



HAL
open science

Mixing of rescaled data and Bayesian inference for earthquake recurrence times

A. Corral

► **To cite this version:**

A. Corral. Mixing of rescaled data and Bayesian inference for earthquake recurrence times. *Nonlinear Processes in Geophysics*, 2005, 12 (1), pp.89-100. hal-00302503

HAL Id: hal-00302503

<https://hal.science/hal-00302503>

Submitted on 18 Jun 2008

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

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

Mixing of rescaled data and Bayesian inference for earthquake recurrence times

A. Corral

Grup de Física Estadística, Departament de Física, Facultat de Ciències, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Barcelona, Spain

Received: 20 July 2004 – Revised: 18 November 2004 – Accepted: 6 January 2005 – Published: 25 January 2005

Part of Special Issue “Seismicity pattern dynamics”

Abstract. The limits of a recently proposed universal scaling law for the probability distributions of earthquake recurrence times are explored. The scaling properties allow to improve the statistics of occurrence of large earthquakes over small areas by mixing rescaled recurrence times for different areas. In this way, the scaling law still holds for events with $M \geq 5.5$ at scales of about 20 km, and for $M \geq 7.5$ at 600 km. A Bayesian analysis supports the temporal clustering of seismicity against a description based on nearly-periodic events. The results are valid for stationary seismicity as well as for the nonstationary case, illustrated by the seismicity of Southern California after the Landers earthquake.

1 Introduction

The statistical properties of seismicity have received considerable interest for many years, in particular the Omori law for aftershocks, which quantifies the temporal decay of the number of aftershocks after a mainshock, and the Gutenberg-Richter law for the number of earthquakes of a given magnitude (Gutenberg and Richter, 1965; Reasenber and Jones, 1989; Kagan, 1994; Utsu et al., 1995; Turcotte, 1997; Turcotte and Malamud, 2002; Utsu, 2002), see also Knopoff (1997). However, less attention has been paid to the statistics of the time between consecutive earthquakes, despite that the potentiality of such quantity in the context of risk assessment and forecasting is beyond any doubt.

The outcome of the relatively scarce number of studies on the distribution of recurrence times, which is how we will refer to the time between consecutive earthquakes, has been miscellaneous. The standard practice consist on studying separately mainshocks and aftershocks (and foreshocks), for which it is necessary an algorithm to unambiguously classify each event; however, such an algorithm does not exist. In

consequence, many recurrence-time distributions for mainshocks have been proposed: exponential (Poisson), lognormal, Weibull, gamma, etc. (Knopoff and Gardner, 1974; Udías and Rice, 1975, Smalley et al., 1987; Koyama et al., 1995; Wang and Lee, 1995; Wang and Kuo, 1998), see also the reference list by Sornette and Knopoff (1997) and Ellsworth et al. (1999). Nevertheless, it is widely believed that large events tend to repeat periodically or nearly-periodically in particular fault segments (Utsu, 1984; Stein, 1995; Kagan and Jackson, 1995; Sieh, 1996; Kagan 1997; Stein 2002; Murray and Segall, 2002; Kerr, 2004). On the other hand, for aftershock sequences following the Omori law it seems clear that the recurrence-time distribution is a power law, although there is some degree of confusion between the Omori exponent, that appears in the relation between the rate of seismic activity and the time since the mainshock, and the exponent of the recurrence-time distribution. It turns out that both exponents are different, although close to one (Utsu et al., 1995; Utsu, 2002).

Fortunately, the picture has changed after the pioneering work of Bak, Christensen, Danon, and Scanlon, where elements of scaling analysis have revealed fundamental to explore the structure of seismicity in space, time, and magnitude, and where an integrated perspective was taken, putting all events on the same footing (no distinction between mainshocks and aftershocks), therefore considering seismicity as a unique process (Bak et al., 2002; Christensen et al., 2002; Corral, 2003, 2004b). In their original paper, Bak et al. (2002) divided Southern California into arbitrary areas (with no relation with seismo-tectonic provinces) and performed a mixture of distributions of recurrence times from different zones to obtain their unified scaling law for earthquakes (that is, recurrence times coming from different areas were counted together in a single distribution). However, Bak et al. did not study the important issue of the distribution of recurrence times for a single zone.

2 Scaling law for single-zone recurrence-time distributions

In this case, a different scaling law has been found if the rate of seismic activity is stationary, or if the time sequence is transformed into a stationary process (Corral, 2004a). Let us consider an arbitrary spatial region, labeled by some coordinates xy ; after intensive data analysis for several seismic catalogs, we concluded that, in the stationary case, the probability density of the recurrence times there can be written as

$$D_{xy}(\tau) = R_{xy} f(R_{xy}\tau), \quad (1)$$

where τ denotes the time between consecutive earthquakes above a certain minimum magnitude M_c in the region xy under consideration (with M_c larger than the threshold of completeness of the data considered). The calculation of the recurrence times τ is straightforward, as each data i is obtained as

$$\tau_i = t_i - t_{i-1}, \quad (2)$$

where t_i is the time of occurrence of the i -th event, $i=1, 2, \dots, N$. Let us stress that, following Bak et al. (2002), the catalogs were not declustered: all events in the selected space-time-magnitude window were considered, independently of the fact that they could be considered mainshocks, aftershocks, or foreshocks.

The scaling factor R_{xy} is the rate of seismic activity, which counts the mean number of earthquakes in the region with magnitude M larger than (or equal to) M_c per unit time. As it is well known, R_{xy} usually fulfills the Gutenberg-Richter law (Gutenberg and Richter, 1965; Kagan, 1994; Turcotte, 1997),

$$R_{xy} = A_{xy}(L)10^{-bM_c}, \quad (3)$$

with the b -value close to 1, and $A_{xy}(L)$ the rate of events with $M \geq 0$ in the region of size L centered at xy (as a result of an extrapolation of the law). A_{xy} has a very complex spatial dependence; a simple assumption is to consider that it follows a fractal distribution (even b seems to show a dependence with xy and the size of the region, though much less abrupt than A_{xy}). The fulfillment of the Gutenberg-Richter relation for a given catalog is a strong support for its completeness. Notice that both R_{xy} and $D_{xy}(\tau)$ must depend (in addition to the region coordinates xy) on M_c and on the size of the region; however, these variables do not appear in the notation just for simplicity, although obviously this dependence is important.

As we have mentioned, this approach is valid in the simple case of stationary seismicity, in which R_{xy} does not significantly change for a moving time window (more rigorously we should talk about a stochastic process homogeneous in time); this implies that a plot of the accumulated number of earthquakes versus time is essentially a straight line. Note that such stationarity does not prevent the existence of aftershock sequences in the data, rather, the seismicity is composed by many small sequences at different places that are

added together to give a stationary rate. As no sequence predominates over the rest, we might refer to this regime as background seismicity as well. Nevertheless, the procedure can be generalized to the case of non-stationary seismicity, just transforming the whole time sequence into a stationary process, as we will see (Corral, 2004a).

It is convenient to remember that the probability density is defined as usual as

$$D_{xy}(\tau) = \frac{\text{Prob}[\tau \leq \tau' < \tau + d\tau]}{d\tau}, \quad (4)$$

which allows an easy estimation of D_{xy} directly from data. In principle, the size of the interval $d\tau$ should tend to zero, but in practice it is necessary a compromise to reach some statistical significance for each interval, as the number of data is not infinite. Moreover, when there are multiple scales involved in the process (in our case recurrence times from seconds to years) it is much more convenient to consider a variable $d\tau$, with the appropriate size for each scale. An easy prescription is to consider the different intervals (in seconds) growing as $[1, c]$, $[c, c^2]$, $[c^2, c^3]$, etc., where the minimum recurrence time is 1 second and $c > 1$. This is somehow equivalent to look at the recurrence times in logarithmic scale, and for this reason sometimes this binning procedure is referred to as logarithmic binning, although the length of a bin increases geometrically with respect to the previous one and depends linearly on the recurrence time.

As a consequence of Eq. (1), a plot of the dimensionless probability density $D_{xy}(\tau)/R_{xy}$ versus the dimensionless time $\theta \equiv R_{xy}\tau$ for different regions and values of M_c makes all the different distributions collapse onto a single curve, showing very clearly the shape of the scaling function f . It is worth mentioning that this approach does not involve the assumption of any model of seismic occurrence, we are only characterizing the process. Sometimes it is mistakenly assumed that the study of the distribution of recurrence times implies an underlying renewal model, (where the recurrence times are independent and identically distributed). On the contrary, any kind of process may be characterized (but not fully) by the probability density, which is in any case the most fundamental quantity of the process (but, we insist, not the only quantity defining the process).

The scaling law given by Eq. (1) is very remarkable, as it reflects the self-similarity of the temporal structure of earthquake occurrence and therefore the fractal nature of seismicity. Indeed, Eq. (1) states that there is a unique shape for the probability density of the recurrence time, independently of the tectonic properties of the region, its size, or the minimum magnitude selected; the only difference is the scale of the distribution: high M_c or small areas have low rates R_{xy} and therefore long recurrence times, which stretches the distributions, whereas for low M_c or large regions the distributions are contracted. The rescaling of these distributions with R_{xy} makes all the corresponding curves collapse onto a single one. But the data collapse holds even for distributions coming from different catalogs, corresponding to regions with

disparate tectonic properties, and in this sense we may state that Eq. (1) constitutes a universal scaling law.

It is noticeable that when we consider the seismicity of a given area and we raise the magnitude threshold M_c and then rescale the recurrence time by a factor R_{xy} , we are performing a transformation analogous to those of the real-space renormalization group (Corral, 2004c). (A similar thing occurs when we change the size of the region L .) Therefore, the scaling function f constitutes a fixed-point of a renormalization-group transformation. A trivial solution to this transformation is provided by the exponential distribution, which moreover seems to have a wide basin of attraction. In fact, Molchan (2005) has shown, imposing the independence of the seismicity between any two regions, that the only possible scaling function is an exponential. Nevertheless, our analysis of the data demands the existence of other fixed-point attractors for the renormalization transformation, related with the presence of strong correlations between events.

In order to model the behavior of the scaling function f , and only to provide a simple approximation to it, we parametrize f as a (truncated) generalized gamma distribution,

$$f(\theta) = \frac{C|\delta|}{a\Gamma(\gamma/\delta)} \left(\frac{\theta}{a}\right)^{\gamma-1} e^{-(\theta/a)^\delta}, \quad \text{for } \theta > \theta_{min} \geq 0, \quad (5)$$

(where $\Gamma(\gamma/\delta)$ is the gamma function, see below). This parametrization allows a very general shape for the scaling function, controlled by its two shape parameters, γ and δ (which have to have the same sign). For instance, for $\delta=1$ we obtain the gamma distribution (which tends to a Gaussian for $\gamma \rightarrow \infty$); for $\gamma=1$, the stretched exponential (or ‘‘contracted’’ if $\delta>1$, which yields the semi-Gaussian if $\delta=2$); $\gamma=\delta$ gives the Weibull distribution (for $\gamma=\delta=2$ it is the Rayleigh distribution); $\gamma=3$ and $\delta=2$ yields the Maxwell distribution; $\gamma=-1/2$ and $\delta=-1$, the random-walk first-return-time distribution, etc. When $\gamma/\delta \rightarrow \infty$ the distribution tends to a log-normal, and of course, $\gamma=\delta=1$ leads to the usual exponential distribution, characteristic of the well-known Poisson process. In general, if both parameters are positive we have a power law for short times and a stretched (or ‘‘contracted’’) exponential decay for long times, whereas if they are negative, the power law turns out to govern long times. The constant C is a correction to normalization due to the fact that the model is not supposed to hold for $\theta \rightarrow 0$, but only for $\theta > \theta_{min}$, whereas the parameter a is a scale parameter (but dimensionless), and could be obtained by using that the mean value of θ , $\bar{\theta}$, is enforced by scaling to be 1; for example, if $\theta_{min}=0$,

$$\bar{\theta} = a \frac{\Gamma\left(\frac{1+\gamma}{\delta}\right)}{\Gamma\left(\frac{\gamma}{\delta}\right)}, \quad (6)$$

for $\gamma>0$ and $\delta>0$, or for $\gamma<-1$ and $\delta<0$ (outside this range of γ the mean is infinite). In fact, the fact of being $\bar{\theta}=1$ is not fully exact, as $D_{xy}(\theta)$ (and also R_{xy}) is always calculated for all values of θ , whereas the scaling function only fits the data which verify $\theta > \theta_{min}$. This yields $\bar{\theta}=1$ for the whole data set

but $\bar{\theta}>1$ for the data for which the fit by the scaling function holds.

There are two reasons that explain why Eq. (1) usually does not hold for small θ ; one is the incompleteness of the catalogs in the short time scale (there is a loss of information after a big earthquake, due to the saturation of the seismometric network), whereas the second reason is the breaking of stationarity in the short time scale (it may be difficult to find a fully stationary process, rather, small aftershock sequences appear). But let us remark that with the consideration of $\theta_{min} > 0$ we are not disregarding events with $\theta < \theta_{min}$, rather, we take into account all the events, but due to the change of behavior of the distributions at about $\theta=0.05$, we restrict the fit of the relatively simple scaling function to $\theta > \theta_{min}$.

Finally, note that the scaling function f turns out to be the probability density of the dimensionless time θ , provided that the scaling relation (1) holds. Be aware that, in order to keep the notation at minimum, we will refer to two different things with the same name $f(\theta)$; these are the probability density of θ and the scaling function that fits it. If the fit were perfect, there would be no distinction between both. We believe this leads to no ambiguities in the text.

Using a least-square fit of $f(\theta)$ to the rescaled probability densities obtained from data from the NEIC worldwide earthquake catalog as well as from regional catalogs for Southern California, Japan, Spain, and the British Islands, an estimation of the parameters of f was accomplished. We stress again that all events in the region above M_c were used, independently of the hypothetical consideration of them as foreshocks, mainshocks, or aftershocks (Corral, 2004a). It turned out to be that δ was significantly close to 1, implying that the gamma distribution is a good model for D_{xy} . Moreover, as a value of γ about 0.7 was obtained, the distribution turned out to be monotonously decreasing for all τ , slowly decreasing for short θ (power law) and more rapidly for long θ (exponential factor). The corresponding coefficient of variation was $cv \simeq 1.2$.

This type of behavior shows a tendency of earthquakes to cluster in time, which was well known for aftershock sequences, but not for the stationary case, which is the subject of our research. In fact, as our data is a mixture of many small aftershock sequences, it is likely that the power-law part of $D_{xy}(\tau)$ is due to these sequences (although the exponent, $1-\gamma \simeq 0.3$, is much smaller than the usual values for aftershock sequences, which is around 1), whereas the exponential decay for long times is due to uncorrelated events. Nevertheless, it is highly surprising to obtain such a regularity in the superposition of the disparate sequences that constitute what we see as stationary seismicity.

A paradoxical consequence of this temporal distribution is that the hazard of earthquake occurrence decreases as time since the last one increases, increasing sharply when a new earthquake takes place, whereas the time one has to wait to the next earthquake does not decrease as time evolves, but just the contrary, it increases with time (Corral, 2005). Note that this is trivial for individual aftershock sequences, where the number of events decreases with time, but it is highly

nontrivial for stationary seismicity. These properties are in sharp contrast with the assumption that background seismicity is a Poisson process.

The validity of the scaling law for the recurrence time distributions, with the universality of the clustering behavior represented by the value of γ was later extended to other catalogs: New Zealand, New Madrid (Corral, 2005), and Northern California (Corral, 2004b). Overall, the law was found to be valid from regions as small as 20 km in linear size, with $M \geq 2$, to the whole globe, with M_c from 5 to 7.5. It is the main purpose of this paper to try to push further the limits of the scaling law given by Eq. (1), for which we will combine maximum-likelihood and Bayesian methods with the scaling properties of the recurrence time distribution. Also, we will test up to which degree the universal description is valid, or it is necessary to introduce different descriptions depending on M_c , the size of the region L or its coordinates xy .

3 Procedure for the data analysis

The goodness-of-fit method described above turns out to be very data consuming (a minimum of about 4 or 5 hundreds events are necessary to plot a nice recurrence-time distribution). In contrast, maximum-likelihood methods perform better with less number of data, and are more appropriate for seismic occurrence, where statistics of large earthquakes over small areas is always very poor.

The likelihood function is simply the probability that a given data set corresponds to a particular set of parameters of a (a priori) given probability distribution, but with the variables reversed: instead of being a function of the random variables (recurrence times in our case), the likelihood is considered to depend on the parameters (Mood et al., 1974). In the case of continuous variables the probability is replaced by the joint probability density, and therefore, we can write for the likelihood function of N recurrence times,

$$\ell_{xy}(\gamma, a) = D_{xy}(\tau_1, \tau_2, \dots, \tau_N | \gamma, a) = \prod_{i=1}^N D_{xy}(\tau_i | \gamma, a), \quad (7)$$

where we have made explicit the dependence of the density on the parameters γ and a , whereas θ_{min} is considered fixed. In fact, this result would seem to be valid only if the recurrence times τ_i are independent of each other (i.e. for a renewal process). Nevertheless, if the number of data is much larger than the typical range of the correlations, the previous formula still holds, as we can be sure of having a complete sampling of the values of the random variable (Cox and Oakes, 1984). Caution must be exerted then when the number of data is low. (An easy test for this effect is just to compute the likelihood for every, let us say, second or third data, and to compare with the full data set.)

As a model of the recurrence time distribution, we will use the truncated gamma distribution,

$$f(\theta) = \frac{1}{a\Gamma(\gamma, \theta_{min}/a)} \left(\frac{\theta}{a}\right)^{\gamma-1} e^{-\theta/a}, \quad \text{for } \theta > \theta_{min}. \quad (8)$$

obtained from Eq. (5) with $\delta=1$ and using the complement of the incomplete gamma function (unnormalized) to determine the value of C ,

$$\Gamma(\gamma, u) = \int_u^\infty z^{\gamma-1} e^{-z} dz \quad (9)$$

with $\Gamma(\gamma, 0) \equiv \Gamma(\gamma)$, the usual (complete) gamma function.

In addition to a good description of the data, the gamma distribution provides an easy parametrization to test the opposite concepts of clustering and nearly-periodicity; indeed, $\gamma < 1$ implies an excess of events separated by short times (with respect to a Poisson process with the same mean), and a consequent deficit for long times, whereas when $\gamma > 1$, the situation is reversed: there is a deficit of events with short recurrence times, and an excess in the long times.

The likelihood for the gamma distribution (8) turns out to be (Gross and Clark, 1975), using the dimensionless recurrence times, θ_i .

$$\ell_{xy}(\gamma, a) = \left(\frac{\theta_G}{a}\right)^{N\gamma} \frac{e^{-N\bar{\theta}/a}}{\Gamma^N(\gamma, \theta_{min}/a)} \quad (10)$$

disregarding factors independent on the variables γ and a , which are irrelevant, as we will see. $\bar{\theta}$ and θ_G are the arithmetic mean and the geometric mean, respectively (and they would be independent on M_c , L , and xy , if the scaling law (1) were exact). To be concrete,

$$\bar{\theta} = \frac{1}{N} \sum_{i=1}^N \theta_i, \quad (11)$$

$$\log \theta_G = \frac{1}{N} \sum_{i=1}^N \log \theta_i, \quad (12)$$

constituting the only way in which the data θ_i enter into the likelihood (obviously, the base of the logarithm does not matter). With this machinery we would be able to estimate the parameters γ and a , just by maximizing the likelihood function with the help of any numerical routine (Gross and Clark, 1975). Note also that there is no problem to calculate the incomplete gamma function by using some numerical recipes (Press et al., 1992).

A refinement that allows to use more information about the data is to consider the time between the last earthquake and the end of the record, $\tau_b = T - t_N$ (the subscript b comes from backward, T is the time at which the record ends and t_N is the time of the last event in the catalog). Therefore, the likelihood function in Eqs. (7) and (10) may be modified by an extra factor

$$S_{xy}(\tau_b | \gamma, a) = \frac{\Gamma(\gamma, R_{xy}\tau_b/a)}{\Gamma(\gamma, \theta_{min}/a)}, \quad (13)$$

where $S_{xy}(\tau)$ is the survivor function, defined as $S_{xy}(\tau) \equiv \text{Prob}[\tau' > \tau] = \int_\tau^\infty D_{xy}(\tau') d\tau'$. It may be important to stress that τ_b may be a continuously changing variable in some cases: if we want to evaluate the present hazard in a given area, the corresponding catalog is not

“closed”, and T is the present time, which obviously increases linearly as time evolves (Davies et al., 1989). Nevertheless, this is not our case, and when the number of events is large the contribution of this factor can be disregarded. This is what we will do, for simplicity.

The procedure just described, based on the likelihood function, has a disadvantage: it does not allow to know how accurate the estimated parameters are. Bayesian methods go one step forward; the key point is to consider the unknown parameters as “random variables”, or more precisely, uncertain quantities, with their own probability distributions, and to use some a priori information to propose the a priori distribution of the parameters, $\mathcal{D}_0(\gamma, a)$ (Loredo, 1990). Then, the Bayes theorem provides the distribution of the parameters conditioned to the observational data, $\mathcal{D}(\gamma, a | \text{data})$, in the following way,

$$\mathcal{D}(\gamma, a | \text{data}) = \mathcal{C} \mathcal{D}_{xy}(\cdots \tau_i \cdots | \gamma, a) \mathcal{D}_0(\gamma, a) = \mathcal{C} \ell_{xy}(\gamma, a) \mathcal{D}_0(\gamma, a), \quad (14)$$

where the probability density that the data take the observed values conditioned to a particular value of the parameters is given just by the likelihood, and \mathcal{C} is simply a normalization constant. If the a priori distribution is uniform (over the domain for which the likelihood is significantly different from zero) we can identify the likelihood with the (unnormalized) a posteriori distribution of the parameters.

In fact, we are mainly interested in the shape parameter γ , which determines the clustering properties of the time series; the scale parameter a is much less important, as it only sets the scale of the dimensionless time [technically, a may be considered as a nuisance or incidental parameter (Loredo, 1990)]. Therefore, we can integrate the joint distribution $\mathcal{D}(\gamma, a | \text{data})$ over a to get the (a posteriori) marginal probability density of γ , $\mathcal{D}(\gamma | \text{data})$, i.e.,

$$\mathcal{D}(\gamma | \text{data}) = \int_0^\infty \mathcal{D}(\gamma, a | \text{data}) da. \quad (15)$$

This integration can be easily performed for the gamma distribution if $\theta_{min}=0$, yielding

$$\begin{aligned} \mathcal{D}(\gamma | \text{data}) &= \mathcal{C} \frac{\Gamma(N\gamma \pm 1)}{\Gamma^N(\gamma)} \left(\frac{\theta_G}{N\bar{\theta}} \right)^{N\gamma} = \\ &= \mathcal{C} \frac{\Gamma(N\gamma \pm 1)}{\Gamma^N(\gamma)} e^{-\ln(N\bar{\theta}/\theta_G)N\gamma}, \end{aligned} \quad (16)$$

where the normalization constant \mathcal{C} has been redefined, and different a priori distributions have been used. The term -1 corresponds to a uniform a priori distribution for a whereas the term $+1$ holds when what is uniform is the distribution of the inverse of a , $1/a$. In the case of a uniform distribution for $\log a$, which corresponds to a $1/a$ -distribution for a , the ± 1 term has to be replaced by zero. This constitutes the Jeffreys priori (Loredo, 1990; Jaynes, 2003), and probably it is the most reasonable a priori distribution for a scale parameter, as it is invariant under inversion: if a is a scale parameter, $1/a$ is a scale parameter as well (in the same way that time and frequency set each one a scale) and if a has

a hyperbolic distribution (power law with exponent -1), so does its inverse, $1/a$. On the other hand, the a priori distribution of γ is always chosen as uniform (notice that if a $1/\gamma$ -prior were selected, this would favor smaller values of γ and therefore the results might be slightly biased towards the clustering case). To put it clear,

$$\mathcal{D}_0(\gamma, a) \propto \frac{1}{a}; \quad (17)$$

nevertheless, the results do not depend strongly on the a priori distributions. In case that we cannot approximate θ_{min} to zero the integration over a can be done numerically, but with care, to avoid overflows. A possible trick is to integrate $\mathcal{D}(\gamma, a | \text{data})$ divided by its maximum value.

In addition, the Bayesian approach allows the estimation of the probability density of the recurrence time not by simply substitution of the single values of the parameters which maximize the likelihood but by using all the information about them coming from the a posteriori distribution. Indeed, we can write

$$D_{xy}(\tau | \text{data}) = \int_0^\infty \int_0^\infty D_{xy}(\tau | \gamma, a) \mathcal{D}(\gamma, a | \text{data}) da d\gamma. \quad (18)$$

When $\theta_{min}=0$ we get, using the $1/a$ -prior and rescaling the distribution,

$$\begin{aligned} f(\theta | \text{data}) &\equiv D_{xy}(\theta | \text{data}) / R_{xy} = \\ &= \frac{\mathcal{C}'}{\theta} \int_0^\infty d\gamma \frac{(\theta_G^N \theta)^\gamma}{(N\bar{\theta} + \theta)^{(N+1)\gamma}} \frac{\Gamma[(N+1)\gamma]}{\Gamma^{N+1}(\gamma)}. \end{aligned} \quad (19)$$

Nevertheless, if the peak of $\mathcal{D}(\gamma, a | \text{data})$ (or that of the likelihood) is sharp enough, $D_{xy}(\tau | \text{data})$ is practically identical to $D_{xy}(\tau | \gamma, a)$.

Finally, we can take advantage that for every region and M_c -value the rescaled recurrence times θ should follow the same distribution, if the scaling law (1) holds. As in this paper we are mainly interested on extending that scaling relation to small spatial regions, we may group different regions of the same size and with the same M_c into a unique data set. In other words, we just paste all the rescaled times θ coming from different regions of size L . The size of the region, L , is measured in degrees, and refers to the span of the region both in latitude and longitude. Note that, as we consider square regions in a naive projection in which both coordinates are translated into a rectangular system, the real shape of the regions is not squared and their areas are not equal. Nevertheless, this has no importance at all in our results, due to the fact that the size of the region L does not enter into the evaluation of the scaling factors R_{xy} . Our procedure would be valid even for regions with disparate values of L , due to the existence of the scaling law (1).

With this mixing procedure, the likelihood function can be written as

$$\ell(\gamma, a) = \prod_{\forall xy} \ell_{xy}(\gamma, a), \quad (20)$$

and it is trivial to show that Eqs. (10), (16), and (19) are still perfectly valid. We will restrict this procedure to regions xy

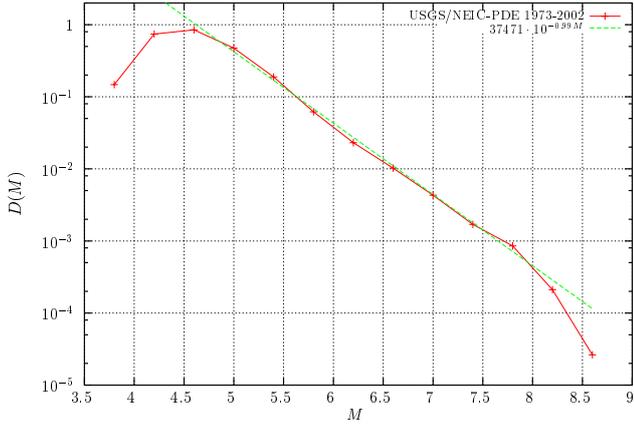


Fig. 1. Probability density of the magnitude of worldwide earthquakes, using the USGS/NEIC PDE catalog, from 1973 to 2002 (included). The size of the bins is fixed at $dM=0.4$. The straight line in the log-linear plot corresponds to an exponential decay, characteristic of the Gutenberg-Richter relation, with a b -value 0.99. The deviations for the smaller values of M are due to the incompleteness of the catalog.

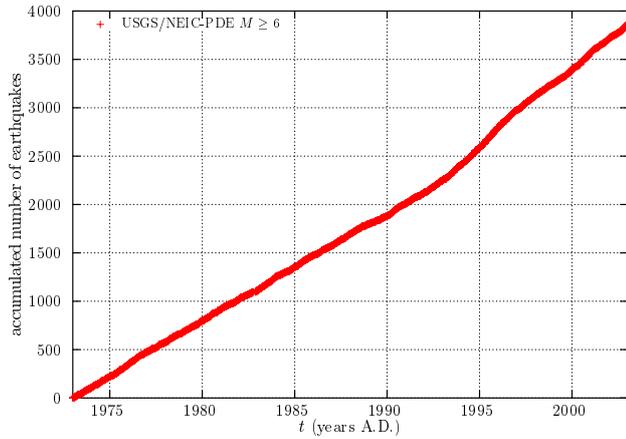


Fig. 2. Accumulated number of worldwide earthquakes with $M \geq 6$ in the USGS/NEIC PDE catalog as a function of time. The linear increase shows the stationarity of the process, with a mean rate of 0.35 earthquakes per day.

for which there are 10 or more recurrences in the period under study, to reduce the error in the calculation of the mean rate, R_{xy} .

The same can be done for the probability density of the rescaled recurrence time θ , which in any case should be given by $f(\theta)$, provided that the scaling law (1) holds. This procedure may be considered as an intermediate point between the one described in the previous section (Corral, 2004a) and the one of Bak et al. (2002). Summarizing, previously we were studying the recurrence times in a single region, and then comparing different regions after rescaling the distributions; in contrast, Bak et al. performed a mixture of recurrence times for different regions. Now we propose a mixture of recurrence times after rescaling the times, however, as the

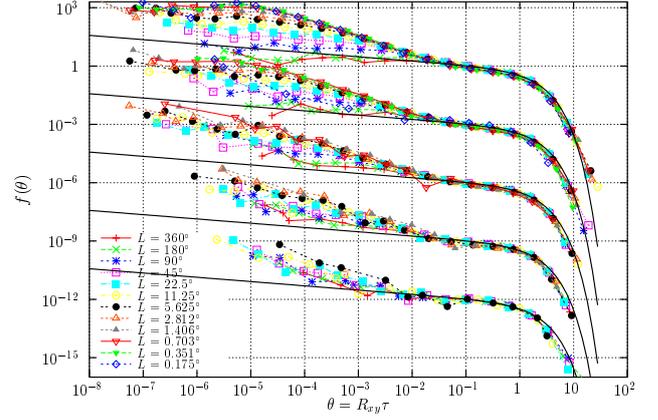


Fig. 3. Probability densities $f(\theta)$ of the rescaled recurrence time θ , mixing regions with different coordinates xy but using the same system sizes L and threshold magnitudes M_c for each distribution. Different L and M_c are also analyzed to obtain the different curves. The upper sets of curves correspond to $M_c=5$, whereas the rest of curves have been shifted, for clarity sake, a factor 10^{-3} , 10^{-6} , 10^{-9} , and 10^{-12} for $M_c=5.5$, 6, 6.5, and 7.5, respectively. The largest system size denoted by $L=360^\circ$ corresponds to a $360^\circ \times 180^\circ$ region which obviously covers the whole Earth. The smallest system sizes are $L=0.175^\circ$ for $M_c=5$ and 5.5, and $L=0.703^\circ$, 1.406° , 5.625° for $M_c=6$, 6.5, 7.5 respectively. The results for $M_c=7.5$ are provided by the catalog of Significant Worldwide Earthquakes from NOAA; the rest of the results come from the USGS/NEIC PDE catalog. The continuous lines represent the proposed gamma distribution, Eq. (8), with $\gamma=0.78$ and $a=1.38$, obtained as the maximum of $\mathcal{D}(\gamma, a | \text{data})$ mixing all the regions of any size and taking $M \geq 6$.

rescaled recurrence times should be identically distributed for any region, this allows to improve the statistics for the determination of $f(\theta)$.

In the next sections we present the results of this approach applied to several seismic catalogs, and extended to the case of nonstationary seismicity.

4 Results for worldwide seismicity

We start our analysis with global seismicity, as it appears as stationary for any time window. Two catalogs will be used, the USGS/NEIC PDE (Preliminary Determinations of Epicenters) and the catalog of Significant Worldwide Earthquakes from NOAA. Both are available at http://wwwneic.cr.usgs.gov/neis/epic/epic_global.html.

For the USGS/NEIC PDE, we study the period 1973–2002. A plot of the probability density of the earthquake magnitude (defined in the same way as in Eq. 4) shows an exponential behavior for about $M \geq 4.75$, see Fig. 1, implying that the Gutenberg-Richter relation is fulfilled above that magnitude (an exponential probability density implies an exponential survivor function, which is the normalized number of earthquakes with magnitude equal or larger than a given value). Therefore, the catalog may be considered as

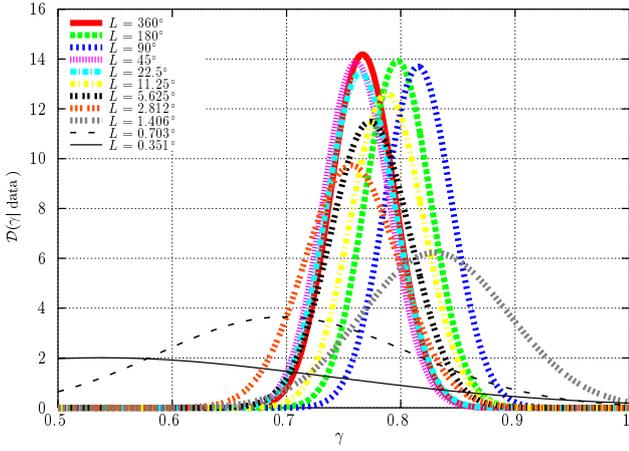


Fig. 4. A posteriori probability densities of the parameter γ , using data for the USGS/NEIC PDE catalog with $M \geq 6$. The probability that the process is not described by $\gamma < 1$ is negligible.

reasonably complete for events with magnitude larger than 5. Further, the linear behavior at Fig. 2 shows clearly how earthquakes take place at an acceptable steady rate, in this case for magnitudes larger than (or equal to) 6. The corresponding mean rate is 3859 earthquakes in 30 years $\approx 0.35 \text{ day}^{-1}$.

First of all, we may calculate the probability densities of the dimensionless recurrence time θ . Mixing the values of θ from regions with different coordinates xy but with the same L and M_c we get the results displayed in Fig. 3. The spatial range of validity of the scaling law (1) turns out to be very broad, from the whole globe to very small L ; for instance, for $M \geq 5$ and 5.5, the smaller L is $L = 0.175^\circ$ (about 20 km) and for $M \geq 6.5$, $L = 1.406^\circ$ (150 km) (see the figure caption for details). With regard the values of θ , the scaling law seems to hold very well for $\theta > 0.01$, whereas the form of the scaling function proposed in Eq. (8) is restricted to $\theta > 0.05$ for high M_c , for which the statistics is poorer. The deviations for small θ are due to the nonstationarity of the process for very short times: small aftershock sequences have little effect in the linear relation between the accumulated number of earthquakes and time, but they yield an excess of events separated by short recurrence times. In this case, as L decreases, the distributions tend to a decreasing power law with exponent very close to 1.

The resulting a posteriori probability densities of the parameter γ are shown in Fig. 4 for data with $M \geq 6$, and different values of L . The parameter θ_{min} has been set to 0.05. From the plot, it is clear that it is very unlikely that the data is generated by a gamma distribution with $\gamma > 1$, as the areas below the curves are negligible for $\gamma > 1$; therefore, we may rule out a nearly-periodic behavior. In general, the most probable γ is in between 0.75 and 0.82 for all system sizes except for the two smallest ones (practically the same values are obtained from the mean value of γ). Nevertheless, the most probable γ is not the only one possible, as its value can be scattered from 0.65 to 0.9 for the sharpest distributions (for which there are more data). For smaller system

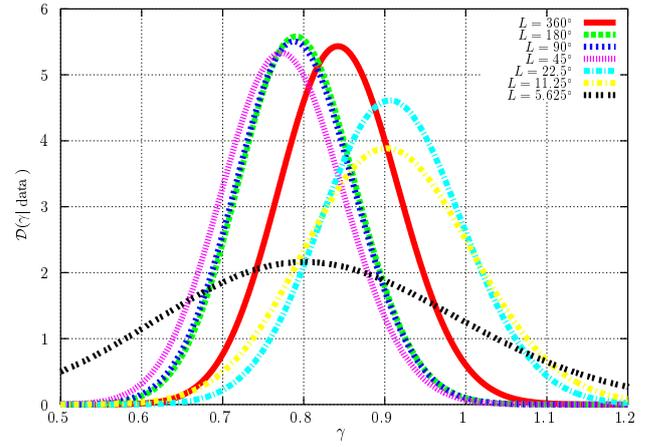


Fig. 5. A posteriori probability densities of the parameter γ , using data for the Significant Worldwide Earthquake catalog from NOAA with $M \geq 7.5$. The probability that the process is not described by $\gamma < 1$ is negligible for large L and very small for smaller L .

sizes the distributions become much broader, due to the decrease in the statistics. For instance, for $L = 1.406^\circ$ there are 57 regions with 10 or more events to give a total number of recurrences of 675; in contrast, for $L = 0.351^\circ$ there are only 6 of such regions, yielding about 70 recurrence times.

The analysis of the catalog of Significant Worldwide Earthquakes from NOAA leads to similar results. The period of analysis is 1897–1994, for which events with $M \geq 7.5$ are recorded at a stationary rate (in fact, there is a little change of the rate at about 1920, but we have not taken it into account). There are 599 of such earthquakes in the period, which corresponds to a rate of 6.1 per year. The fact that the occurrence appears as stationary for a period of about 100 years supports the assumption that the catalog can be considered as complete for $M \geq 7.5$ (if we accept that the last years of the record are indeed complete). On the other hand, the analysis of the magnitude probability density does not provide clear answers, as the range of magnitudes for which the Gutenberg-Richter law could hold is short; nevertheless we will take $M_c = 7.5$ as the threshold of completeness of the catalog.

The rescaled recurrence time distributions (mixing different xy) for $M \geq 7.5$ are shown in Fig. 3, bottom curve. The system sizes range from the whole planet to $L = 5.625^\circ$ (600 km). The flat part of the densities below $\theta = 0.3$ could be an indication of exponential behavior; however, our Bayesian analysis clearly rejects that possibility for the cases where more data are available (large regions, $L \geq 45^\circ$). Indeed, the a posteriori distributions for the parameter γ for $\theta_{min} = 0.05$ appear in Fig. 5, where for large L the worst situation corresponds to $L = 360^\circ$, which yields a probability of $\gamma < 1$ of about 98%. For the smallest system sizes less data are available: from 20 to 10 regions with a maximum number of events from about 40 to about 10 (the minimum is kept not below 10). In all these cases the probability of clustering is always above 80%.

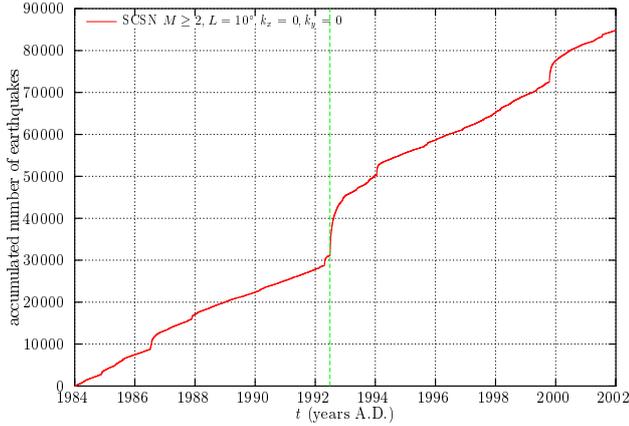


Fig. 6. Accumulated number of earthquakes in Southern California with $M \geq 2$ during the period 1984–2001, using the SCSN catalog. The seismicity consists on stationary periods “punctuated” by sudden increases of the activity, provoked by large earthquakes. The vertical line marks the occurrence of the Landers earthquake.

5 Results for regional seismicity: Southern California

As an illustration of the validity of the scaling law (1) for regional catalogs, we analyze the Southern California Seismic Network (SCSN) catalog, available at http://www.data.scec.org/catalog_search/index.html/. We will concentrate on the period 1984–2001, and on a spatial window of size 10° , located at $(-123^\circ, -113^\circ) \times (30^\circ, 40^\circ)$; further, from the fulfillment of the Gutenberg-Richter relation, the catalog can be considered practically complete for $M \geq 2$, all this yields a total of 84 772 events. Figure 6 shows the accumulated number of earthquakes versus time, being apparent that the behavior consists of linear increases alternated with more abrupt changes. These sudden changes correspond to aftershock sequences generated by large earthquakes, and they seem to be present at all scales, from very large bursts to much smaller disturbances of the constant rate. The largest of these bursts corresponds to the Landers earthquake, analyzed in detail in the next section and marked by a vertical line in the figure.

The most important periods of stationarity in the whole area of Southern California are, roughly, 1984–1985, 1988–1991, 1994–1999, and 2001. Calculating the rescaled recurrence times for these periods, for regions of different coordinates x, y , and defining one single distribution for each L and M_c (that is, we mix data not only for different x, y but also for different time periods, but not for different L and M_c), we get the results of Fig. 7 (upper set of curves), where the values of L range from $L = 10^\circ$ (1100 km) to $L = 0.009^\circ$ (10 km). Unfortunately, from the figure we see that the scaling law (1) is not verified for large θ .

At least we know the reason of this failure, which is that stationarity does not hold when we decrease L , at variance with the worldwide case studied in the previous section. Indeed, although the rate appears as stationary in the whole region of size $L = 10^\circ$, at smaller scales the fluctuations of the rate become important and invalidate our approach. So,

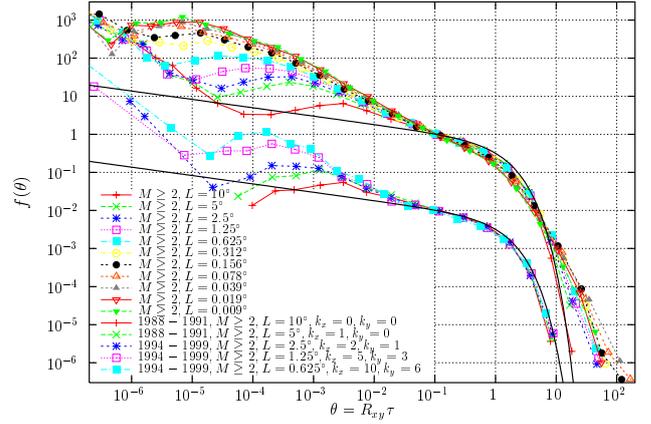


Fig. 7. Top curves: Probability densities of the rescaled recurrence times from the SCSN catalog, mixing data for different stationary time periods and different areas of the same size. In this case the scaling law (1) does not hold, as the stationarity of the process is not maintained at smaller length scales. Bottom curves: Probability densities of θ , shifted a factor 10^{-2} , now for single regions with stationary seismicity (no mixing). For each L , the regions are labeled by the discrete coordinates k_x, k_y , which take values from 0 to $10^\circ/L - 1$, and increase with longitude and latitude, respectively. For both sets of curves, the continuous lines represent the same function as in Fig. 3, with the same parameters.

the problem is that we are mixing regions that the majority of them are nonstationary. It would be necessary to look at each small region in detail to separate the stationary behavior from the nonstationary. However, although stationary regions are difficult to find, some of them appear in Fig. 7 (bottom set). In this case, the agreement with the scaling law (1) is total.

The a posteriori probability densities of γ for the stationary regions, which are shown in Fig. 8, support these conclusions, as the maximum of the distributions range from 0.63 to 0.72, and the probability of $\gamma > 1$ is practically 0. So, we conclude that the scaling law (1) works very well for stationary seismicity, and in the incoming section we will see how to overcome the difficulty of a nonstationary seismic rate.

6 Results for nonstationary seismic rate: seismicity after the Landers earthquake

The Landers earthquake, with magnitude $M = 7.3$ at the SCSN catalog, is the largest event in Southern California for several decades. It took place on 28 June 1992 at 34.12° N, 116.26° W. We will analyze the seismicity in Southern California after the Landers earthquake, which shows the usual behavior of seismicity after large shallow events: a sudden enormous increase in the number of earthquakes and a consequent slow decay with time. In our procedure, as in previous sections, we will not distinguish mainshocks from aftershocks, on the contrary, we will look at the complete seismicity in the region under study after the Landers “mainshock”.

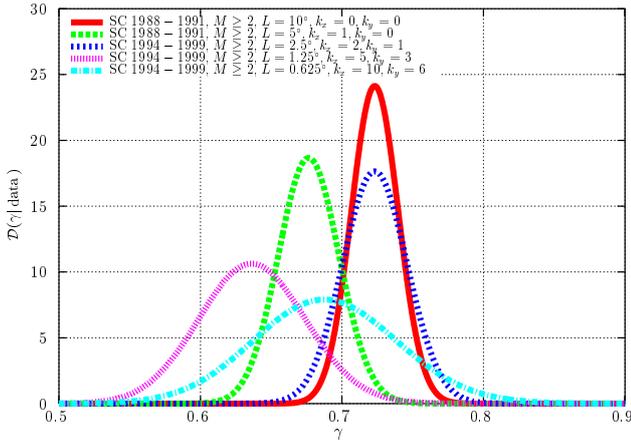


Fig. 8. A posteriori probability densities for the stationary periods and single regions corresponding to the bottom curves in the previous figure (note that there is no mixing of regions). In this case the agreement with the scaling law (1) is very good.

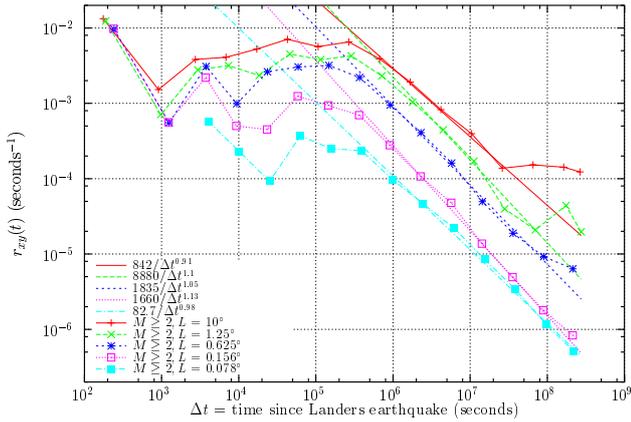


Fig. 9. Seismic rate $r_{xy}(t)$ after the Landers earthquake for different regions of size L including that event, using the SCSN catalog. Only events with $M \geq 2$ are considered. The straight lines correspond to power-law decays, compatible with the modified Omori formula. The values of p and K_{xy} (in s^{p-1}) appear in the legend; in all cases the exponent p is around 1.

Figure 9 shows the temporal decay of the “instantaneous” seismic rate, $r_{xy}(t)$, defined as the number of earthquakes per unit time for relatively short time windows. Regions of different size L are considered, all of them including the Landers earthquake; therefore, in this section we study a unique region for each system size, and in this way no mixing of data between regions of different coordinates is performed. When the logarithm of the rate versus the logarithm of the elapsed time since the mainshock Δt is plotted, a straight line appears for a certain period, corresponding to a power-law decay:

$$r_{xy}(t) = \frac{K_{xy}}{\Delta t^p}, \quad (21)$$

with $\Delta t \equiv t - t_{Landers}$, $t_{Landers} \simeq 1992.49$ years. This power law is of course the essence of the modified Omori law (Utsu

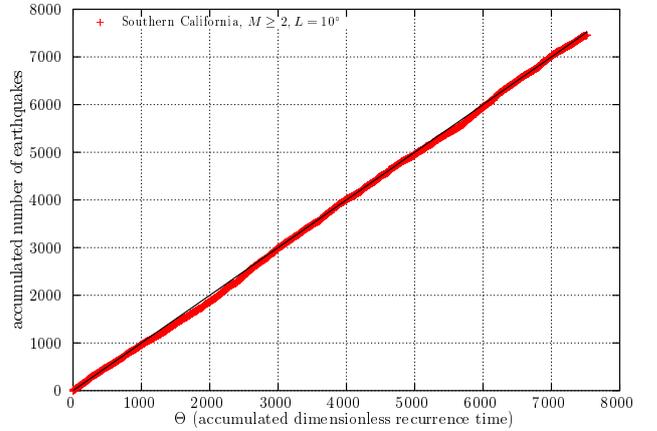
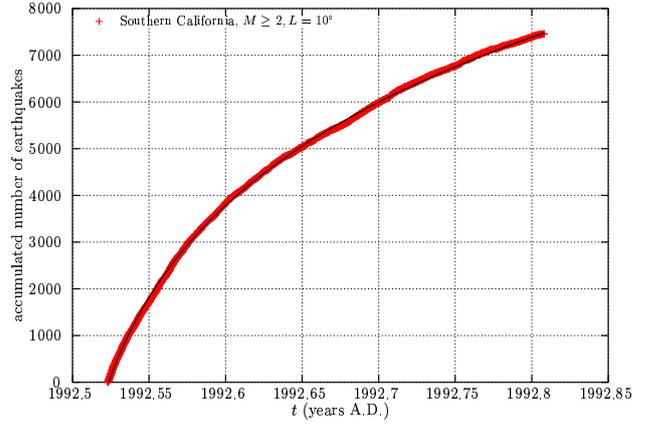


Fig. 10. (a) Accumulated number of earthquakes in Southern California during the power-law regime of the rate after the Landers event, using the SCSN catalog with $M \geq 2$. Note that, in years, the Landers earthquake took place at $t_{Landers} \simeq 1992.49$. The continuous line is the integral of the power-law rate corresponding to the modified Omori law, with $p=0.91$, $K_{xy}=842 s^{p-1}$ (see previous figure) and an integration constant $=-32519$. (b) Same data as in previous case, but transforming the x -axis to obtain a stationary process. The straight line is a linear relation with unity slope.

et al., 1995). Clearly, in this case the seismicity is not stationary; nevertheless, the approach explained in previous sections (Corral, 2004a) can be easily generalized. We only need to rescale the recurrence time as

$$\theta \equiv r_{xy}(t)\tau, \quad (22)$$

with $r_{xy}(t)$ given by the previous power law, in order to obtain the rescaled, dimensionless recurrence time θ .

The accumulated number of earthquakes, which is just the integral of $r_{xy}(t)$, is shown in Fig. 10a, only for the time period of power-law decay of the rate. The nonstationarity of the process is apparent; in contrast, Fig. 10b shows the accumulated number of earthquakes versus the accumulated rescaled recurrence time Θ , defined as $\Theta_i = \theta_1 + \theta_2 + \dots + \theta_i$. The clear straight line shows how we have accomplished the transformation of the process into a stationary one. The analysis of these data is now identical to the previous stationary cases. Moreover, the fact that we consider only the period of

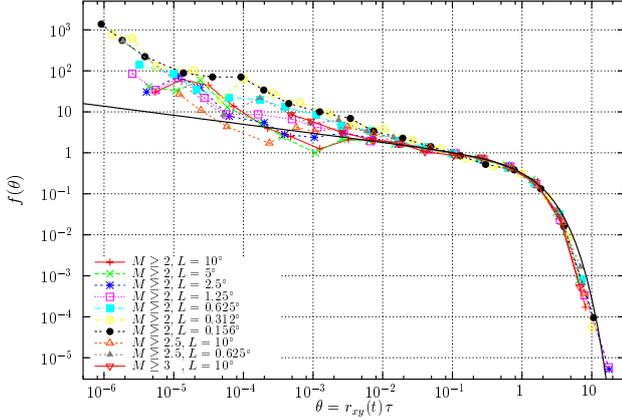


Fig. 11. Probability densities $f(\theta)$ of the rescaled recurrence time $\theta \equiv r_{xy}(t)\tau$, for different areas in Southern California after the Landers earthquake, using several magnitude thresholds M_c . The data comes from the SCSN catalog. The continuous curve is the same function as in Figs. 3 and 7. The results are in agreement with a secondary clustering structure inside the primary clustering.

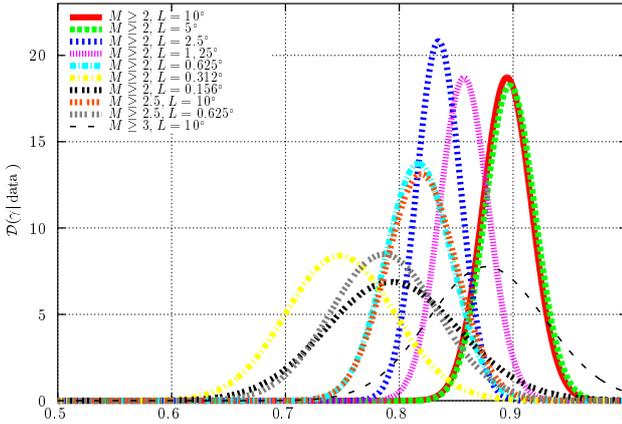


Fig. 12. A posteriori probability densities of the parameter γ for the same rescaled data after the Landers event as in the previous figure. The fact that γ is clearly distributed below 1 is unveils the secondary clustering structure inside the primary clustering.

power-law decay of the rate guarantees reasonably the completeness of the data; indeed, the deviations from the power law at small time intervals Δt are believed to be caused by the incompleteness of the record short after the mainshock.

The probability densities of the rescaled recurrence times appear in Fig. 11 in surprising agreement with the results for stationary seismicity. Therefore, note that this implies the existence of a secondary clustering structure inside the main sequence, due to the fact that any large aftershock may generate its own aftershocks (Ogata, 1999; Helmstetter and Sornette, 2002). What is remarkable is that this structure seems to be identical to the one corresponding to stationary seismicity (studied in the previous sections). The deviations from the scaling law at short rescaled times are a manifestation of the same phenomenon but on a different scale: also

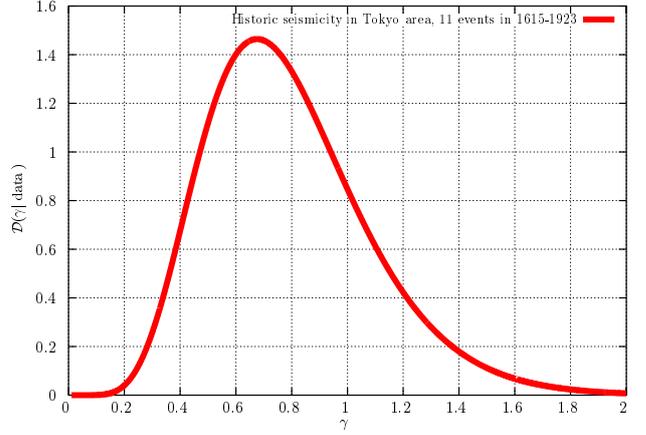


Fig. 13. A posteriori probability density of the parameter γ for the historic seismicity around Tokyo (Utsu, 1984). The clustering option is clearly the more likely (78% in front of 22% for nearly-periodicity).

in this case there are relatively important sequences of aftershocks, which our homogenization (performed by a simple power-law decay of the rate) has not taken into account.

The a posteriori distribution of the parameter γ also supports this conclusion, see Fig. 12. As a consequence of this, the assumption about the validity of the modified Omori law (in the limiting case of a power law) leads to the fact that aftershock sequences cannot be described as a nonhomogeneous Poisson process, as in that case one should obtain an exponential distribution for $f(\theta)$.

7 Results for historical earthquakes in the Tokyo area

In this last section of results, as a corollary, we work with a different kind of data. We use the list of strong earthquakes in the Tokyo area provided by Utsu (1984), consisting in 11 events, from 1615 to 1923. Our analysis differs from the one by Utsu in that, following our approach, we will consider the full data set. (Utsu considered that 3 events occurring within about 2 years constituted a correlated sequence, represented only by the largest event. Remember that from our point of view the correlations between events extend beyond the usual aftershock range.)

The a posteriori distribution for γ is shown in Fig. 13, where we have taken $\theta_{min}=0$. It is clear that the parameter γ may take any value in a broad range, but the probability of $\gamma < 1$, representing clustering, is significantly larger than the probability of $\gamma > 1$, which would be the signature of a nearly-periodic process (78% versus 22%). Of course, it is obvious that the inclusion of two recurrence times much smaller than the rest makes γ decrease, but it is remarkable that it decreases (in the most likely situation) below the limiting value $\gamma=1$, while the uncertainty of γ is not extremely large. The maximum value of the probability density, which would correspond with the maximum-likelihood estimation of γ if the priori were uniform, is $\gamma=0.68$; for comparison,

the mean value of γ turns out to be $\bar{\gamma}=0.79$; this difference is due to the asymmetry of $\mathcal{D}(\gamma| \text{data})$, in fact, we observe that the shape of $\mathcal{D}(\gamma| \text{data})$ is much more asymmetric than the shape obtained in the previous sections. In any case, the estimated value of γ is in surprising agreement with the previous cases.

8 Conclusions

We have taken advantage of the scaling properties of the recurrence time distribution to improve the statistics of occurrence of large earthquakes over small regions. In this way, we have found evidence of clustering for earthquakes with $M \geq 7.5$ in areas of size about 600 km, and for earthquakes with $M \geq 5.5$ in areas of about 20 km. If we could improve these results even further, they could be very relevant in the debate of seismic-gap models versus clustering occurrence (Kagan and Jackson, 1999).

It is remarkable that small earthquakes (let us say $M \geq 2$), in small areas lead to identical results than large events, if the seismicity is stationary. The difference is, however, that in the first case stationarity is much difficult to fulfill. Nevertheless, we provide a procedure to analyze cases in which the process is not stationary, if the seismic rate follows the Omori law. When seismicity is transformed into a stationary process we get the astonishing result that the properties of seismicity turn out to be the same as in the stationary case. Therefore, one may conclude that the behavior of seismicity is fully determined by the seismic rate, $r_{xy}(t)$, which is the only parameter of the process (Corral, 2004a).

This equivalence between big and small earthquakes, in large or reduced areas, and for stationary seismicity or sequences decaying as the Omori law has induced us to use the term universality. In fact, our plots of the a posteriori distributions of the parameter γ (the most important parameter of the distribution, measuring the degree of clustering or nearly-periodicity) show a certain scattering in its more likely value, as different values of γ are more appropriate for each set of values of M_c , L , and xy . Nevertheless, the a posteriori distributions overlap in most of the cases, or are close to overlap, and the distinction between different values of γ in the fit of the scaling function to the rescaled probability density is negligible. Therefore, we consider very useful to have a unified description of the earthquake occurrence phenomenon in terms of the universality concept, and, in order to reach some progress in the understanding of this complex problem, we take as a first step this unifying global picture, postponing the details for future works, when the important implications of these ideas have become clear. So, we have assumed the following philosophy: first consider the “physics of the 90%”, and only when the fundamentals have been grasped, undertake the study of the remaining 10%.

The scale-invariant properties of seismicity, exemplified by the scaling relation (1), supports the view of the seismic processes operating at a critical point (Bak, 1996; Jensen, 1998; Turcotte, 1999; Hergarten, 2002). Further research

in this line will clarify the relation between seismology and statistical physics, phase transitions, and critical phenomena. Indeed, the seminal research of Bak et al. (2002) has led, directly or indirectly, to a renewed interest in the statistics of time between earthquakes, which we hope will shed light on the complexity of this problem (Corral, 2003, 2004a, 2004b, 2004c, 2005; Mega et al., 2003; Baiesi and Paczuski, 2004; Baiesi, 2004; Helmstetter and Sornette, 2004; Molchan, 2005; Molchan and Kronrod, 2004; Scafetta and West, 2004; Davidsen and Goltz, 2004).

Acknowledgements. The author acknowledges the leading efforts of the late P. Bak to change the general view about earthquakes and to look at them from a unifying perspective. Interaction with M. Muncheta, S. Monguith, and M. de la Seu has allow the author a more efficient profitment of time. All this has been possible thanks to the Ramón y Cajal program and research projects BFM2003-06033 (MCyT) and 2001-SGR-00186 (DGRGC).

Edited by: C. Goltz

Reviewed by: two referees

References

- Baiesi, M.: Scaling and precursor motifs in earthquake networks, <http://arxiv.org/abs/cond-mat/0406198>, 2004.
- Baiesi, M. and Paczuski, M.: Scale-free networks of earthquakes and aftershocks, *Phys. Rev. E*, 69, 066106, 2004.
- Bak, P.: *How Nature Works: The Science of Self-Organized Criticality*, Copernicus, New York, 1996.
- Bak, P., Christensen, K., Danon, L., and Scanlon, T.: Unified scaling law for earthquakes, *Phys. Rev. Lett.* 88, 178501, 2002.
- Christensen, K., Danon, L., Scanlon, T., and Bak, P.: Unified scaling law for earthquakes, *Proc. Natl. Acad. Sci. USA* 99, 2509–2513, 2002.
- Corral, A.: Local distributions and rate fluctuations in a unified scaling law for earthquakes, *Phys. Rev. E* 68, 035102(R), 2003.
- Corral, A.: Long-term clustering, scaling, and universality in the temporal occurrence of earthquakes, *Phys. Rev. Lett.* 92, 108501, 2004a.
- Corral, A.: Universal local versus unified global scaling laws in the statistics of seismicity, *Physica A*, 340, 590–597, 2004b.
- Corral, A.: Correlations and invariance of seismicity under renormalization-group transformations, <http://arxiv.org/abs/cond-mat/0409124>, 2004c.
- Corral, A.: Time-decreasing hazard and increasing time until the next earthquake, *Phys. Rev. E*, 71, 017101, 2005.
- Cox, D. R. and Oakes, D.: *Analysis of Survival Data*, Chapman and Hall, London, 1984.
- Davidsen, J. and Goltz, C.: Are seismic waiting time distributions universal?, *Geophys. Res. Lett.*, accepted, 2004.
- Davis, P. M., Jackson, D. D., and Kagan, Y. Y.: The longer it has been since the last earthquake, the longer the expected time till the next? *Bull. Seism. Soc. Am.* 79, 1439–1456, 1989.
- Ellsworth, W. L., Matthews, M. V., Nadeau, R. M., Nishenko, S. P., Reasenber, P. A., and Simpson, R. W.: A physically-based earthquake recurrence model for estimation of long-term earthquake probabilities, U.S. Geological Survey Open-File Report 99-522, 1999.

- Gardner, J. K. and Knopoff, L.: Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian? *Bull. Seism. Soc. Am.* 64, 1363–1367, 1974.
- Gross, A. J. and Clark, V. A.: *Survival Distributions: Reliability Applications in the Biomedical Sciences*, Wiley, New York, 1975.
- Gutenberg, B. and Richter, C. F.: *Seismicity of the Earth*, Hafner Pub. Co., New York, 1965.
- Helmstetter, A. and Sornette, D.: Diffusion of epicenters of earthquake aftershocks, Omori's law, and generalized continuous-time random walk models, *Phys. Rev. E*, 66, 061 104, 2002.
- Helmstetter, A. and Sornette, D.: Comment on "Power-law Time distribution of large earthquakes", *Phys. Rev. Lett.*, 92, 129 801, 2004.
- Hergarten, S.: *Self-Organized Criticality in Earth Systems*, Springer, Berlin, 2002.
- Jaynes, E. T.: *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, 2003.
- Jensen, H. J.: *Self-Organized Criticality*, Cambridge University Press, Cambridge, 1998.
- Kagan, Y. Y.: Observational evidence for earthquakes as a nonlinear dynamic process, *Physica D*, 77, 160–192, 1994.
- Kagan, Y. Y.: Statistical aspects of Parkfield earthquake sequence and Parkfield prediction experiment, *Tectonophysics* 270, 207–219, 1997.
- Kagan, Y. Y. and Jackson, D. D.: New seismic gap hypothesis: five years after, *J. Geophys. Res. B*, 100, 3943–3959, 1995.
- Kagan, Y. Y. and Jackson, D. D.: Worldwide doublets of large shallow earthquakes, *Bull. Seismol. Soc. Amer.*, 89, 114–7-1155, 1999.
- Kerr, R. A.: Parkfield keeps secrets after a long-awaited quake, *Science* 306, 206–207, 2004.
- Knopoff, L.: Scale invariance of earthquakes, in *Scale Invariance and Beyond*, edited by: Dubrulle, B., Graner, F., and Sornette, D., 159–172, EDP Sciences – Springer, 1997.
- Koyama, J., Shaoxian, Z., and Ouchi, T.: Stochastic modeling of complex earthquake activity, in: *Advance in Mathematical Seismology*, edited by: Koyama, J. and Deyi, F., Seismological Press, Beijing, 1995.
- Loredo, T. J.: From Laplace to supernova SN 1987 A: Bayesian inference in astrophysics, in: *Maximum Entropy and Bayesian Methods*, edited by: Fougère, P. F., Kluwer Academic Publishers, Dordrecht, 1990.
- Mega, M. S., Allegrini, P., Grigolini, P., Latora, V., Palatella, L., Rapisarda, A., and Vinciguerra, S.: Power-law time distribution of large earthquakes, *Phys. Rev. Lett.*, 90, 188 501, 2003.
- Molchan, G.: Interevent time distribution in seismicity: a theoretical approach, *Pure Appl. Geophys.*, accepted, 2005.
- Molchan, G. and Kronrod, T.: On the frequency-magnitude law for fractal seismicity, <http://arxiv.org/abs/physics/0409142>, 2004.
- Mood, A. M., Graybill, F. A., and Boes, D. C.: *Introduction to the theory of statistics*, McGraw-Hill, New York, ed. 3, 1974.
- Murray, J. and Segall, P.: Testing time-predictable earthquake recurrence by direct measurement of strain accumulation and release, *Nature* 419, 287–291, 2002.
- Ogata, Y.: Seismicity analysis through point-process modeling: a review, *Pure Appl. Geophys.*, 155, 471–507, 1999.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P.: *Numerical Recipes in FORTRAN*, Cambridge University Press, Cambridge, ed. 2, 1992.
- Reasenber, P. A. and Jones, L. M.: Earthquake hazard after a mainshock in California, *Science*, 243, 1173–1176, 1989.
- Scafetta, N. and West, B. J.: Multiscaling comparative analysis of time series and a discussion on "earthquake conversations" in California, *Phys. Rev. Lett.*, 92, 138 501, 2004.
- Sieh, K.: The repetition of large-earthquake ruptures, *Proc. Natl. Acad. Sci. USA* 93, 3764–3771, 1996.
- Smalley, R. F., Chatelain, J.-L., Turcotte, D. L., and Prévot, R.: A fractal approach to the clustering of earthquakes: applications to the seismicity of the New Hebrides, *Bull. Seism. Soc. Am.*, 77, 1368–1381, 1987.
- Sornette, D. and Knopoff, L.: The paradox of the expected time until the next earthquake, *Bull. Seism. Soc. Am.*, 87, 789–798, 1997.
- Stein, R. S.: Characteristic or haphazard? *Nature*, 370, 443–444, 1995.
- Stein, R. S.: Parkfield's unfulfilled promise, *Nature*, 419, 257–258, 2002.
- Turcotte, D. L.: *Fractals and Chaos in Geology and Geophysics*, Cambridge University Press, Cambridge, ed. 2, 1997.
- Turcotte, D. L.: Self-organized criticality, *Rep. Prog. Phys.*, 62, 1377–1429, 1999.
- Turcotte, D. L. and Malamud, B. D.: Earthquakes as a Complex System, in: *International Handbook of Earthquake and Engineering Seismology, Part A*, edited by: Lee, W. H. K., Kanamori, H., Jennings, P. C., and Kisslinger, C., 209–227, Academic Press, 2002.
- Udías, A. and Rice, J.: Statistical analysis of microearthquake activity near San Andreas Geophysical Observatory, Hollister, California, *Bull. Seism. Soc. Am.*, 35, 809–827, 1975.
- Utsu, T.: Estimation of parameters for recurrence models of earthquakes, *Bull. Earthq. Res. Inst. Univ. Tokyo*, 59, 53–66, 1984.
- Utsu, T.: Representation and analysis of the earthquake size distribution: a historical review and some new approaches, *Pure Appl. Geophys.* 155, 509–535, 1999.
- Utsu, T.: Statistical features of seismicity, in: *International Handbook of Earthquake and Engineering Seismology, Part A*, edited by: Lee, W. H. K., Kanamori, H., Jennings, P. C., and Kisslinger, C., Academic Press, 719–732, 2002.
- Utsu, T., Ogata, Y., and Matsu'ura, R. S.: The centenary of the Omori formula for a decay law of aftershock activity, *J. Phys. Earth*, 43, 1–33, 1995.
- Wang, J.-H. and Lee, C.-W.: Fractal characterization of an earthquake sequence, *Physica A*, 221, 152–158, 1995.
- Wang, J.-H. and Kuo, C.-H.: On the frequency distribution of interoccurrence times of earthquakes, *J. Seism.*, 2, 351–358, 1998.