

Using the ETAS Model for Catalog Declustering and Seismic Background Assessment

R. CONSOLE,¹ D. D. JACKSON,² and Y. Y. KAGAN²

Abstract—The concept of background seismicity is strictly related to the identification of spontaneous and triggered earthquakes. The definition of foreshocks, main shocks and aftershocks is currently based on procedures depending on parameters whose values are notoriously assumed by subjective criteria. We propose a method for recognizing the background and the induced seismicity statistically. Rather than using a binary distinction of the events in these two categories, we prefer to assign to each of them a probability of being independent or triggered. This probability comes from an algorithm based on the ETAS model. A certain degree of subjectivity is still present in this procedure, but it is limited by the possibility of adjusting the free parameters of the algorithm by rigorous statistical criteria such as maximum likelihood. We applied the method to the seismicity of southern California and analyzed the sensitivity of the results to the free parameters in the algorithm. Finally, we show how our statistical declustering algorithm may be used for mapping the background seismicity, or the moment rate in a seismic area.

Key words: Seismic background, declustering, epidemic model, maximum likelihood.

1. Introduction

Temporal clustering, as commonly observed during aftershock sequences, constitutes strong evidence for time-dependent behavior of the seismic process and its departure from a simple spatially-variable, time-independent Poisson process. Nevertheless, seismicity is often modeled by a process with a rate constant in time, and most earthquake damage mitigation measures in seismic areas are based on this assumption. The tectonic processes causing seismic activity apparently change only on a geologic time scale, so that the driving stress and seismic moment

rate can be considered constant over a long-term observation.

Regardless of the behavior of the seismicity in time, it is widely recognized that the magnitude distribution of earthquakes follows the well-known Gutenberg–Richter (G–R) law (GUTENBERG and RICHTER, 1944) modeled by Eq. 5 in the Appendix. The value of the b parameter in the G–R law is typically close to 1.0 in different areas of the world and for different magnitude ranges, so expressing a universal property of the seismicity (BIRD and KAGAN, 2004).

The validity of the G–R law is limited to a range of magnitude where the lower limit is generally connected to the minimum magnitude reported with completeness by the observation system, and the upper limit is related to the maximum size of the rupture produced by the seismogenic process in a given region. Without an upper limit to the magnitude distribution, the total moment rate of any region would be infinite. Therefore, several modifications to the linear trend have been proposed (see KAGAN, 2002 and BIRD and KAGAN, 2004 for a thorough analysis of this issue).

Seismologists have traditionally labeled earthquakes as “foreshocks”, “main shocks” or “aftershocks”, assuming that only the main shocks can be described by a time-independent stationary process (see e.g., GARDNER and KNOPOFF, 1974 and REASENBERG, 1985). However, a quantitative physical definition of main shock has never been given, and the separation of the above mentioned three classes of earthquakes is based on empirical subjective definitions that are checked against the stationarity of the main shock rate in time.

In order to overcome the above-mentioned problems, a number of people (e.g., KAGAN and KNOPOFF,

¹ Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy.
E-mail: console@ingv.it

² University of California, Los Angeles, USA.

1976; KAGAN, 1991; KAGAN and JACKSON, 2000; ZHUANG *et al.*, 2002, 2004, 2008; ZHUANG and OGATA, 2006) have already described some form of stochastic declustering. Their methods are based on the application of models whose parameters are fit by maximum likelihood criteria. While KAGAN and KNOPOFF (1976) used a branching model of earthquake occurrence, ZHUANG *et al.* (2002) made use of a similar model called Epidemic Type Aftershock Sequence (ETAS), in which the background and the clustered structure are estimated in non-parametric and parametric ways, respectively (see Sect. 3 for more details on this model). In this study we used the same kind of method used by ZHUANG *et al.* (2002), with the only difference being that they used variable kernel functions to estimate the background rate, while our method uses FRANKEL'S (1995) method. More recently, MARSAN and LONGLINÉ (2008) introduced an iterative procedure that finds the best declustering algorithm in a way that the result is somewhat independent of the initial maximum likelihood model parameters. In their algorithm the background and the clustering structure are both estimated in non-parametric ways.

Based on criteria already applied in previous studies, our method makes use of an iterative procedure for obtaining a spatially variable model of the seismicity, suitable for seismological applications and seismic hazard estimates. However, we are not aiming at producing a catalog where some of the events are removed, but rather a catalog where the events receive a weight proportional to the probability of being independent.

The data used for all the numerical applications are drawn from the seismic catalog of the Southern California Earthquake Data Center (SCEDC) (http://www.data.scec.org/catalog_search/date_mag_loc.php). Southern California is one of the most densely sampled seismic regions of the world.

2. A Smoothing Algorithm for a Spatially Variable Poisson Model

Figure 1a shows the epicenter distribution of 60,480 earthquakes with magnitude equal to or larger than 2.0 reported by SCEDC from 1984 to 2002. The

area taken for the analysis is a rectangle 360 km × 440 km wide, centered on the point of geographical coordinates 34.5°N and 117.0°W. As clearly shown in Fig. 1b, the catalog can be considered complete in the period of time chosen for the analysis. Even if some small magnitude events are missing, especially soon after other larger magnitude earthquakes, the

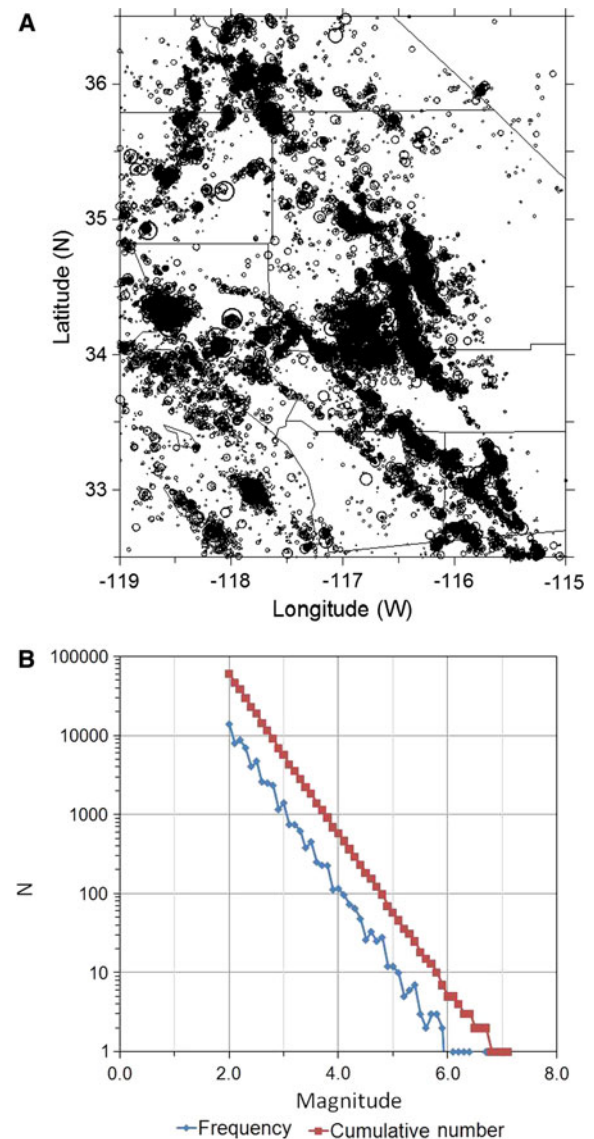


Figure 1

a Epicentral distribution of the earthquakes with magnitude $M \geq 2.0$ reported by the Southern California Earthquake Data Center in the time period 1984–2002. **b** Frequency-magnitude distribution of the earthquake catalog considered in this study

circumstance does not affect the results in terms of background seismicity distribution. This is because, as explained in the introduction, and discussed later on in this paper, these early events, if recorded, would be strongly down-weighted by the algorithm for the computation of the background seismicity.

In order to obtain a continuous rate density $\lambda_0(x,y)$, we applied an algorithm for smoothing the discrete epicenter distribution. Applications of algorithms making use of smoothing kernels for seismic hazard assessment can be found in KAGAN and JACKSON (1994, 2000), JACKSON and KAGAN (1999), and HELMSTETTER *et al.* (2007), whose ideas are largely reflected in this study.

Here we compute a gridded smooth geographical distribution of the seismic rate density at each node k of a regular grid through the method introduced by FRANKEL (1995):

$$\tilde{N}_k = \frac{\sum_l N_l \exp(-\Delta_{kl}^2/d^2)}{\sum_l \exp(-\Delta_{kl}^2/d^2)}, \quad (1)$$

where N_l is the number of events in each cell centered on the l th node, Δ_{kl} is the distance between nodes k and l , and d is a free parameter.

The spatially variable rate density $\lambda_0(x,y)$ (used in Eq. 5 of the Appendix) is computed dividing by the observation time interval the value obtained from linear interpolation of \tilde{N}_k among the four nearest grid nodes. The free parameter d is determined by maximizing the likelihood of the seismicity contained in the later half of the catalog under the model obtained from the earlier half. In this study the best fit of the d value has been carried out by maximizing the likelihood of the SCEDC earthquake catalog ($M \geq 2.0$) from 1993 to 2002 under the time-independent Poisson model obtained from the same catalogue in the period from 1984 to 1992. Figure 2 shows the epicenter distributions for the two separate time periods. These two maps show a remarkable similarity, indicating that the spatial distribution of the seismic activity is rather stable in time, though the map of the first period exhibits stronger activity in the area of the Landers (1992) earthquake and the map of the second period shows concentration of epicenters in the areas of the Northridge (1994) and Hector Mine (1999) earthquakes.

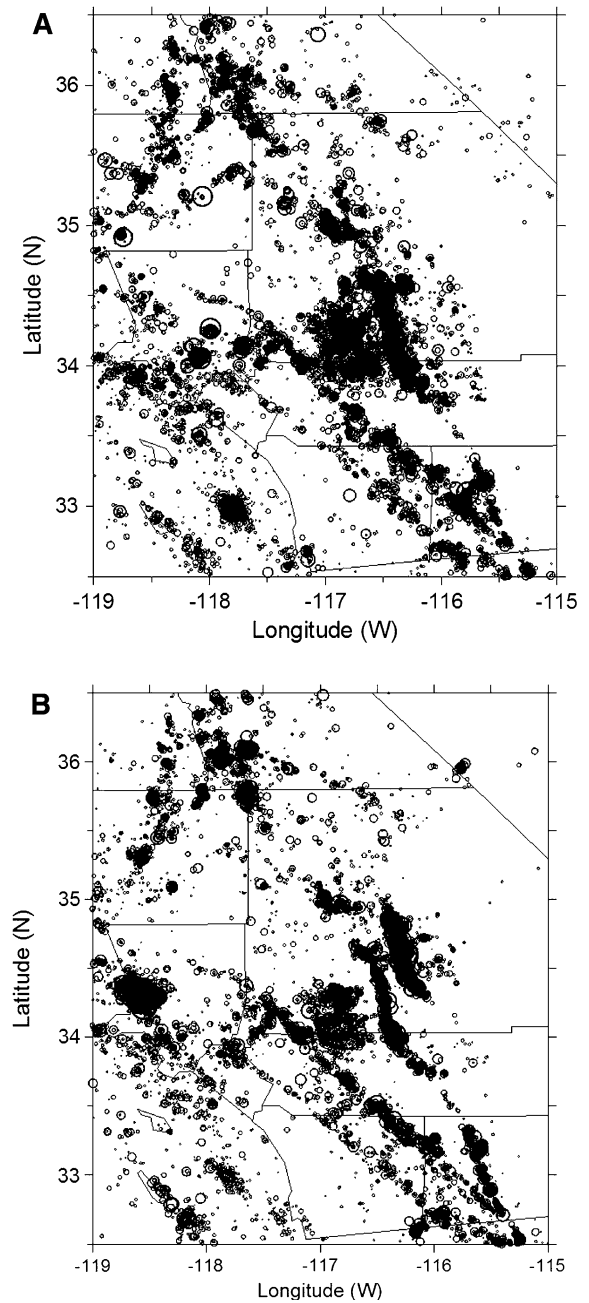


Figure 2

a Epicentral distribution of the earthquakes with magnitude $M \geq 2.0$ observed in Southern California in the time period 1984–1992. **b** As in **a**, for the time period 1993–2002

Figure 3 shows the plot of the likelihood of the second part of the catalog estimated from the first part, as a function of the correlation distance d . A maximum is visible around $d = 5.0$ km, which is the value taken for the following analysis. The smoothed

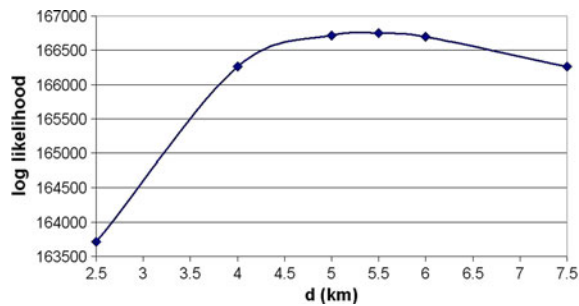


Figure 3

Plot of the likelihood of the Southern California 1993–2002 earthquake catalog under the time-independent Poisson model representing the smoothed spatial distribution of the 1984–1993 catalog, versus different values of the d parameter of the smoothing algorithm

geographical distribution so obtained for the whole data set is shown in Fig. 4.

3. Modeling the Earthquake Catalog by the ETAS Model

As shown in Fig. 5 (line a), the cumulative number of all the earthquakes in the catalog is strongly irregular, because of the presence of after-shock sequences. We may suppose, in agreement with assumptions underlying the ETAS model (see e.g., OGATA, 1998, 1999 CONSOLE and MURRU, 2001; HELMSTETTER and SORNETTE, 2002, 2003; CONSOLE *et al.*, 2007 and references therein) that the seismicity has two components. The first is assumed to have a time-independent Poisson behavior, and the second is represented by the triggered earthquakes. A short outline of the ETAS model is reported in Appendix. Note that stationarity of the background rate and the parameters of the ETAS model is assumed in this, as in many other studies, though it cannot be really validated.

We aim at a spatial distribution that does not include the triggered component of the seismicity, still preserving the total seismic moment released by the seismicity, which is approximately proportional to the total number of earthquakes. With this purpose, we applied an iterative process based on a clustering (epidemic) time dependent model (see also MARSAN and LONGLINÉ, 2008):

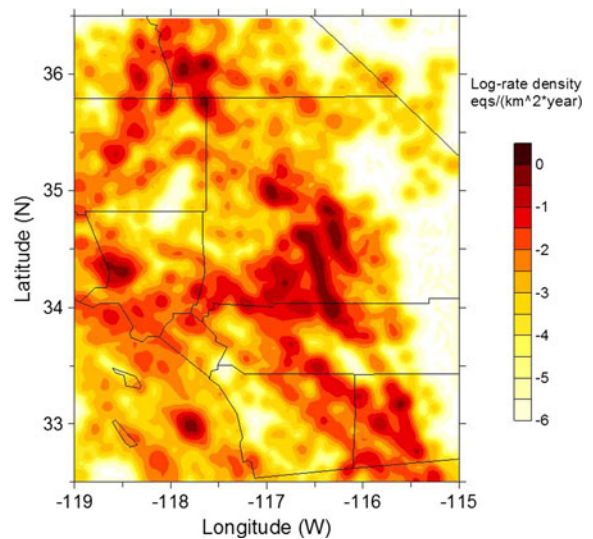


Figure 4

Smoothed distribution of the Southern California (1984–2002) seismicity obtained by the smoothing algorithm of Eq. 1 with a correlation distance $d = 5.0$ km. The color scale represents the number of events occurred in cells $2 \text{ km} \times 2 \text{ km}$ wide over the total duration of the catalog (i.e., 19 years)

1. We started by finding the maximum likelihood set of free parameters using the initial distribution of the smoothed seismicity $\lambda_0(x,y)$ by interpolation of the gridded distribution obtained through Eq. 1;
2. We computed the probability of independence p_i as the ratio between the independent component $f_r \cdot \lambda_0(x_i, y_i, m_i)$ and the composite rate density $\lambda(x_i, y_i, m_i, t_i)$ for every event i in Eq. 4 of the Appendix;
3. Then, without adjusting the d parameter again, we computed a new distribution of $\lambda_0(x,y)$ through Eq. 1 as in step 1, but introducing the weights p_i as a multiplying factor for each event, and dividing by f_r to normalize the result to the total number of events in the catalog;
4. The new smoothed distribution was used in a new maximum likelihood best fit of the free parameters;
5. We proceeded again from step 2, and so on, until a reasonable convergence was reached.

In order to limit the time needed for the maximum likelihood best fit, which is rather computer demanding, step 1 of the above-mentioned process was carried out using a threshold of magnitude 3.0 both for the triggering and the triggered earthquakes

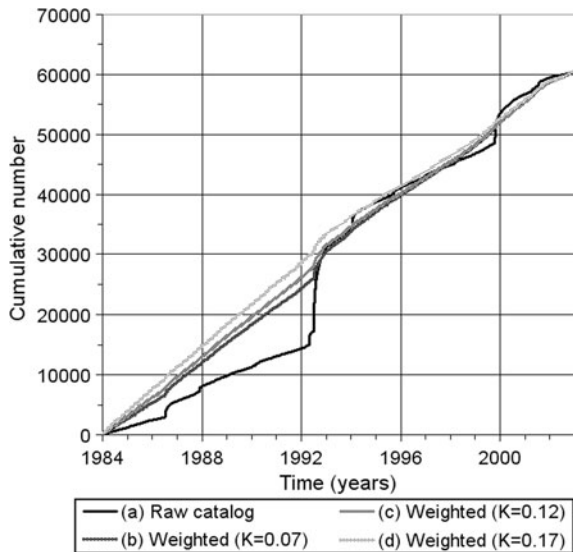


Figure 5

a Cumulative distribution of the number of events in the Southern California (1984–2002) earthquake catalog. **b** Cumulative distribution of the weights p_i over the time spanned by the Southern California (1984–2002) earthquake catalog, normalized to the total number of events; see point 2 in Sect. 3 of the text for the definition of p_i ; the model parameters used for obtaining this plot are those shown in the last column of Table 1. **c** As in **b**; the model parameters used for obtaining this plot are those shown in the last column of Table 1, with the substitution of the value $K = 0.12$ for this parameter. **d** As in **b**; the model parameters used for obtaining this plot are those shown in the last column of Table 1, with the substitution of the value $K = 0.17$ for this parameter

(for a total of 5,733 events). The time required for carrying out the same computations, but with a magnitude threshold one unit lower, would have been about 100 times longer. The smoothed distribution of the spatial density, however, was computed on the basis of the entire data set with a lower magnitude threshold equal to 2.0. The best fit values of the

parameters obtained after five iterations of the procedure described above are reported in Table 1. The parameters affected by the largest variations are those modeling the spatial distribution (d_0 and q), while the parameters related to the time decay are considerably more stable. Moreover, there is a negative correlation (as expected) between the productivity parameter K and d_0 , because both are related to the total number of triggered events. Parameter d_0 is probably influenced also by the location errors of the epicenters reported in the catalog, which is not taken in consideration in our simple algorithm.

Figure 6 shows the map of the new smoothed distribution, using the algorithm described in step 3 and the model parameters obtained after the final iteration. Comparing this map with that shown in Fig. 4, we note a more diffuse distribution of the seismicity, and a better correspondence to the fault system.

As it can be clearly noted in Fig. 5 (line b), the cumulative distribution of p_i over the time spanned by the catalog, normalized to the total number of events, is much closer to a linear trend than the starting cumulative number of events. This gives a good indication of the ability of our algorithm to remove the contribution of the triggered seismicity, without really deleting any event from the catalogue.

4. Constraining the Weighting Process for the Most Uniform Total Rate

The total earthquake rate over the region is represented by the slope of the cumulative number of events versus time. Our iterative procedure results in

Table 1

Values obtained for the parameters of the epidemic model in the progressive adjustment of the background seismicity

	Initial	Iter. 2	Iter. 3	Iter. 4	Iter. 5
K (days $^{p-1}$)	0.0637	0.0673	0.0674	0.0678	0.0698
d_0 (km)	0.504	0.502	0.499	0.498	0.478
q	1.884	1.859	1.853	1.850	1.812
c (days)	0.0154	0.0154	0.0152	0.0155	0.0154
p	1.080	1.079	1.081	1.081	1.082
α	0.563	0.544	0.546	0.546	0.549
f_r	0.0772	0.0838	0.0840	0.0817	0.0822
$\ln L$	49,985.4	50,086.5	50,101.1	50,104.4	50,107.1

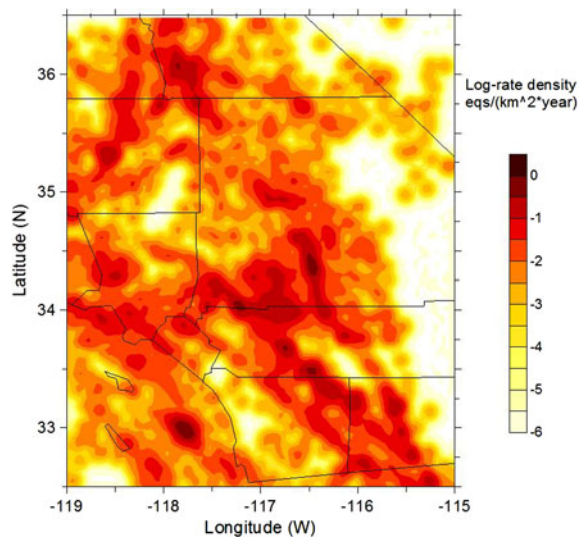


Figure 6

Smoothed distribution of the Southern California (1984–2002) seismicity obtained by the smoothing algorithm applied to the weighted catalog as described in Sect. 3 of the text and the model parameters shown in the last column of Table 1

a fairly uniform total rate of background events, yet, it is quite evident that a bump is left in the cumulative distribution. This bump shows that the effect of the aftershock sequence of the Landers (1992) earthquake has not been completely removed. We make the arbitrary assumption that an ideal weighting algorithm should achieve a uniform total rate of background seismicity. In this respect, we are interested in exploring the idea that a modification of the parameters of the ETAS model could achieve a more uniform background rate without compromising the likelihood too much.

We define the parameter D_n (the same used in the Kolmogorov–Smirnov or K–S test; see for instance GIBBONS and CHAKRABORTI, 2003) to express the mismatch between the actual cumulative distribution and the desired linear trend with the same total earthquake count:

$$D_n = \max |\hat{F}_n(i) - F_0(i)|, \quad (2)$$

where $\hat{F}_n(i)$ is the observed value of the normalized cumulative distribution at the event i , and $F_0(i)$ is the theoretical normalized linear trend, so that $\hat{F}_n(n) = F_0(n)$. The value of this parameter for the ETAS model parameters reported in the last column of

Table 1 and the cumulative distribution shown in Fig. 5 (line b), is $D_n = 0.083$.

Through a set of empirical tests, we searched for the value of K in the ETAS model that minimizes D_n . Figure 7, where the results of this search are reported, shows how, increasing the K value, D_n decreases substantially, and then it increases again after having reached a minimum ($D_n = 0.033$) at $K = 0.12$ (almost twice the value obtained from the maximum likelihood best fit, see Table 1).

The new cumulative distribution, obtained using the previous set of best fit parameters, but with the substitution of $K = 0.12$, is shown in Fig. 5 (line c). Figure 8 shows the corresponding map of the smoothed spatial distribution, using the algorithm described in Sect. 3 and the model parameters with the value of K adjusted at $K = 0.12$. A comparison between the maps of Figs. 6 and 8 shows a very slight change in the smoothed seismicity distribution, though the value of the K productivity parameter has been raised by more than 70%. This can be explained considering that the ETAS model using the maximum likelihood estimate of K is already capable of suppressing nearly all the clustered activity. However, the map of Fig. 8 is characterized by a degree of smoothing slightly higher than that of Fig. 6.

We computed the log-likelihood of the same catalog used in the previous analysis, with a lower magnitude threshold equal to 3.0, using the final best fit parameters of the ETAS model reported in the last column of Table 1, but with the smoothed seismicity distribution obtained using the value of K adjusted at $K = 0.12$ as shown in Fig. 8. This test provided a

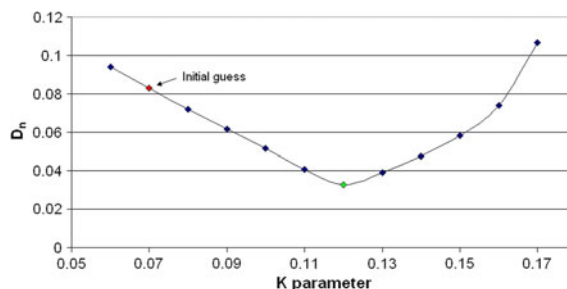


Figure 7

Plot of the D_n parameter, showing a measure of the departure of the cumulative distribution of events from a theoretical linear trend, versus the K value of the ETAS model used for giving a weight to each event in the seismic catalog

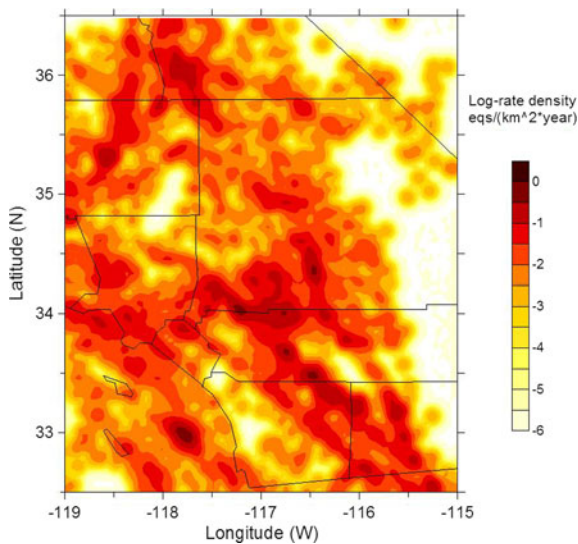


Figure 8

Smoothed distribution of the Southern California (1984–2002) seismicity obtained by the smoothing algorithm applied to the weighted catalog as described in Sect. 3 of the text and the model parameters shown in the last column of Table 1, with the substitution of the value $K = 0.12$ for this parameter

value of 50,111.4 for the log-likelihood, which is slightly better than that of the best fit obtained with the smoothed density distribution of Fig. 6, obtained from the best fit value of $K = 0.07$. Of course, we use $K = 0.12$ only for the preparation of the smoothed seismicity, because this value, if used in the ETAS model, would yield a much smaller value of the likelihood.

A visual inspection shows that the new total background rate of Fig. 5 (line c) is not yet uniform. This is also confirmed by the fact that the K–S test for a linear distribution is not passed at a 90% confidence level. A further increase of the parameter K , together with the complete disappearance of the effect of the aftershock sequences on the cumulative distribution, produces a decrease of the slope just at the time of the Landers (1992) earthquake. As an example, Fig. 5 (line d) shows the plot of the cumulative distribution obtained for $K = 0.17$.

5. Testing the Model on Independent Data

It is generally acknowledged that testing a forecast model requires a data set which is completely

independent of the one used in the learning process (see e.g., CONSOLE, 2001; CONSOLE *et al.*, 2003; HELMSTETTER *et al.*, 2007). In line with this rule, we have tested the quality of the ETAS model relative to a time-independent, spatially variable model, using a new data set for the same area of southern California, spanning the time period 2003–2007. At the same time, we have also tested the impact of the different choices for the seismicity smoothing algorithm used for obtaining the background spatial distribution. In these tests we have considered both the lower magnitude threshold $m_0 = 3.0$ used in the learning phase (including only $N_3 = 185$ target events), and an even lower magnitude threshold $m_0 = 2.0$ (including $N_2 = 1772$ target events), which allows a more robust test on a number of events about ten times larger. The main results are reported in Table 2.

Table 2 shows in a clear way the overall better performance of the models including the weighting technique for the smoothed seismicity, compared with the standard unweighted smoothing algorithm. However, if we look at the log-performance factor, obtained as the difference between the log-likelihood of the ETAS model and that of the Poisson model (and at its average, also called probability gain per earthquake), we notice that it decreases from the model using the standard unweighted smoothing algorithm to those including the weighting technique. This is clearly a consequence of the better performance of the new background distribution applied to a time-independent model. If we compare, instead, the performances of the two weighted background distributions (the one with $K = 0.07$ and the other with $K = 0.12$), we notice very little difference. The overall likelihood is improved from the former to the latter for $m_0 = 3.0$, but gets worse for $m_0 = 2.0$. The performance factor and the probability gain per earthquake obtained from these two approaches are also not significantly different.

We may conclude that the use of the procedure described in Sect. 3 is useful for obtaining a better time-independent spatial distribution of the seismicity, which significantly improves the likelihood of the earthquakes catalogs under this model. It also gives a moderate improvement to the resulting ETAS model, which includes the effect of triggered seismicity. However, the arbitrary change of the K parameter in

Table 2

Performance of the ETAS model on the seismicity of the test period (2003–2007) with different models of the time-independent background seismicity distribution

	Unweighted smoothing	Weighted smoothing with $K = 0.07$	Weighted smoothing with $K = 0.12$
$\ln L_e$ ($m_0 = 3.0$)	1,560.8	1,586.5	1,588.4
$\ln L_e - \ln L_0$ ($m_0 = 3.0$)	1,739.3	1,704.4	1,704.4
$(\ln L_e - \ln L_0)/N_3$ ($m_0 = 3.0$)	9.40	9.21	9.21
$\ln L_e$ ($m_0 = 2.0$)	20,109.0	20,149.0	20,147.5
$\ln L_e - \ln L_0$ ($m_0 = 2.0$)	18,622.0	18,071.4	18,022.9
$(\ln L_e - \ln L_0)/N_2$ ($m_0 = 2.0$)	10.51	10.20	10.17

L_0 —Likelihood of the time-independent, spatially variable Poisson model

L_e —Likelihood of the time-dependent ETAS model

the smoothing algorithm, even producing a more straight cumulative number of events in time, does not significantly affect the spatial seismicity model, and does not represent an advantage for the performance of the ETAS model.

6. Spatial Distribution of Seismic Moment

The information about the geographical distribution of the earthquake rate λ_0 may have relevant consequences for the seismic hazard assessment in a given area. For instance, assuming that all tectonic stress is fully released by earthquakes, and that the magnitude distribution is a truncated Gutenberg–Richter distribution, it is possible to obtain the expected seismic moment rate \dot{M}_0 from the number of earthquakes observed per unit space and time λ_0 (WARD, 1994; CATALI *et al.*, 2008):

$$\dot{M}_0 \cong \lambda_0 M_0^* S \frac{b}{1.5 - b} \left[10^{(1.5-b)(m_{\max} - m_0)} - 1 \right] \quad (3)$$

where:

b is the parameter of the Gutenberg–Richter magnitude distribution,

m_0 is the lower magnitude threshold of the catalog

m_{\max} is the assumed maximum possible magnitude

M_0^* is the seismic moment of an earthquake of magnitude m_0 , and S is the area of the seismogenic zone

Taking into account that only a fraction of the tectonic moment rate is released by earthquakes, Eq. 3 must be retained as a constraint putting a lower limit on

the estimate of the moment rate. The practical estimate of the moment rate \dot{M}_0 is conditioned by the constraint that the seismogenic volume used for the application of Eq. 3 should be small enough as to allow the assumption of a spatially uniform value of \dot{M}_0 . At the same time the number of observed earthquakes should be large enough to justify the assumption of stationarity for the seismic process. These constraints are obviously in conflict within each other, so that a reasonable compromise should be chosen.

A map of the moment rate can be obtained through Eq. 3, after having multiplied the seismic moment of each event by a weight p_i equal to its probability of independence. A value of the maximum magnitude $m_{\max} = 8.0$ (BIRD and KAGAN, 2004) has been assumed in this application. As for the rate density, the moment rate was then normalized by a factor equal to the ratio between the total number of events in the catalog and the sum of p_i over all these events. This was done in order to preserve the total seismic moment released by the earthquakes in the period of time covered by the catalog. Figure 9 shows the map obtained for the area considered in our study, using the spatial distribution of the occurrence rate shown in Fig. 6 ($K = 0.07$). This map is suitable for a comparison with analogous maps obtained from geodetic observations.

7. Discussion and Conclusions

Following the method proposed by ZHUANG *et al.* (2002), and adopting an iterative procedure as

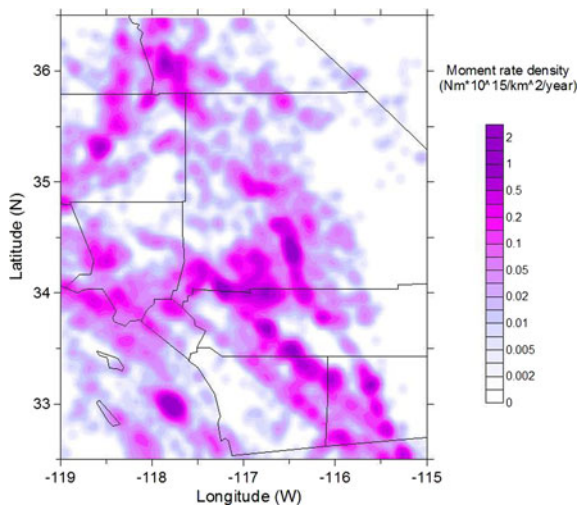


Figure 9

Map of the seismic moment rate obtained by Eq. 3 from the earthquakes observed in Southern California in the time period 1984–2002, using the value $K = 0.07$ for the productivity parameter

suggested by MARSAN and LONGLINÉ (2008), our method finds the optimal set of parameters for the ETAS model, taking also into account their effect on the best spatial distribution. The results reported in Table 1 show that the parameters related to the spatial distribution of the seismicity are most affected by the iterative procedure, while the temporal parameters are more robust.

Moreover, we have shown also that the maximum likelihood set of parameters does not guarantee a constant background rate. A more linear trend can be achieved by an appropriate change of the productivity parameter K of the ETAS model.

The method developed in this study was chosen to produce an earthquake catalog that fulfils the property of stationarity for the seismic process without the need of removing events from the catalog. Each event is assigned a weight proportional to the probability of independence according to the ETAS model with suitable values of the parameters. The probability is a number that can assume any value between 0 and 1. Figure 10 shows the distribution of the probability of independence for the 60,480 events of the SCEDC catalog analyzed in this study. It can be noted that about 50% of the events are characterized by a probability of independence close to zero. Note that among these events even earthquakes traditionally

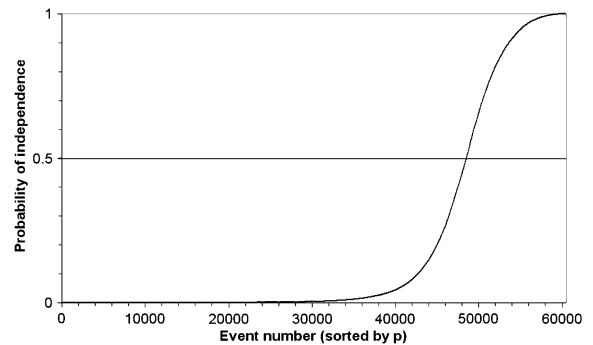


Figure 10

Distribution of the probability of independence for the 60,480 events of the Southern California earthquake catalog

defined as main shocks could be included, if these main shocks had been preceded by significant seismic activity commonly defined as foreshocks. For instance, the main shock of the Landers ($M = 7.3$) 1992 sequence has been assigned a probability of independence $p = 0.00007$ only, because of the influence of some previous foreshocks. The event that most probably triggered the main shock was a small earthquake of magnitude 2.3, which occurred about 20 h before the main shock, with an epicenter shifted by 0.5 km to the west. This event had a probability of independence $p = 0.38$. These results clearly contradict the usual geophysicist's intuition that the main shocks should be independent events. Note also that the initial $M 2.3$ earthquake (definable as a foreshock in a strict sense) triggered 2.6 aftershocks in its first generation (in probabilistic sense), while the magnitude 7.3 main shock triggered 211.9 aftershocks in its first generation. The total number of $M \geq 2.0$ aftershocks in the whole sequence was larger than 15,000. This means that the ETAS model assigns most of the events to next generations. This is consistent with the low value found for the productivity parameter α by the maximum likelihood best fit ($\alpha = 0.55$). In our ETAS model the numerous low magnitude events trigger a total number of events larger than the number of events triggered by the few big ones. This result, as for other ETAS results, is difficult to interpret in geophysical terms, because this model is based on a statistical, rather than a geophysical parameterization.

This analysis shows that for a very large fraction of earthquakes the probability of independence does not justify a clear separation between independent

and triggered events. About 20% of earthquakes in the analyzed catalog exhibit a probability of independence larger than 0.5, but for few of them the probability is very close to 1. For comparison, MARSAN and LONGLINÉ (2008) found that the background rate of $M \geq 3$ earthquakes corresponds to 19.5% of the total rate.

The background seismicity distribution obtained by down-weighting the events of the catalog according to their probability of independence still exhibits large peaks of the spatial density. This is clearly the case for the Landers and Oceanside areas, as shown in both Figs. 6 and 8. A detailed analysis of the probability of independence of the earthquakes included in these small zones shows that these high values of the spatial density do not come from the aftershock sequences, but from a rather persistent moderate seismic activity. For instance, for a square of 10 km \times 10 km centered on the epicenter of the Landers (1992) main shock (34.2°N, 116.4°W), the analysis shows that the total weight of the events included in the first day of the aftershock series (101 events) is only 0.0082. For the first 10 days (447 events) the total weight is 0.097, and extending the count to the first 100 days (941 events) we obtained a total weight equal to 0.94. Note that the weight is fairly proportional to the time duration of the considered period, and not to the number of events contained in every period. The smaller proportion of activity in the first day can be ascribed to the increased detection threshold of the first hours in the network.

In the case of the Oceanside area (33.0°N, 117.8°W), the same kind of analysis shows a seismic activity even more sparse in time, with many more events of relevant probability of independence distributed all over the observation period. Therefore, the ETAS model achieving the maximum likelihood accommodates this circumstance with a relatively high level of the time-independent component.

Any attempt to derive a background earthquake rate using a short catalog is going to encounter serious difficulties and challenges. It is clear from the density maps in this paper, as well as from many maps published by other researchers in the 2007(1) SRL special issue (FIELD, 2007), that such a background rate model is not unique. Even when we used all of the available historic data to produce the map (KAGAN *et al.*, 2007),

one can still see that the rate is likely to be too low in certain places, and too high in others. Perhaps, only by using geodetic and geologic information (SHEN *et al.*, 2007; BIRD and LIU, 2007) one could obtain a map of a really long-term background rate. However, such a geodetic/geologic map would fail to correctly predict earthquakes on the 5- or 15-year time-span, since these events are likely to occur near sites of recent activity. Thus, the “background” rate needs a proper definition — what kind of forecast is being considered, is it 5-, 15-, 30-, or 50-year? KAGAN and JACKSON (1994) argue that the time horizon of a prediction should be comparable to the length of an earthquake catalog used in the forecast.

KAGAN *et al.* (2010) discuss general drawbacks of the ETAS model. Particularly, because the c value in the Omori’s law parameterization is not scaled with magnitude, the model fit to earthquake patterns might be strongly biased. This is related to aftershock number deficiency after strong earthquakes (see e.g., KAGAN, 2004; HELMSTETTER *et al.*, 2006). However, KAGAN *et al.* (2010) (this issue) removed all of the close-in-time aftershocks and recalculated the model parameters, and they found that the results do not change significantly. This is consistent with the strong similarity we have noted between the background seismicity distribution obtained by a large range of values for the K parameter.

Even taking into account the above-mentioned difficulties in the search for the background rate, and meaning of the background rate itself, maps of the smoothed seismicity produced by our algorithm may represent a spatially variable, time-independent Poisson model useful as a null hypothesis against which to test short-term or medium-term time-dependent forecast models.

Acknowledgments

The authors are grateful for comments by reviewers Jiancang Zhuang and Agnes Helmstetter and by the Associate Editor David Rhoades that have been helpful in the revision of the manuscript. The UCLA authors (DDJ and YYK) appreciate partial support from the Southern California Earthquake Center (SCEC). SCEC is funded by NSF Cooperative

Agreement EAR-0529922 and the U.S. Geologic Survey (USGS) Cooperative Agreement 07HQAG0008. Publication 1257, SCEC.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix

The expected occurrence rate density of earthquakes, $\lambda(x,y,t,m)$, at any time and location, is modeled as the sum of the independent, or time-invariant “spontaneous” activity and the contribution of every previous event:

$$\lambda(x, y, t, m) = f_r \cdot \lambda_0(x, y, m) + \sum_{j=1}^N H(t - t_j) \cdot \lambda_j(x, y, m, t), \quad (4)$$

where $\lambda_0(x,y,m)$ is the rate density of the long-term average seismicity, expressing the Gutenberg–Richter magnitude distribution as

$$\lambda_0(x, y, m) = \lambda_0(x, y) \beta e^{-\beta(m-m_0)}; \quad (5)$$

f_r is the failure rate (fraction of spontaneous events over the total number of events) of the process;

β is related to the most widely known b value by the relationship $\beta = b \ln 10$;

m_0 is a reference magnitude;

$H(t)$ is the step function;

and $\lambda_j(x,y,t,m)$ is a kernel function that depends on the magnitude of the triggering earthquake, the spatial distance from the triggering event, and the time interval between the triggering event and the time of interest. We factor this function in three terms depending, respectively, on time, space and magnitude, as:

$$\lambda_j(x, y, t, m) = K \cdot f(x - x_j, y - y_j) \cdot h(t - t_j) \cdot \beta e^{-\beta(m-m_0)}, \quad (6)$$

where K is a constant parameter, while $f(x,y)$ and $h(t)$ represent the space and time distributions, respectively.

The spatial distribution of the triggered seismicity is modeled by a function with circular symmetry around the point of coordinates (x_j, y_j) . This function in polar coordinates (r, θ) can be written as:

$$f(r, \theta) = \left(\frac{d_j^2}{r^2 + d_j^2} \right)^q, \quad (7)$$

where r is the distance from the point (x_j, y_j) , q is a free parameter modeling the decay with distance, and d_j is the characteristic triggering distance. We assume that d_j is related to the magnitude m_j of the triggering earthquake:

$$d_j = d_0 10^{\alpha(m_j - m_0)/2}, \quad (8)$$

where d_0 is the characteristic triggering distance of an earthquake of magnitude m_0 and α is a free parameter, i.e., the distance distribution is scaled with magnitude.

For the time dependence we adopt the modified Omori law (OGATA, 1983):

$$h(t) = (t + c)^{-p} (p > 1), \quad (9)$$

where c and p are characteristic parameters of the process.

The free parameters for the ETAS model actually estimated in this study are the following:

- K (productivity coefficient),
- d_0 (characteristic triggering distance),
- q (exponent of the spatial distribution of triggered events),
- α (coefficient of the exponential magnitude productivity law),
- c (time constant of the generalized Omori law) and
- p (exponent of the generalized Omori law).

The fraction of spontaneous events over the total number of events of the process, f_r , is constrained by these free parameters. The b value is estimated from the entire catalog independently from the other parameters.

We may note that neither the spatial (Eq. 7) nor the temporal (Eq. 9) kernel distributions are normalized. This might produce biases in the parameters of the model that control the productivity of triggered seismicity. In particular, this is the case for K and the spatial parameters d_0 , q and α . In spite of these biases, the best fit process always converges to stable results,

independently of the guess choices for the parameters' values to initiate the optimization algorithm.

REFERENCES

- BIRD, P. and KAGAN, Y. Y. (2004), *Plate-tectonic analysis of shallow seismicity: Apparent boundary width, beta, corner magnitude, coupled lithosphere thickness, and coupling in seven tectonic settings*, Bull. Seismol. Soc. Am. 94(6), 2380–2399 (plus electronic supplement).
- BIRD, P. and LIU, Z. (2007), *Seismic hazard inferred from tectonics: California*, Seism. Res. Lett. 78(1), 37–48.
- CATALI, F., COCCO, M., CONSOLE, R. and CHIARALUCE, L. (2008), *Modeling seismicity rate changes during the 1997 Umbria-Marche sequence (central Italy) through a rate- and state-dependent model*, in preparation.
- CONSOLE, R., *Testing earthquake forecasting hypotheses* (2001), Tectonophysics, 338, 261–268.
- CONSOLE, R., MURRU, M., and LOMBARDI, A. M. (2003), *Refining earthquake clustering models*, J. Geophys. Res. 108, 2468, doi: [10.1029/2002JB002130](https://doi.org/10.1029/2002JB002130).
- CONSOLE, R. and MURRU, M. (2001), *A simple and testable model for earthquake clustering*, J. Geophys. Res. 106, 8699–8711.
- CONSOLE, R., MURRU, M., CATALI, F., and FALCONE, G. (2007), *Real time forecasts through an earthquake clustering model constrained by the rate-and-state constitutive law: Comparison with a purely stochastic ETAS model*, Seism. Res. Lett. 78, 49–56.
- FIELD, E. H. (2007), *Overview of the Working Group for the Development of Regional Earthquake Likelihood Models (RELM)*, Seism. Res. Lett. 78(1), 7–16.
- FRANKEL, A. (1995), *Mapping seismic hazard in the central and eastern United States*, Seism. Res. Lett. 66, 8–21.
- GARDNER, J. K. and KNOPOFF, L. (1974), *Is the sequence of earthquakes in California, with aftershocks removed, Poissonian?*, Bull. Seism. Soc. Am. 64, 1363–1367.
- GIBBONS, J. D. and CHAKRABORTI, S. *Non-parametric Statistical Inference*, 4th ed. (New York: Marcel Dekker 2003), 645 pp.
- GUTENBERG, B. and RICHTER, C. (1944), *Frequency of earthquakes in California*, Bull. Seism. Soc. Am. 34, 185–188.
- HELMSTETTER, A. and SORNETTE, D. (2002), *Subcritical and supercritical regimes in epidemic models of earthquake aftershocks*, J. Geophys. Res. 107, 2237, doi: [10.1029/2001JB001580](https://doi.org/10.1029/2001JB001580).
- HELMSTETTER, A. and SORNETTE, D. (2003), *Importance of direct and indirect triggered seismicity in the ETAS model of seismicity*, Geophys. Res. Lett. 30, 1576, doi: [10.1029/2003GL017670](https://doi.org/10.1029/2003GL017670).
- HELMSTETTER, A., KAGAN, Y. Y., and JACKSON, D. D. (2006), *Comparison of short-term and time-independent earthquake forecast models for southern California*, Bull. Seism. Soc. Am. 96(1), 90–106.
- HELMSTETTER, A., KAGAN, Y. Y., and JACKSON, D. D. (2007), *High-resolution time-independent grid-based forecast for $M \geq 5$ earthquakes in California*, Seism. Res. Lett. 78(1), 78–86.
- JACKSON, D. D. and KAGAN, Y. Y. (1999), *Testable earthquake forecasts for 1999*, Seism. Res. Lett. 70(4), 393–403.
- KAGAN, Y. Y. (1991), *Likelihood analysis of earthquake catalogs*, Geophys. J. Int. 106(1), 135–148.
- KAGAN, Y. Y. (2002), *Seismic moment distribution revisited: II. Moment conservation principle*, Geophys. J. Int. 149(3), 731–754.
- KAGAN, Y. Y. (2004), *Short-term properties of earthquake catalogs and models of earthquake source*, Bull. Seism. Soc. Am. 94(4), 1207–1228.
- KAGAN, Y. Y., BIRD, P., and JACKSON, D. D. (2010), *Earthquake Patterns in Diverse Tectonic Zones of the Globe*, Pure Appl. Geophys. (this issue).
- KAGAN, Y. Y. and JACKSON, D. D. (1994), *Long-term probabilistic forecasting of earthquakes*, J. Geophys. Res. 99, 13,685–13,700.
- KAGAN, Y. Y. and JACKSON, D. D. (2000), *Probabilistic forecasting of earthquakes, (Leon Knopoff's Festschrift)*, Geophys. J. Int. 143, 438–453.
- KAGAN, Y. Y., JACKSON, D. D., and RONG, Y. F. (2007), *A testable five-year forecast of moderate and large earthquakes in southern California based on smoothed seismicity*, Seism. Res. Lett. 78(1), 94–98.
- KAGAN, Y. Y. and KNOPOFF, L. (1976), *Statistical search for non-random features of the seismicity of strong earthquakes*, Phys. Earth Planet. Inter. 12(4), 291–318.
- MARSAN, D. and LONGLINÉ, O. (2008), *Extending earthquakes' reach through cascading*, Science 319, 1076–1079.
- OGATA, Y. (1983), *Estimation of the parameters in the modified Omori formula for aftershock frequencies by the maximum likelihood procedure*, J. Phys. Earth 31, 115–124.
- OGATA, Y. (1998), *Space-time point-process models for earthquake occurrences*, Ann. Inst. Statist. Math. 50(2), 379–402.
- OGATA, Y. (1999), *Seismicity analysis through point-process modeling: A review*, Pure Appl. Geophys. 155, 471–507.
- REASENBERG, P. (1985), *Second-order moment of Central California seismicity*, J. Geophys. Res. 90, 5479–5495.
- SHEN, Z. K., JACKSON, D. D., and KAGAN, Y. Y. (2007), *Implications of geodetic strain rate for future earthquakes, with a five-year forecast of $M5$ earthquakes in southern California*, Seism. Res. Lett. 78(1), 116–120.
- WARD, S. N. (1994), *A multidisciplinary approach to seismic hazard in California*, Bull. Seism. Soc. Am. 84, 1293–1309.
- ZHUANG, J. and OGATA, Y. (2006), *Properties of the probability distribution associated with the largest event in an earthquake cluster and their implications to foreshocks*, Phys. Rev. E, doi: [10.1103/PhysRevE.73.046134](https://doi.org/10.1103/PhysRevE.73.046134).
- ZHUANG, J., OGATA, Y., and VERE-JONES, D. (2002), *Stochastic declustering of space-time earthquake occurrences*, J. Am. Stat. Ass. 97, 458, 369–380.
- ZHUANG, J., OGATA, Y., and VERE-JONES, D. (2004), *Analyzing earthquake clustering features by using stochastic reconstruction*, J. Geophys. Res. 109 (B5), B05301, doi: [10.1029/2003JB002879](https://doi.org/10.1029/2003JB002879).
- ZHUANG, J., CHRISTOPHERSEN, A., SAVAGE, M. K., VERE-JONES, D., OGATA, Y., and JACKSON, D. D. (2008), *Differences between spontaneous and triggered earthquakes: Their influences on foreshock probabilities*, J. Geophys. Res. 113, B11302, doi: [10.1029/2008JB005579](https://doi.org/10.1029/2008JB005579).