological Society of America, v. 92, p. 1154–1170.

Hauksson, E., Stock, J., Hutton, K., Yang, W., Vidal-Villegas, J.A., and Kanamori, H., 2010, The 2010 M_w 7.2 El Mayor-Cucapah earthquake sequence, Baja California, Mexico and southernmost California, USA—active seismotectonics along the Mexican Pacific margin: Pure and Applied Geophysics, doi:10.1007/s00024-010-0209-7.

Helmstetter, A., and Sornette, D., 2002, Subcritical and supercritical regimes in epidemic models of earthquake aftershocks: Journal of Geophysical Research, v. 107, no. B10, p. 2237, doi:10.1029/2001JB001580.

Hudnut, K.W., Seeber, L., and Pacheco, J., 1989, Cross-fault triggering in the November 1987 Superstition Hills earthquake sequence, southern California: Geophysical Research Letters, v. 16, p. 199–202.

Marsan, D., and Lengliné, O., 2010, A new estimation of the decay of aftershock density with distance to the mainshock: Journal of Geophysical Research, v. 115, B09302, doi:10.1029/2009JB007119.

McLaren, M.K., Hardebeck, J.L., van der Elst, N., Unruh, J., Bawden, G.W., and Blair, J.L., 2008, Complex faulting associated with the 22 December 2003 M_w 6.5 San Simeon, California earthquake, aftershocks and postseismic deformation: Bulletin of the Seismological Society of America, v. 98, p. 1659–1680.

Meng, X., Peng, Z, and Hardebeck, J.L., 2010, Detecting missing earthquakes on the Parkfield section of the San Andreas fault following the 2003 M_w 6.5 San Simeon earthquake: Abstract S43D-08, presented at 2010 American Geophysical Union Fall Meeting, San Francisco, California, 13-17 Dec.

Mori, J., Wald, D.J., and Wesson, R.L., 1995, Overlapping fault planes of the 1971 San Fernando and 1994 Northridge, California earthquakes: Geophysical Research Letters, v. 22, p. 1033–1036.

Ogata, Y., 1988, Statistical models for earthquake occurrences and residual analysis for point processes: Journal of the American Statistical Association, v. 83, p. 9–27.

Perfettini, H., Stein, R.S., Simpson, R., and Cocco, M., 1999, Stress transfer by the 1988–1989 M=5.3 and 5.4 Lake Elsman foreshocks to the Loma Prieta fault—unclamping at the site of peak mainshock slip: Journal of Geophysical Research, v. 104, p. 20169–20182.

Sornette, D., and Werner, M.J., 2005, Apparent clustering and apparent background earthquakes biased by undetected seismicity: Journal of Geophysical Research, v. 110, no. B9, doi:10.1029/ 2005JB003621.

Utsu, T., 1965, A method for determining the value of *b* in a formula *log n = a – bM* showing the magnitude frequency for earthquakes: Geophysical Bulletin of Hokkaido University, v. 13, p. 99–103.

Werner, M.J., Helmstetter, A., Jackson, D.D., and Kagan, Y.Y., 2011, High-resolution long-term and short-term earthquake forecasts for California: Bulletin of the Seismological Society of America, v. 101, p. 1630–1648.

Wessel, P., and Smith, W.H., 1998, New, improved version of generic mapping tools released: Eos, American Geophysical Union Transactions, v. 79, no. 47, p. 579.

Wong, K., and Schoenberg, F.P., 2009, On mainshock focal mechanisms and the spatial distribution of aftershocks: Bulletin of the Seismological Society of America, v. 99, p. 3402–3412.

Zhuang, J., Ogata, Y., and Vere-Jones, D., 2002, Stochastic declustering of space-time earthquake occurrences: Journal of the American Statistical Association, v. 97, no. 458, p. 369–380.