The occupation of a box as a toy model for the seismic cycle of a fault

Álvaro González[a)] and Javier B. Gómez[b)]
Departamento de Ciencias de la Tierra, Universidad de Zaragoza, C. Pedro Cerbuna 12, 50009 Zaragoza, Spain

Amalio F. Pacheco[c)]
Departamento de Física Teórica and BIFI, Universidad de Zaragoza, C. Pedro Cerbuna 12, 50009 Zaragoza, Spain

(Received 30 December 2004; accepted 10 July 2005)

We illustrate how a simple statistical model can describe the quasiperiodic occurrence of large earthquakes. The model idealizes the loading of elastic energy in a seismic fault by the stochastic filling of a box. The emptying of the box after it is full is analogous to the generation of a large earthquake in which the fault relaxes after having been loaded to its failure threshold. The duration of the filling process is analogous to the seismic cycle, the time interval between two successive large earthquakes in a particular fault. The simplicity of the model enables us to derive the statistical distribution of its seismic cycle. We use this distribution to fit the series of earthquakes with magnitude around six that occurred at the Parkfield segment of the San Andreas fault in California. Using this fit, we estimate the probability of the next large earthquake at Parkfield and devise a simple forecasting strategy.
[DOI: 10.1119/1.2013310]

I. INTRODUCTION

In reporting the mechanism of the great California earthquake of 1906, Reid[1] presented the elastic rebound theory. It asssumes that an earthquake is the result of a sudden relaxation of elastic strain by rupture along a fault (a rupture surface between two rock blocks that move past each other). This theory extended earlier insights into the relation between earthquakes and faults by other geologists,[2] especially Gilbert,[3,4] McKay,[5] and Koto.[6] Since its formulation, it has been the basis for interpreting the earthquakes that occur in faults in the Earth's upper, fragile crust.

According to Reid's theory, elastic energy slowly accumulates on a fault over a long time after the occurrence of an earthquake, as the rock blocks on both sides of the fault are strained by tectonic forces. When the strain is large enough, the system relaxes by fast rupture and/or frictional sliding along the fault during the next earthquake. The elastic waves generated by this sudden event are the seismic waves that seismometers detect.

The tectonic loading and relaxation process of a fault is cyclic. The seismic cycle is the time interval between two successive large earthquakes on the same fault, frequently called characteristic earthquakes.[7] If the seismic cycle were periodic, earthquake prediction would be easy. There is increasing information about earthquake occurrences in the seismic record, compiled with historical data and recognition of ancient large earthquakes on faults.[2] These data show that the seismic cycle of any given fault is not strictly periodic. The reason is that the tectonic loading and relaxation of a fault are complex nonlinear processes.[8] Moreover, faults occur in topologically complex networks,[9] and an earthquake occurring in a fault influences what occurs in other faults.[10]

The duration of the seismic cycle is not constant, but follows a statistical distribution that can be empirically deduced from the earthquake time series.[11] This distribution, if it were known, could be used to estimate the probability of the next earthquake. However, it is not well known, because there are little data (typically less than ten) in the earthquake time series for any given fault or fault segment.[11] To use this probabilistic approach, it is convenient to fit the data to a theoretical statistical distribution.

Especially since the 1970s,[12–18] earthquake recurrence is frequently considered as a renewal process,[19,20] in which the times between successive events, in this case the large earthquakes in a fault, are assumed to be independent and identically distributed random variables. In this interpretation, the expected time of the next event does not depend on the details of the last event, except the time it occurred. In combination with elastic rebound theory, the probability of another earthquake would be low just after a fault-rupturing earthquake, and would then gradually increase, as tectonic deformation slowly stresses the fault again. When an earthquake finally occurs, it resets the renewal process to its initial state. Several well-known statistical distributions (such as the gamma,[21] log-normal,[22] and Weibull[16,21,23,24]) have been used to describe the duration of the seismic cycle and to calculate the conditional probabilities of future earthquakes. These distributions also have been used as failure models for reliability and time-to-failure problems.[25]

More recently, many numerical models have been devised for simulating the tectonic processes occurring on a seismic fault.[26,27] These models can generate as many synthetic earthquakes as desired,[28] so the statistical distribution of the time intervals between them can be fully characterized.[29] Two highly idealized models are the Brownian passage time model,[30] and the minimalist model.[31–33] Their seismic cycle distributions have been used as renewal models, to fit actual earthquake series and estimate future earthquake probabilities.[30,33,34] They, as well as the gamma, log-normal, and Weibull distributions, provide a reasonably good fit to the existing data.[21,33,34] The renewal models have been widely applied, particularly in Japan[35] and in the United States,[36] to estimate the probabilities of the next large earthquake for particular faults.

This paper aims to explain how a renewal model can fit the series of seismic cycles in a particular fault, and how it can be used to estimate the probability of the next large earthquake in the fault. For this purpose we will use the process of stochastic occupation of a box to visualize the progressive loading of a seismic fault. This box model will be used to fit the series of characteristic earthquakes, with magnitude around 6, which have occurred on the Parkfield segment of the San Andreas fault in California.

In Sec. II we present the data of the Parkfield series, and compute its mean, standard deviation, and aperiodicity (coefficient of variation). The presentation of these data is important for appreciating the design and tuning of the subsequent model. Section III is devoted to a detailed presentation of the box model. In Sec. IV the parameters of the model will be tuned to fit the Parkfield data series. The comparison between the model and the data is made in Sec. V, and the annual probability of occurrence of the next large shock at Parkfield is calculated. In Sec. VI we introduce a simple forecasting strategy for the box model and illustrate its effectiveness for the Parkfield sequence. In Sec. VII we present our conclusions.

II. THE PARKFIELD SERIES

The San Andreas fault runs for 1200 km, from the Gulf of California (Mexico) to just north of San Francisco, where it enters the Pacific Ocean. Fortunately, it does not slide or break in its whole length as a single earthquake. Rather, as for other long faults, each earthquake ruptures only one or a few sections (segments) of its length. During the last century and a half, several earthquakes with magnitude around 6 have occurred along a 35-km-long segment of the San Andreas fault that crosses the tiny town of Parkfield, CA. The apparent temporal regularity of this series has lead to extensive seismic monitoring in the area.[37,38] Including the most recent event, this Parkfield series[37,39] consists of seven shocks which occurred on 9 January 1857; 2 February 1881; 3 March 1901; 10 March 1922; 8 June 1934; 28 June 1966; and 28 September 2004. The durations (in years) of the six observed seismic cycles are $c_1=24.07$, $c_2=20.08$, $c_3=21.02$, $c_4=12.25$, $c_5=32.05$, and $c_6=38.25$. The mean of this series is

$$m=\frac{1}{6}\sum_{i=1}^{6} c_i = 24.62 \text{ yr}, \tag{1}$$

and its sample standard deviation (square root of the bias-corrected sample variance) is

$$s=\left[\frac{1}{6-1}\sum_{i=1}^{6}(c_i-m)^2\right]^{1/2}=9.25 \text{ yr}. \tag{2}$$

The coefficient of variation, or aperiodicity, is

$$\alpha=\frac{s}{m}=0.3759. \tag{3}$$

III. THE BOX MODEL

In this section we will introduce a renewal model based on a simple cellular automaton. Cellular automata models are frequently used to model earthquakes and other natural hazards.[40] These models evolve in discrete time steps, and consist of a grid of cells, where each cell can be only in a finite number of states. Each cell's state is updated at each time step according to rules that usually depend on the state of the cell or that of its neighbors in the previous time step. For example, a grid of cells can represent a discretized fault plane, and the rules can be designed according to certain friction laws,[27] and include stress transfer[8,41,42] and the mechanical effects of fluids.[43] In the simplest models,[31,44] these details are ignored: the model is driven stochastically, there are only two possible states for each cell, and the earthquakes are generated according to simplified breaking rules. The model proposed here is of this last type. It is simple, easy to describe analytically, and generates a temporal distribution of seismic cycles comparable to that of an actual fault.

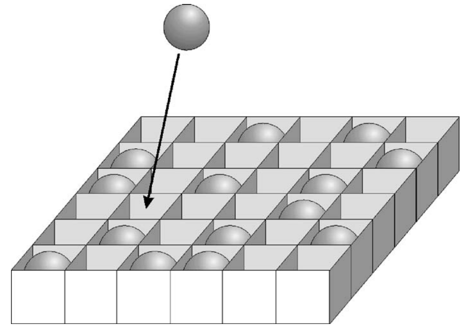

Fig. 1. Sketch of the box model. Balls are thrown at random until all the cells of the box are full. Then the box is emptied and a new cycle starts.

A. The rules

Consider an array of N cells. The position of the cells is irrelevant, but we can assume that they are arranged in the shape of a box (see Fig. 1). At the beginning of each cycle, the box is completely empty. At each time step, one ball is thrown, at random, to one of the cells in the box. That is, each cell has equal probability, $1/N$, of receiving the ball. If the cell that is chosen is empty, it will become occupied. If it was already occupied, the thrown ball is lost. (Thus, each cell can be either occupied by a ball or empty.) When a new throw completes the occupation of the N cells of the box, it topples, becoming completely empty, and a new cycle starts. The time elapsed since the beginning of each cycle, expressed by the number of thrown balls, will be called n. The duration of the cycles is statistically distributed according to a discrete distribution function $P_N(n)$.

The box represents the area of the fault or fault segment, and the random throwing of balls represents the increase of regional stress. This randomness is a way of dealing with the complex stress increase on actual faults. The occupation of a cell by a ball stands for the elastic strain induced by the stress in a patch or element of the fault plane. The loss of the balls that land on already occupied cells mimics stress dissipation on this plane. The total elastic strain (or conversely the total potential elastic energy) accumulated in the fault is represented by the number of occupied cells. This number gradually grows up to the limit N (analogous to the failure threshold of the fault), and the toppling of the box represents the occurrence of the characteristic earthquake in the fault. It is easy to simulate the evolution of this system with a Monte Carlo approach.

This model is similar to that introduced by Newman and Turcotte in Ref. 44. The difference is that their model is a square grid of cells in which the topology is relevant: they consider that the characteristic earthquake occurs when a percolating cluster[45] spans the grid. This cluster happens before the grid is completely full.

B. Some formulas of the box model

To describe the box model analytically, it is convenient to recall some elements of the geometric distribution. Consider the probability that exactly x independent Bernoulli trials, each with a probability of success p, will be required until the first success is achieved. The probability that $(x-1)$ failures will be followed by a success is $(1-p)^{x-1}p$. The resulting probability function,

$$f(x;p) = (1-p)^{x-1}p, \tag{4}$$

is known as the geometric distribution. Its mean and variance are

$$\langle x \rangle = \frac{1}{p}, \quad \sigma^2 = \frac{1-p}{p^2}. \tag{5}$$

We now consider the box model further. In each cycle, the filling of the box proceeds sequentially and continues until the Nth cell is occupied. Because each of these sequential steps is an independent process, the mean number of throws to completely fill the box will be

$$\langle n \rangle_N = \langle x_1 \rangle_N + \langle x_2 \rangle_N + \cdots + \langle x_N \rangle_N, \tag{6}$$

where $\langle x_i \rangle_N$ is the mean number of throws it takes to fill the ith cell.

Because the filling of the ith cell is geometrically distributed with $p_i = (N+1-i)/N$, it follows that

$$\langle x_i \rangle_N = \frac{N}{N+1-i} \quad (i = 1,2,\ldots,N) \tag{7}$$

and therefore

$$\langle n \rangle_N = 1 + \sum_{i=2}^{N} \frac{N}{N+1-i}. \tag{8}$$

Relations similar to Eqs. (6) and (8) can be written for the variance of the number of thrown balls to fill the box, namely

$$\sigma_N^2 = \sigma_1^2 + \sigma_2^2 + \cdots = 0 + \sum_{i=2}^{N} \frac{1 - \dfrac{N+1-i}{N}}{\left(\dfrac{N+1-i}{N}\right)^2}, \tag{9}$$

and, consequently, the standard deviation is

$$\sigma_N = \left[\sum_{i=2}^{N} \frac{N(i-1)}{(N+1-i)^2} \right]^{1/2}. \tag{10}$$

The aperiodicity of the series, α_N, for a given N is

$$\alpha_N = \frac{\sigma_N}{\langle n \rangle_N}. \tag{11}$$

The mean and the standard deviation of the box model can be calculated by summing the $N-1$ terms of Eqs. (8) and (10). For $N \geq 10$, these two equations can be approximated (with an absolute error <0.01) by their asymptotic expressions:[46]

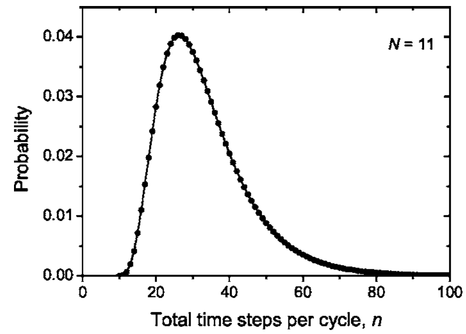

Fig. 2. Discrete distribution function for the duration (in time steps, n) of the seismic cycle in the box model with $N=11$.

$$\langle n \rangle_N \xrightarrow[N\to\infty]{} N(C + \ln N) + \tfrac{1}{2}, \tag{12}$$

where $C \simeq 0.577\,215\,7$ is Euler's constant, and

$$\sigma_N \xrightarrow[N\to\infty]{} N\left[\frac{\pi^2}{6} - \frac{1 + C + \ln N}{N} \right]^{1/2}, \tag{13}$$

and the aperiodicity can be estimated by using Eq. (11) with Eqs. (12) and (13).

The function $P_N(n)$ is not as easy to obtain as its mean and standard deviation. It is given by

$$P_N(n) = \sum_{j=1}^{N-1} (-1)^{j+1} \binom{N-1}{j-1} \left(1 - \frac{j}{N}\right)^{n-1}, \tag{14}$$

and the accumulative probability function, $A_N(n)$, is

$$A_N(n) = \sum_{j=N}^{n} P_N(j) = 1 - \sum_{j=1}^{N-1} (-1)^{j+1} \binom{N-1}{j-1} \left(1 - \frac{j}{N}\right)^n \frac{N}{j}. \tag{15}$$

We have deduced Eq. (14) by means of a Markov chain approach analogous to the one used in Refs. 31 and 32. This derivation is omitted here because of its length.

IV. FITTING THE PARAMETERS OF THE BOX MODEL

We will fit the Parkfield series to the accumulative probability function, Eq. (15), using the method of moments.[21] Another method that could be used is that of maximum likelihood.[21] We have seen in Sec. III that the aperiodicity in the box model depends only on N. Thus, we will choose N for which the aperiodicity is the closest to that of the Parkfield series, that is, $\alpha \simeq 0.3759$. The result is $N=11$, for which, from Eq. (11), the aperiodicity is $\alpha = 0.3752$.

From Eq. (8) the mean value of n for $N=11$ is $\langle n \rangle_{N=11} = 33.22$. Because the actual mean of the Parkfield series is $m = 24.62$ yr, one ball throw in the model is equivalent to $\tau = 24.62\text{ yr}/33.22 = 0.74\text{ yr} \approx 9$ months. The discrete distribution function for the duration of the seismic cycle in a box model with $N=11$, $P_{11}(n)$ is shown in Fig. 2.

In Fig. 3 we plot the evolution of the number of occupied cells for ten cycles with $N=11$. Note the fluctuations in the duration of the cycles, which are consistent with the mean and the standard deviation of the series. Note also that the occupation increases rapidly just after a toppling, and then slows down. This increase is due to the fact that, as a cycle

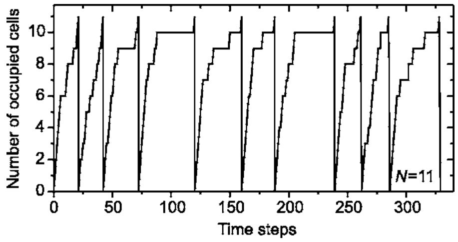

Fig. 3. Plot of the number of occupied cells during ten cycles of a box model with $N=11$.

progresses, there are more occupied cells, and thus it is less probable for an incoming ball to land on an empty cell. If ρ_n is the fraction of occupied cells at time step n, there is a probability $1-\rho_n$ for the next ball to be thrown to an empty cell. Because such a throw would increase ρ by $1/N$, the mean ρ at step $n+1$ is

$$\langle \rho_{n+1} \rangle = \langle \rho_n \rangle + \frac{1}{N}[1 - \langle \rho_n \rangle]. \tag{16}$$

The box is empty at the beginning of the cycle ($\rho_0=0$), so from Eq. (16), the mean value of ρ_n is

$$\langle \rho_n \rangle = 1 - \left(1 - \frac{1}{N}\right)^n, \tag{17}$$

which approaches one asymptotically.

In real faults, the strain also does not increase uniformly during the seismic cycle. Instead, it follows a trend similar to that of the number of occupied cells in the box model: the loading rate is faster just after a large earthquake, and decreases over time.[47]

The relaxation of a real fault by means of a large earthquake reduces the stress in the system. Thus a minimum time has to elapse before the fault accumulates enough stress to produce the next large earthquake. This effect is called stress shadow.[10] In the box model there exists a stress shadow: a characteristic earthquake cannot occur until the Nth step in the cycle. According to the box model, this minimum time for the Parkfield series is $\tau N \simeq 8$ yr.

V. EARTHQUAKE PROBABILITIES AT PARKFIELD

We now evaluate the quality of the box model fit for the Parkfield series and estimate the probability of the next earthquake in this fault segment. In Fig. 4(a), the empirical distribution function of the Parkfield series is plotted. It is an accumulative step function ranging from 0 to 1.0, with a jump 1/6 at each of the six observed recurrence intervals c_i. The accumulated distribution of the box model in Eq. (15) for $N=11$ with $\tau=0.74$ yr also is drawn. In Fig. 4(b), we show the residuals of the fit, which do not surpass 7.5%. The equivalent fits to these data, made by using the renewal models cited in Sec. I, give very similar results.[34]

Now we calculate the yearly probability of the next earthquake, that is, the conditional probability of the next shock occurring in a certain year, given that it has not occurred previously. For the box model it is

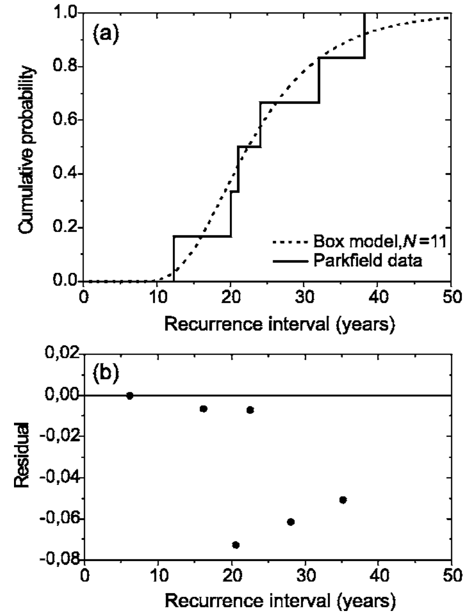

Fig. 4. (a) Fit of the accumulative distribution of the box model to the accumulated histogram of the Parkfield earthquake sequence. (b) Residuals of the fit, evaluated at the midpoints of the horizontal segments of the accumulated histogram.

$$P_\tau(N,n) = \frac{A_N(n+1/\tau) - A_N(n)}{1 - A_N(n-1)}. \tag{18}$$

Note that $1/\tau$ is the number of time steps of the box model corresponding to one year. After calculating P_τ from Eq. (18), it is necessary to rescale the abscissas, n, to actual years, $n\tau + t_0$, where t_0 is the calendar year at which the last earthquake occurred ($t_0=2004.75$ for the Parkfield series). In Fig. 5 we plot the yearly probability for the new cycle at Parkfield according to the box model. During the first eight years after the last earthquake at Parkfield (which occurred in September 2004), the box model indicates that another big shock should not be expected. From that time on, the probability of the next earthquake increases, tending to a constant equal to 11%.

In the seismological literature there is a well-known question about the yearly probability for a time much longer than the mean value of the series:[48] "The longer it has been since the last earthquake, the longer the expected time till the

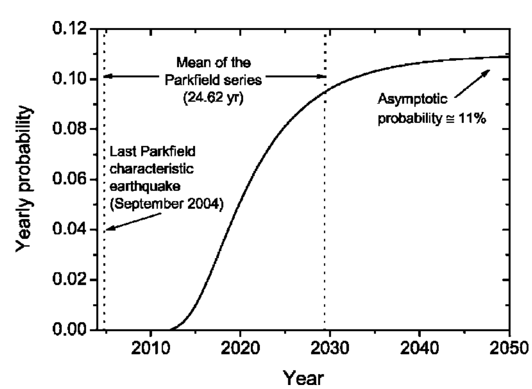

Fig. 5. Yearly probability of the next characteristic earthquake at Parkfield, according to the box model.

next?" Sornette and Knopoff[49] have discussed some statistical distributions that lead to affirmative, negative, or neutral answers to it. The result shown in Fig. 5 leads us to conclude that the box model produces a neutral answer. The reason is that for a long cycle duration (large n), the $P_N(n)$ of the box model decays exponentially, and asymptotically the box model behaves as a Poisson model, in which the conditional probability of occurrence of the next earthquake is a constant.

VI. A SIMPLE FORECASTING STRATEGY

In earthquake forecasting an "alarm" is sometimes turned on when it is estimated that there is a high probability for a large earthquake to occur.[50] If a large shock takes place when the alarm is on, the prediction is considered to be a success. If it takes place when the alarm is off, there has been a failure to predict. The fraction of errors, f_e, is the number of prediction failures divided by the total number of large earthquakes. The fraction of alarm time, f_a, is the ratio of the time during which the alarm is on to the total time of observation. A good strategy of forecasting must produce both small f_e and f_a, because both the prediction failures and the alarms are costly. Depending on the trade-off between the costs and benefits of forecasting,[51] we can try to minimize a certain loss function, L. For simplicity, we will consider a simple loss function defined as

$$L = f_a + f_e. \tag{19}$$

A random guessing strategy (randomly turning the alarm on and off) will yield $L=1$, a result which can be easily understood. The alarm will be on, randomly, during a certain fraction of time, f_a. Thus, there will be a probability equal to f_a for it being on when an earthquake eventually occurs (and a probability of $1-f_a$ for it being off). The result is that $f_e = 1-f_a$. As a trivial special case, if the alarm is always on ($f_a=1$), then all the earthquakes are "forecasted" ($f_e=0$). Conversely, all the earthquakes are failures to predict if the alarm is always off. The random guessing strategy is considered as a baseline, so a forecasting procedure makes sense only if it gives $f_a+f_e<1$.

We can use the box model fit to the Parkfield series to implement a simple earthquake forecasting strategy. The strategy consists of turning on the alarm at a fixed value of n time steps after the last big earthquake and maintaining the alarm on until the next one.[32,44] Then the alarm is turned off and the same strategy is repeated, evaluating f_a and f_e for all the cycles. The result is

$$f_e(n) = \sum_{n'=1}^{n} P(n'), \tag{20}$$

and

$$f_a(n) = \frac{\sum_{n'=n}^{\infty} P(n')(n'-n)}{\sum_{n'=0}^{\infty} P(n')n'}. \tag{21}$$

These relations are illustrated in Fig. 6(a), together with $L = f_a+f_e$. For each value of N, $L(n)$ has a minimum at a specific value of n, $n^*(N)$. As can be seen in Fig. 6(a), $n^*(11) = 20$, for which

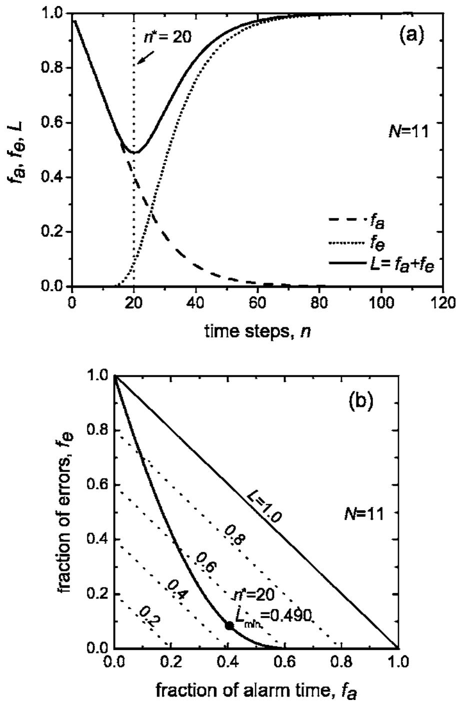

Fig. 6. (a) Fraction of errors (f_e), fraction of alarm time (f_a), and loss function ($L=f_a+f_e$) as a function of n for the forecasting strategy in a box model with $N=11$. (b) Error diagram for this strategy. Each point on the curve is the result of using a different value of n. The large dot corresponds to n^*, for which the loss function reaches a minimum. The diagonal lines are isolines of L. A random guessing strategy would render $L=1$.

$$f_a(n^*) = 0.405, \quad f_e(n^*) = 0.085, \quad L(n^*) = 0.490. \tag{22}$$

For the Parkfield sequence, n^* corresponds to

$$\tau n^* = 14.8 \text{ yr}. \tag{23}$$

If the distribution derived from the box model correctly describes the recurrence of large earthquakes at Parkfield, the alarm connected at this time since the beginning of the cycles and disconnected just after the occurrence of each shock would yield the results given in Eq. (22). Note that this time is substantially smaller than the mean duration of the cycles, $m=24.62$ yr.

The quality of the model-earthquake forecast also can be understood visually by means of an error diagram, in which f_e is plotted versus f_a.[51] This kind of plot is similar to the receiver operating characteristic diagram, used, for example, to test the success of weather forecasts.[52]

VII. SUMMARY

The generation of large earthquakes in a seismic fault involves slow loading of elastic strain (or, conversely, elastic energy), and release through rupture and/or frictional sliding during an earthquake. The duration of this process is the seismic cycle, which is repeated indefinitely, leading to a series of recurrent shocks. We have illustrated this process with a very simple model. The loading of elastic strain is represented by the stochastic filling of a box with N cells. The emptying of the box after its complete filling is analogous to the generation of a large earthquake, in which the fault relaxes after having being loaded to its failure threshold. The duration of the filling process is thus equivalent to the seismic cycle.

The statistical distribution of seismic cycles in the box model (just as the distributions of the Brownian passage time model[30] and the minimalist model[31–33]) can be used to fit actual earthquake series and to estimate earthquake probabilities. The conditional probability of the box model has two interesting features. First, after a large earthquake, there is a period of stress shadow during which a new large earthquake cannot occur. Second, from this time on the probability continuously increases, approaching a constant asymptotic value. By using a simple forecasting strategy, we have shown that the earthquakes in the model are predictable to some extent.

We hope that our discussion will be a useful educational tool for introducing several important geophysical and statistical concepts to graduate and undergraduate students. It could illustrate how to make quantitative estimates of a natural phenomenon as popular and as mystifying as earthquakes.

ACKNOWLEDGMENTS

AFP thanks Jesús Asín, Marisa Rezola, Leandro Moral, and José G. Esteve for useful comments. This research was funded by the Spanish Ministry of Education and Science, through Project No. BFM2002-01798 and Research Grant No. AP2002-1347 held by ÁG.

[a]Electronic mail: alvaro.gonzalez@unizar.es
[b]Electronic mail: jgomez@unizar.es
[c]Electronic mail: amalio@unizar.es

[1] Harry Fielding Reid, "The mechanics of the earthquake," in *The California Earthquake of April 18, 1906, Report of the State Earthquake Investigation Comission* (Carnegie Institution, Washington, DC, 1910), Vol. 2, pp. 1–192.

[2] Robert S. Yeats, Kerry Sieh, and Clarence R. Allen, *The Geology of Earthquakes* (Oxford U.P., New York, 1997).

[3] Grove Karl Gilbert, "A theory of the earthquakes of the Great Basin, with a practical application," Am. J. Sci. Ser. 3 **27**(157), 49–54 (1884).

[4] Grove Karl Gilbert, "Earthquake forecasts," Science **XXIX**(734), 121–138 (1909).

[5] Alexander McKay, "On the earthquakes of September 1888, in the Amuri District," N. Z. Geol. Surv. Rep. Geol. Explor. 1888-1889 20, 1–16 (1890).

[6] Bunjiro Koto, "On the cause of the great earthquake in central Japan, 1891," J. Coll. Sci., Imp. Univ. Tokyo **5**, 296–353 (1893).

[7] David P. Schwartz and Kevin J. Coppersmith, "Fault behavior and characteristic earthquakes: Examples from the Wasatch and San Andreas fault zones," J. Geophys. Res. **89**(B7), 5681–5698 (1984).

[8] *GeoComplexity and the Physics of Earthquakes*, edited by John B. Rundle, Donald L. Turcotte, and William Klein (American Geophysical Union, Washington, DC, 2000).

[9] Eric Bonnet, Olivier Bour, Noelle E. Odling, Philippe Davy, Ian Main, Patience Cowie, and Brian Berkowitz, "Scaling of fracture systems in geological media," Rev. Geophys. **39**(3), 347–383 (2001).

[10] Ruth A. Harris, "Earthquake stress triggers, stress shadows, and seismic hazard," Curr. Sci. **79**(9), 1215–1225 (2000).

[11] James C. Savage, "Empirical earthquake probabilities from observed recurrence intervals," Bull. Seismol. Soc. Am. **84**(1), 219–221 (1994).

[12] David Vere-Jones, "Stochastic models for earthquake occurrence (with discussion)," J. R. Stat. Soc. Ser. B. Methodol. **32**, 1–62 (1970).

[13] Tokuji Utsu, "Large earthquakes near Hokkaido and the expectancy of the occurrence of a large earthquake of Nemuro," Rep. Coord. Comm. Earthq. Predict. **7**, 7–13 (1972).

[14] Tokuji Utsu, "Aftershocks and earthquake statistics (IV)," J. Fac. Sci. Hokkaido Univ. Ser. VII **4**, 1–42 (1972).

[15] Tsuneji Rikitake, "Probability of earthquake occurrence as estimated from crustal strain," Tectonophysics **23**(3), 299–312 (1974).

[16] Yukio Hagiwara, "Probability of earthquake occurrence as obtained from a Weibull distribution analysis of crustal strain," Tectonophysics **23**(3), 313–318 (1974).

[17] Tsuneji Rikitake, "Recurrence of great earthquakes at subduction zones," Tectonophysics **35**(4), 335–362 (1976).

[18] David Vere-Jones, "Earthquake prediction: A statistician's view," J. Phys. Earth **26**, 129–146 (1978).

[19] Erhan Çinlar, *Introduction to Stochastic Processes* (Prentice Hall, New York, 1975).

[20] Daryl J. Daley and David Vere-Jones, *An Introduction to the Theory of Point Processes* (Springer-Verlag, New York, 2003), 2nd ed.

[21] Tokuji Utsu, "Estimation of parameters for recurrence models of earthquakes," Bull. Earthquake Res. Inst., Univ. Tokyo **59**, 53–66 (1984).

[22] Stuart P. Nishenko and Raymond Buland, "A generic recurrence interval distribution for earthquake forecasting," Bull. Seismol. Soc. Am. **77**(4), 1382–1399 (1987).

[23] Anne S. Kiremidjian and Thalia Anagnos, "Stochastic slip-predictable model for earthquake occurrences," Bull. Seismol. Soc. Am. **74**(2), 739–755 (1984).

[24] Kerry Sieh, Minze Stuiver, and David Brillinger, "A more precise chronology of earthquakes produced by the San Andreas fault in southern California," J. Geophys. Res. **94**(B1), 603–623 (1989).

[25] Nancy R. Mann, Ray E. Schafer, and Nozer D. Singpurwalla, *Methods for Statistical Analysis of Reliability and Life Data* (Wiley, New York, 1974).

[26] Ian Main, "Statistical physics, seismogenesis, and seismic hazard," Rev. Geophys. **34**(4), 433–462 (1996).

[27] Yehuda Ben-Zion, "Dynamic ruptures in recent models of earthquake faults," J. Mech. Phys. Solids **49**(9), 2209–2244 (2001).

[28] Steven N. Ward, "San Francisco Bay area earthquake simulations: A step toward a standard physical earthquake model," Bull. Seismol. Soc. Am. **90**(2), 370–386 (2000).

[29] Russell Robinson, "Potential earthquake triggering in a complex fault network: The Northern South Island, New Zealand," Geophys. J. Int. **159**(2), 734–748 (2004).

[30] Mark V. Matthews, William L. Ellsworth, and Paul A. Reasenberg, "A Brownian model for recurrent earthquakes," Bull. Seismol. Soc. Am. **92**(6), 2233–2250 (2002).

[31] Miguel Vázquez-Prada, Álvaro González, Javier B. Gómez, and Amalio F. Pacheco, "A minimalist model of characteristic earthquakes," Nonlinear Processes Geophys. **9**(5-6), 513–519 (2002).

[32] Miguel Vázquez-Prada, Álvaro González, Javier B. Gómez, and Amalio F. Pacheco, "Forecasting characteristic earthquakes in a minimalist model," Nonlinear Processes Geophys. **10**(6), 565–571 (2003).

[33] Javier B. Gómez and Amalio F. Pacheco, "The minimalist model of characteristic earthquakes as a useful tool for description of the recurrence of large earthquakes," Bull. Seismol. Soc. Am. **95**(5), 1960–1967 (2004).

[34] Álvaro González, Javier B. Gómez, and Amalio F. Pacheco, "Updating seismic hazard at Parkfield," J. Seismol., in press, physics/0411003.

[35] Masajiro Imoto, "Probability gains expected for renewal process models," Earth, Planets Space **56**(6), 563–571 (2004).

[36] Working Group on California Earthquake Probabilities, "Earthquake probabilities in the San Francisco Bay Region: 2002-2031," Open-File Report 03-214 (United States Geological Survey, Menlo Park, CA, 2003).

[37] William H. Bakun and Allan G. Lindh, "The Parkfield, California, earthquake prediction experiment," Science **229**, 619–624 (1985).

[38] Evelyn Roeloffs, "The Parkfield, California earthquake experiment: An update in 2000," Curr. Sci. **79**(9), 1226–1236 (2000).

[39] William H. Bakun, "History of significant earthquakes in the Parkfield area," Earthq. Volcano. 20(2), 45–51 (1988).

[40] Bruce D. Malamud and Donald L. Turcotte, "Cellular-automata models applied to natural hazards," Comput. Sci. Eng. **2**(3), 42–51 (2000).

[41] Zeev Olami, Hans Jacob S. Feder, and Kim Christensen, "Self-organized criticality in a continuous, nonconservative cellular automaton modeling earthquakes," Phys. Rev. Lett. **68**(8), 1244–1247 (1992).

[42] Eric F. Preston, Jorge S. Sá Martins, John B. Rundle, Marian Anghel, and William Klein, "Models of earthquake faults with long-range stress transfer," Comput. Sci. Eng. **2**(3), 34–41 (2000).

[43] Stephen A. Miller, "Properties of large ruptures and the dynamical influence of fluids on earthquakes and faulting," J. Geophys. Res. **107**, 2182 (2002).

[44] William I. Newman and Donald L. Turcotte, "A simple model for the earthquake cycle combining self-organized complexity with critical point

behavior," Nonlinear Processes Geophys. **9**(5-6), 453–461 (2002).
[45] Dietrich Stauffer and Amnon Aharony, *Introduction to Percolation Theory* (Taylor and Francis, London, 1994), 2nd ed.
[46] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1965).
[47] Andrew J. Michael, "Viscoelasticity, postseismic slip, fault interactions, and the recurrence of large earthquakes," EOS Trans. Am. Geophys. Union **84**(46), Fall Meet. Suppl., Abstract S31G-04 (2003). The paper corresponding to this abstract has been submitted to Bull. Seismol. Soc. Am.
[48] Paul M. Davis, David D. Jackson, and Yan Y. Kagan, "The longer it has been since the last earthquake, the longer the expected time till the next?," Bull. Seismol. Soc. Am. **79**(5), 1439–1456 (1989).
[49] Didier Sornette and Leon Knopoff, "The paradox of the expected time until the next earthquake," Bull. Seismol. Soc. Am. **87**(4), 789–798 (1997).
[50] *Nonlinear Dynamics of the Lithosphere and Earthquake Prediction*, edited by Vladimir I. Keilis-Borok and Alexandre A. Soloviev (Springer-Verlag, Berlin, Germany, 2003).
[51] George M. Molchan, "Earthquake prediction as a decision-making problem," Pure Appl. Geophys. **149**(1), 233–247 (1997).
[52] Ian B. Mason, "Binary forecasts," in *Forecast Verification: A Practitione's Guide in Atmospheric Science*, edited by Ian T. Joliffe and David B. Stephenson (Wiley, West Sussex, UK, 2003), pp. 37–76 (references on pp. 215–226).

Erratum: "The occupation of a box as a toy model for the seismic cycle of a fault" [Am. J. Phys. 73 (10), 946–952 (2005)]

Álvaro González[a)] and Javier B. Gómez[b)]
Departamento de Ciencias de la Tierra, Universidad de Zaragoza, C. Pedro Cerbuna 12, 50009 Zaragoza, Spain

Amalio F. Pacheco[c)]
Departamento de Física Teórica and BIFI, Universidad de Zaragoza, C. Pedro Cerbuna 12, 50009 Zaragoza, Spain

(Received 20 November 2006; accepted 1 December 2006)
[DOI: 10.1119/1.2426356]

The values of f_e plotted in Fig. 6 and used to estimate n^* were determined by a Monte Carlo simulation of the model instead of using Eq. (20). If, at a given time step in the simulation, the "alarm" was sounded and the model earthquake occurred, the latter was deemed as successfully forecasted. This assumption is incorrect, and leads to a value of f_e smaller than the true one in Eq. (20). Given that n is the number of time steps before sounding the alarm, if the earthquake occurs at the nth time step, the alarm has still not been sounded, and the earthquake should be considered a prediction failure. An earthquake in the box model cannot occur before the Nth time step of each cycle, so $f_e=0$ if and only if $n<N$. This error caused $f_e=0$ also for $n=N$.

We give here a revised version of Fig. 6. The correct values of f_e and $L=f_a+f_e$ are only slightly higher than those previously published. This correction changes the value of n^* (19 time steps, instead of 20). It also modifies the results in Eq. (22),

$$f_a(n^*) = 0.432, \quad f_e(n^*) = 0.084, \quad L(n^*) = 0.516, \tag{1}$$

and the value in Eq. (23) for the Parkfield sequence:

$$\tau n^* = 14.1 \text{ yr}. \tag{2}$$

We apologize for this error and hope that this note will serve to clarify the convention for calculating f_e with discrete probability distributions.

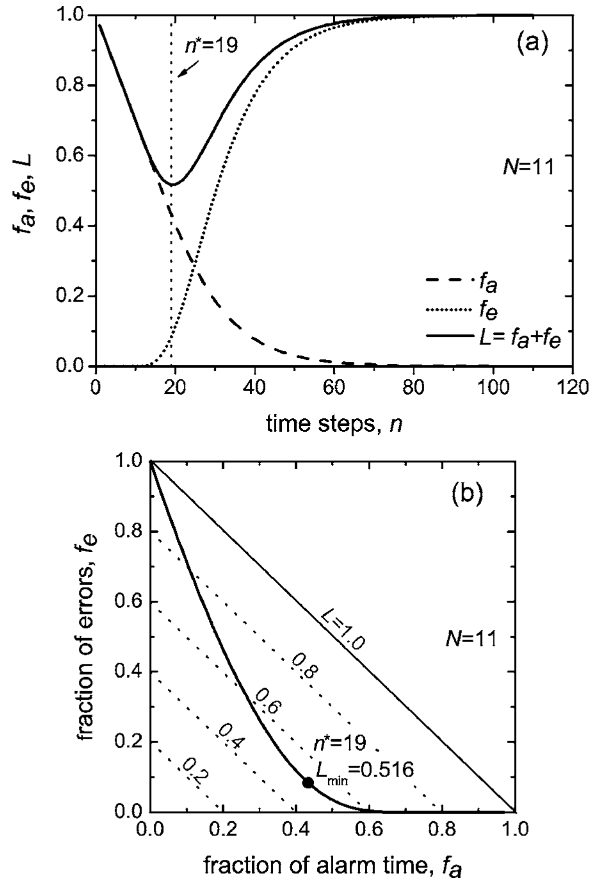

Fig. 6. Corrected version of Fig. 6 of the original article.

[a)]Electronic address: alvaro.gonzalez@unizar.es
[b)]Electronic address: jgomez@unizar.es
[c)]Electronic address: amalio@unizar.es

Appendix to "The occupation of a box as a toy model for the seismic cycle of a fault" [*Am. J. Phys.*, 73(10), 946-952]

Álvaro González* and Javier B. Gómez†
Departamento de Ciencias de la Tierra,
Universidad de Zaragoza. C. Pedro Cerbuna 12. 50009 Zaragoza, Spain

Amalio F. Pacheco‡
Departamento de Física Teórica and BIFI,
Universidad de Zaragoza. C. Pedro Cerbuna 12. 50009 Zaragoza, Spain

This is an informal appendix to the paper "The occupation of a box as a toy model for the seismic cycle of a fault" (*American Journal of Physics*, 73(10), 946-952), where we illustrated how a simple statistical model can describe the quasiperiodic occurrence of large earthquakes in a seismic fault. This appendix describes some proofs that could not be included in the original paper because of their length. Namely, we deduce here:
(1) the discrete probability distribution for the duration of the seismic cycle in the model;
(2) the asymptotic mean and standard deviation of that distribution (when the number of cells in the model tends to infinity); and
(3) the asymptotic conditional probability in this model (when the time since the last earthquake tends to infinity).

I. DISCRETE PROBABILITY DISTRIBUTION FOR THE DURATION OF THE SEISMIC CYCLE

The discrete probability distribution for the duration of the seismic cycle in the box model was named $P_N(n)$, and written in Eq. 14 of the original paper[1].

The box model is a Markov chain[2], and this enables to deduce $P_N(n)$ by using a technique[3] that we already applied to the Minimalist Model[4], which is also Markovian. A Markov chain is a stochastic process defined by a discrete random variable X that 1) can only take a finite number of values, and 2) whose value at the next time step depends only upon the value at the present time step, being independent of the way in which the present value arose from its predecessors. In other words, a Markov chain has no memory: the evolution of a Markov system at any time depends only on the state of the system at that time and not on the history of how the state was achieved.

In a box model with N cells, the state is only determined by the number of occupied cells, that here will be called ν. The succession of values of this random variable defines the stochastic process of the box model. Note that exactly which cells are occupied is not relevant, but only how many of them are. The number of stable states in the model is N; in each of them ν takes one value in the set $\{0, 1, 2, 3 \ldots (N-1)\}$. If N cells become occupied, the system instantly changes to the empty state. It does not reside any time step in the state of complete occupancy, so this is not a stable state.

The value of ν in the next time step only depends on the value of ν in the current time step, so it follows the definition of a Markov chain. For example, if the system is empty ($\nu = 0$), in the next time step, for sure (with probability equal to 1) it will move to the state of $\nu = 1$. In this second step the fraction of occupied cells is $1/N$, and the fraction of empty cells is $(N-1)/N$. So, in the third time step, with probability $(N-1)/N$ another cell will be occupied by a ball (ν becoming equal to 2), or the model will remain in $\nu = 1$ with probability $1/N$ (the probability of the incoming ball landing in the only occupied cell). In general, for $\nu < N-1$, there is a probability $(N-\nu)/N$ of moving to $\nu + 1$ in the next time step. If $\nu = N-1$, there is a probability $(N-1)/N$ of moving to $\nu = 0$ (passing through $\nu = N$, but not residing any moment there). In each time step there is a probability ν/N of residing in the same state during the next time step.

As for any other Markov chain, for the box model we can define a transition matrix $\mathbf{M}$, a table that contains all the transition probabilities of passing, in one time step, from any of the states of the system to any of the others or to itself. Each element of the matrix will be denoted in the standard way as $\mathbf{M}(i,j)$, being i the row (from top to bottom), and j the column (from left to right). Each element gives the probability of moving from the state $X = i$ in the time step n to the state $X = j$ in the step $n+1$:

$$\mathbf{M}(i,j) = P(X_{n+1} = j \mid X_n = i). \qquad (1)$$

The transition matrix of the box model is different for each N: as shown above, the transition probabilities depend on N, and because there are N stable states, the size of the matrix is $N \times N$. Thus we will denote the matrix for the box model as $\mathbf{M}_N$. Denoting the occupation state with ν as above, the element $\mathbf{M}_N(i,j)$ will be the transition probability from $\nu = i-1$ to $\nu = j-1$:

$$\mathbf{M}_N(i,j) = P(\nu_{n+1} = j-1 \mid \nu_n = i-1). \qquad (2)$$

The cause for this difference in notation is that ν ranges from 0 to $N-1$, while i and j range from 1 to N.

Let us now deduce the discrete probability distribution for the duration of the seismic cycle, using the formalism of Markov chains. The discrete distribution $P_N(n)$ defines the probability that, for a box model of N cells, the seismic cycle lasts n time steps. The seismic cycle starts when $\nu = 0$, and lasts until $\nu = 0$ again. Except for $N = 1$, which is a trivial, special case of the model, there is no possible transition in one time step from $\nu = 0$ to $\nu = 0$ (remember that this impossibility causes the stress shadow in the model). Speaking more generally, in the first $n-1$ time steps of the cycle there is no transition to the state $\nu = 0$. Because of this, to calculate $P_N(n)$ we will first deduce the probability that the system evolves from $\nu = 0$ to $\nu = N-1$ in $n-1$ time steps, without having passed through $\nu = 0$ in between. To calculate the probability that the system evolves from $\nu = N-1$ to $\nu = 0$ is simpler. At the beginning of the n-th step the system has $\nu = N-1$. Then a new particle is added to the only one empty cell, so the occupation becomes $\nu = N$, but instantly drops to $\nu = 0$ at the end of the step. The transition in the time step is thus from $\nu = N-1$ to $\nu = 0$. The probability for this to happen is $1/N$, the chance for the incoming particle to land in the only empty cell of the array when $\nu = N-1$.

Thus, the deduction of $P_N(n)$ proceeds as follows:

1. Deduce the probabilities of passing between the different states of the system in one time step. These transition probabilities will be tabulated in the transition matrix $\mathbf{M}_N$.
2. Remove from $\mathbf{M}_N$ the possibility of intermediate transitions to $\nu = 0$. The resulting matrix will be called $\mathbf{M}'_N$.
3. Calculate the transition probabilities of passing between the different states in $n-1$ time steps and neglecting the possibility of passing through the state with $\nu = 0$. The result is a new matrix, $\mathbf{T}_N = \mathbf{M}'^{n-1}_N$.
4. One of the elements of this matrix will indicate the probability of passing from $\nu = 0$ to $\nu = N-1$ in $n-1$ time steps without having passed through $\nu = 0$ in between. Multiplying this probability by $1/N$ we will obtain $P_N(n)$.

Let us proceed in this order. The transition matrices for N equal to 2, 3 and 4 are as follows:

For $N = 2$,

$$\mathbf{M}_2 = \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 0 & 2 \\ 1 & 1 \end{pmatrix}; \tag{3}$$

for $N = 3$,

$$\mathbf{M}_3 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & 0 & \frac{2}{3} \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 0 & 3 & 0 \\ 0 & 1 & 2 \\ 1 & 0 & 2 \end{pmatrix}; \tag{4}$$

and for $N = 4$,

$$\mathbf{M}_4 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{3}{4} & 0 \\ 0 & 0 & \frac{2}{4} & \frac{2}{4} \\ \frac{1}{4} & 0 & 0 & \frac{3}{4} \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 0 & 4 & 0 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 2 & 2 \\ 1 & 0 & 0 & 3 \end{pmatrix}. \tag{5}$$

All the elements of these matrices are nonnegative (they are probabilities) and the sum of all the elements of any row is always 1. These two are necessary and sufficient properties of transition matrices of Markov chains. These matrices show evident regularities, which enable to deduce by inspection that the matrix for any N is

$$\mathbf{M}_N = \frac{1}{N}\begin{pmatrix} 0 & N & 0 & 0 & \dots & 0 \\ 0 & 1 & N-1 & 0 & \dots & 0 \\ 0 & 0 & 2 & N-2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & N-2 & 2 \\ 1 & 0 & 0 & 0 & 0 & N-1 \end{pmatrix}. \tag{6}$$

Note that the matrix multiplied by $1/N$ has only three non-null diagonals, all of them trivial. The first one is the sequence $N, N-1, N-2 \dots 2$, the second one is the sequence $0, 1, 2 \dots N-1$, and the third one is only the bottom left element, which is always 1.

To calculate $P_N(n)$ the next step (the second one in the list above) is to prune from this matrix the transitions to $\nu = 0$. The only possible transition to $\nu = 0$ is from $\nu = N-1$, and the probability for this transition is given by the bottom left element $\mathbf{M}_N(N, 1)$. Nullifying this element, the resulting matrix, $\mathbf{M}'_N$, is particularly simple, because it has only two trivial, non-null diagonals:

$$\mathbf{M}'_N = \frac{1}{N}\begin{pmatrix} 0 & N & 0 & 0 & \dots & 0 \\ 0 & 1 & N-1 & 0 & \dots & 0 \\ 0 & 0 & 2 & N-2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & N-2 & 2 \\ 0 & 0 & 0 & 0 & 0 & N-1 \end{pmatrix}. \tag{7}$$

Now (third step of the list) it is necessary to compute a new matrix, $\mathbf{T}_N$, which indicates all the transition probabilities, in $n-1$ time steps, between all the states, with the restriction that passing from $\nu = N-1$ to $\nu = 0$ is forbidden. In Markov chains, the m-step transition probability matrix is the m-th power of the transition matrix[3]. So the new matrix is

$$\mathbf{T}_N = \mathbf{M}'^{n-1}_N. \tag{8}$$

This operation is done through the Jordan decomposition of $\mathbf{M}'_N$. The element $\mathbf{T}_N(1, N)$ of this matrix is the transition probability from $\nu = 0$ to $\nu = N-1$ in $n-1$ time steps and with the transition $\nu = N-1 \to \nu = 0$ forbidden. As the probability of passing, in the next time step, from $\nu = N-1$ to $\nu = 0$ is $1/N$, $P_N(n)$ is obtained by multiplying that element by $1/N$. The results, for $N = 2$ and $N = 3$ are as follows.

For $N = 2$,

$$\frac{1}{2^{n-1}} = \frac{2}{2^n} = \frac{1}{2^n}\sum_{j=0}^{2-1}\left[\binom{2}{j} j^{n-1}(-1)^{1-j}(2-j)\right] = \\ = \frac{1}{2^n}[0+2]; \qquad (9)$$

and for $N = 3$,

$$\frac{2}{3^{n-1}}(2^{n-2}-1) = \frac{1}{3^n}\sum_{j=0}^{3-1}\left[\binom{3}{j} j^{n-1}(-1)^{1-j}(3-j)\right]. \qquad (10)$$

By inspection, the result for a generic N is

$$P_N(n) = \frac{1}{N^n}\sum_{j=0}^{N-1}\left[\binom{N}{j} j^{n-1}(-1)^{1-j}(N-j)\right] = \\ = \sum_{j=1}^{N-1}(-1)^{j+1}\binom{N-1}{j-1}\left(1-\frac{j}{N}\right)^{n-1} \qquad (11)$$

(Eq. 14 of the original paper).

II. ASYMPTOTIC MEAN OF $P_N(n)$

The mean duration of the cycle in the box model was indicated in Eq. 8 of the original paper:

$$\langle n\rangle_N = 1 + \sum_{i=2}^{N}\frac{N}{N+1-i} = \\ = 1 + \frac{N}{N-1} + \frac{N}{N-2} + \cdots + \frac{N}{2} + \frac{N}{1} = \\ = N\left[\frac{1}{N} + \frac{1}{N-1} + \frac{1}{N-2} + \cdots + \frac{1}{2} + \frac{1}{1}\right] \qquad (12)$$

The asymptotic value of this expression can be obtained considering that[5]

$$\sum_{i=1}^{N}\frac{1}{i} \xrightarrow[N\to\infty]{} C + \ln N + \frac{1}{2N} - 0\left(\frac{1}{N}\right)^2, \qquad (13)$$

where $C \simeq 0.5772157$ is Euler's constant. Multiplying this equation by N we obtain the asymptotic mean of $P_N(n)$,

$$\langle n\rangle_N \xrightarrow[N\to\infty]{} N(C + \ln N) + \frac{1}{2} \qquad (14)$$

(Eq. 12 of the original paper). The absolute error of this approximation is < 0.01 for $N \geq 9$.

III. ASYMPTOTIC STANDARD DEVIATION OF $P_N(n)$

The variance of $P_N(n)$ was indicated in Eq. 9 of the original paper:

$$\sigma_N^2 = \sum_{i=1}^{N}\frac{1-\dfrac{N+1-i}{N}}{\left(\dfrac{N+1-i}{N}\right)^2} = \\ = \sum_{i=1}^{N}\frac{1}{\left(\dfrac{N+1-i}{N}\right)^2} - \sum_{i=1}^{N}\frac{1}{\dfrac{N+1-i}{N}}. \qquad (15)$$

To simplify the sums, we can change the variable to $k \equiv N + 1 - i$. Because i ranges from 1 to N, k will range from N to 1. Then the above equation can be rewritten as

$$\sigma_N^2 = \sum_{k=1}^{N}\frac{1}{\left(\dfrac{k}{N}\right)^2} - \sum_{k=1}^{N}\frac{1}{\dfrac{k}{N}} = \sum_{k=1}^{N}\frac{N^2}{k^2} - \sum_{k=1}^{N}\frac{N}{k} = \\ = N^2\sum_{k=1}^{N}\frac{1}{k^2} - N\sum_{k=1}^{N}\frac{1}{k}. \qquad (16)$$

The first sum in the right-hand side of this expression can be simplified to

$$\sum_{k=1}^{N}\frac{1}{k^2} = \sum_{k=1}^{\infty}\frac{1}{k^2} - \int_N^{\infty}\frac{1}{k^2}\mathrm{d}k = \\ = \frac{\pi^2}{6} - \left[-\frac{1}{k}\right]_N^{\infty} = \frac{\pi^2}{6} - \frac{1}{N}. \qquad (17)$$

Inserting Eq. 13 and this result, Eq. 16 in the limit of $N \to \infty$ can be written as

$$\sigma_N^2 \xrightarrow[N\to\infty]{} N^2\left(\frac{\pi^2}{6} - \frac{1}{N}\right) - \\ -N\left[C + \ln N + \frac{1}{2N} - 0\left(\frac{1}{N}\right)^2\right] = \\ = N^2\frac{\pi^2}{6} - N - CN - N\ln N - \frac{1}{2} = \\ = N^2\frac{\pi^2}{6} - N(1 + C + \ln N) - \frac{1}{2} = \\ = N^2\left[\frac{\pi^2}{6} - \frac{1 + C + \ln N}{N} - \frac{1}{2N^2}\right]. \qquad (18)$$

The asymptotic standard deviation is the square root of the above equation,

$$\sigma_N \xrightarrow[N\to\infty]{} N\left[\frac{\pi^2}{6} - \frac{1 + C + \ln N}{N} - \frac{1}{2N^2}\right]^{1/2}. \qquad (19)$$

Because $N \to \infty$, the term $-1/2N^2$ can be dropped, so the equation can be further simplified to

$$\sigma_N \xrightarrow[N\to\infty]{} N \left[\frac{\pi^2}{6} - \frac{1 + C + \ln N}{N}\right]^{1/2} \qquad (20)$$

(Eq. 13 of the original paper). This approximation has an absolute error < 0.01 for $N \geq 3$.

The asymptotic aperiodicity, obtained by dividing Eq. 20 by Eq. 14, has an absolute error < 0.0001 for $N \geq 10$.

IV. ASYMPTOTIC CONDITIONAL PROBABILITY

To deduce the asymptotic conditional probability in the box model we will first consider the asymptotic value of $P_N(n)$ when $n \to \infty$. This value is the first, largest term in the sum (when $j = 1$ in Eq. 11), namely

$$P_N(n) \xrightarrow[n\to\infty]{} \left(1 - \frac{1}{N}\right)^{n-1} = a^{n-1}, \qquad (21)$$

where we have denoted $a \equiv 1 - 1/N$.

For calculating the asymptotic conditional probability we need to deduce the value of the cumulative probability distribution, $A_N(n)$, for that large n. This is easier to do by defining the sum

$$A'_N(n) = \sum_{i=n}^{\infty} a^{i-1} = \frac{a^{n-1}}{1-a}. \qquad (22)$$

Considering that n is large enough, $P_N(n)$ can be replaced by its asymptotic value (Eq. 21), which is the term summed in $A'_N(n)$. Then it holds that

$$A_N(n-1) = \sum_{i=1}^{n-1} P_N(n) \xrightarrow[n\to\infty]{} 1 - A'_N(n), \qquad (23)$$

so

$$A_N(n) \xrightarrow[n\to\infty]{} 1 - A'_N(n+1). \qquad (24)$$

The conditional probability (Eq. 18 of the original paper) is:

$$P_\tau(N,n) = \frac{A_N(n+1/\tau) - A_N(n)}{1 - A_N(n-1)}. \qquad (25)$$

Inserting Eqs. 22 to 24, it results that

$$\begin{aligned} P_\tau(N,n) \xrightarrow[n\to\infty]{} & \frac{1 - A'_N(n+1/\tau+1) - [1 - A'_N(n+1)]}{A'_N(n)} = \\ &= \frac{A'_N(n+1) - A'_N(n+1/\tau+1)}{A'_N(n)} = \\ &= \frac{\dfrac{a^{n+1-1} - a^{n+1/\tau+1-1}}{1-a}}{\dfrac{a^{n-1}}{1-a}} = \\ &= a - a^{1+1/\tau} = a\left(1 - a^{1/\tau}\right) = \\ &= \left(1 - \frac{1}{N}\right)\left[1 - \left(1 - \frac{1}{N}\right)^{1/\tau}\right]. \end{aligned} \qquad (26)$$

In the original paper we were interested in the yearly conditional probability for the next large earthquake at Parkfield. In order to fit the series of previous earthquakes, N was chosen as 11 cells, and one time step corresponded to $\tau = 0.74$ years. Substituting these values in the above equation, the asymptotic yearly probability when a long time has passed since the last large earthquake is 0.11 (the value of 11% cited in the original paper).

Acknowledgments

The Spanish Ministry of Education and Science funded this research by means of the project FIS2005–06237, and the grant AP2002–1347 held by Á.G.

* Electronic address: alvaro.gonzalez@unizar.es
† Electronic address: jgomez@unizar.es
‡ Electronic address: amalio@unizar.es

[1] Álvaro González, Javier B. Gómez, and Amalio F. Pacheco, "The occupation of a box as a toy model for the seismic cycle of a fault," *American Journal of Physics* **73**(10), 946–952 (2005). http://arxiv.org/abs/physics/0502048

[2] This kind of stochastic processes is named after the Russian mathematician Andrey Andreyevich Markov (1856-1922), who was the first in describing it (in 1906). The foundation of a general theory was provided during the 1930s by Andrey Nikolaevich Kolmogorov.

[3] See for example R. Durrett, *Essentials of Stochastic Processes* (Springer Verlag, Berlin, Germany, 1999).

[4] Miguel Vázquez-Prada, Álvaro González, Javier B. Gómez, and Amalio F. Pacheco, "A minimalist model of characteristic earthquakes," *Nonlinear Processes in Geophysics* **9**(5-6), 513–519 (2002). http://www.copernicus.org/EGU/npg/9/513.htm

[5] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products* (Academic Press, New York, 1965).