Pure Appl. Geophys. 173 (2016), 197–203

DOI 10.1007/s00024-015-1056-3

Pure and Applied Geophysics

A Bayesian Assessment of Seismic Semi-Periodicity Forecasts

F. Nava,[1] C. Quinteros,[1] E. Glowacka,[1] and J. Frez[1]

Abstract—Among the schemes for earthquake forecasting, the search for semi-periodicity during large earthquakes in a given seismogenic region plays an important role. When considering earthquake forecasts based on semi-periodic sequence identification, the Bayesian formalism is a useful tool for: (1) assessing how well a given earthquake satisfies a previously made forecast; (2) re-evaluating the semi-periodic sequence probability; and (3) testing other prior estimations of the sequence probability. A comparison of Bayesian estimates with updated estimates of semi-periodic sequences that incorporate new data not used in the original estimates shows extremely good agreement, indicating that: (1) the probability that a semi-periodic sequence is not due to chance is an appropriate estimate for the prior sequence probability estimate; and (2) the Bayesian formalism does a very good job of estimating corrected semi-periodicity probabilities, using slightly less data than that used for updated estimates. The Bayesian approach is exemplified explicitly by its application to the Parkfield semi-periodic forecast, and results are given for its application to other forecasts in Japan and Venezuela.

Key words: Earthquake forecasting, Bayesian probability, semi-periodicity.

1. Introduction

An important approach to earthquake prediction is the search for statistical regularities in the time occurrence of large earthquakes. Indeed, a simplified application of the elastic rebound model (Reid 1910, as referenced in Richter 1958) with plate tectonics (e.g., Morgan 1968; Cox 1973) as the (constant rate) strain source, would lead one to expect periodic behavior (e.g., Lomnitz 1966; Rikitake 1976; and references therein).

However, seismic processes involve complex and highly non-linear systems featuring feedback, thusly depending heavily on its history. As such, this process involves self-organized criticality (SOC) (e.g., Bak *et al.* 1988; Bak and Tang 1989; Bak and Chen 1991; Turcotte 1992; Márquez 2012) with essentially random occurrences of small events and semi-periodic occurrences of large events.

The occurrence times of a semi-periodic sequence of K earthquakes are of the form:

$$t_k = t_0 + k\,\tau + \eta_k; \quad k = 1, \ldots, K, \tag{1}$$

where τ is the period and η_k is a realization of a random variable such that $\eta \ll \tau$.

Many studies have searched for semi-periodicity, with mixed results. Of these, we will use as an example the one by Bakun and Lindh (1985) which predicted an earthquake in the region of Parkfield, California, USA, on the basis of recurrence times from a series of six earthquakes, and missed the occurrence time of the next earthquake by some 17 years.

Nava *et al.* (2014) and Quinteros *et al.* (2014), hereafter referred to as Paper I and Paper II, respectively, realized that in a given seismogenic region there may be more than one semi-periodic process, so that the observed seismicity may contain more than one semi-periodic sequence, and may also include events from long-period sequences that cannot be identified because of the limited observation time. As a result, they can be considered to occur randomly. Not all earthquakes that occur in a given seismogenic region are necessarily part of a seismic sequence. Paper I proposes a method to identify semi-periodic sequences in earthquake occurrence time series, assess their significance, and use the results for earthquake forecasting; it also illustrates the method by applying it to the Parkfield sequence. Paper II addresses some aspects of catalogue processing and presents applications to Japan and Venezuela; Quinteros and Nava (2013), hereafter referred to as Paper III, presents the application to a recent event in

[1] CICESE, Ensenada, Baja California, Mexico. E-mail: fnava@cicese.mx

Birkhäuser

Venezuela. We will use results from the abovementioned papers, where all details of the pertinent data are given, to apply and illustrate the proposed Bayesian estimation.

We will now briefly review the basic characteristics of the semi-periodicity forecasting method.

The time series of large earthquake occurrence times, considered as a point process in time, is

$$\tilde{t} = \{t_j;\ j = 1, \ldots, N\},\quad N \geq K;$$

from this series, a function is built as

$$f(t) = \sum_{j=j_1}^{j_2} \delta(t - t_j);$$

recognizing this function as a segment of an infinite series, the analytic Fourier transform (e.g., BRACEWELL 1965)

$$F(s) = \sum_{j=j_1}^{j_2} \mathrm{e}^{-i\,2\pi\, t_j\, s}$$

allows identification of dominant frequencies corresponding to semi-periodic sequences within the point process of earthquake occurrences in time. Figure 1 illustrates the Parkfield time series and the process of sequence identification.

Once a dominant frequency is identified in the spectrum (Fig. 1), a periodic sequence in time (referred to as a "comb") is built based on the identified spectral period τ and phase. Events of the time series possibly corresponding to comb "teeth" are identified and the rest are eliminated. The process is repeated three times using a stricter acceptance criterion during each pass (in the last pass, acceptable occurrence times t_k have to differ from that of the corresponding comb tooth t_k^c by less than $\tau/6$). Goodness of fit is measured as the root-mean-square (RMS) error of fit between sequence and comb

$$\sigma = \sqrt{\frac{\sum_{k=1}^{K} (t_k^c - t_k)}{K - 2}}. \quad (1)$$

For the Parkfield example, $\tau_p = 36.36$ year and $\sigma = 4.55$ year.

Based on the estimated comb and the error, a forecast can be made as

$$t_f = t_{0p} + K\tau \pm q\sigma, \quad (2)$$

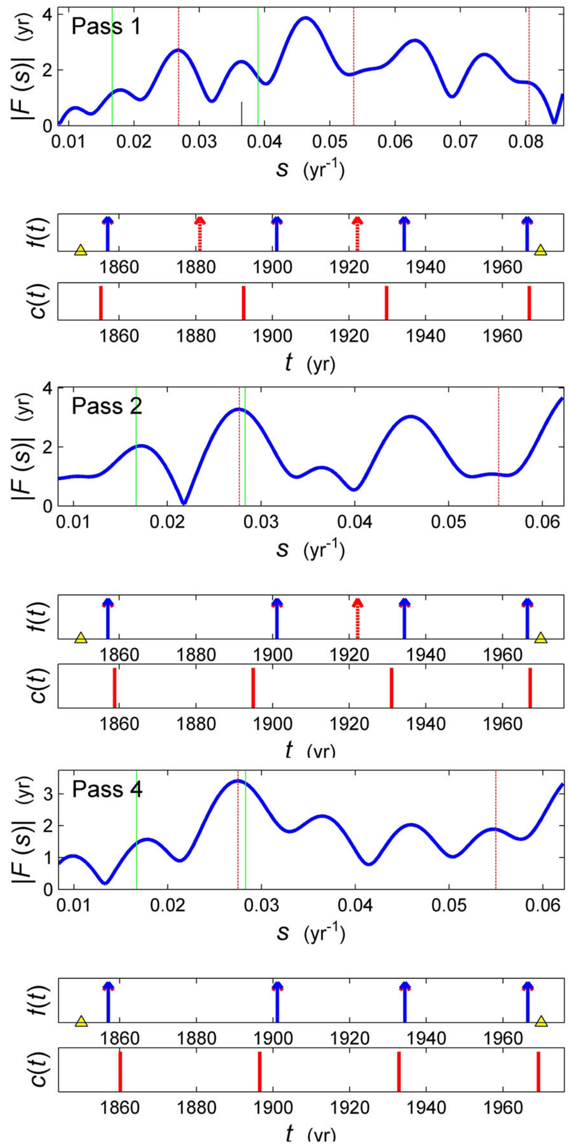

Figure 1
Example of sequence and comb determination for Parkfield (Paper I). For each pass, a section of the (absolute) spectrum is shown on top; the *vertical continuous lines* indicate the range of acceptable frequencies, and the *dotted vertical lines* indicate the chosen frequency (*left*) and some of its multiples. Below the spectra are the occurrence time series (*arrows*); *dotted arrows* show ineligible events. Below the time series the estimated comb (*vertical lines*). Pass 3 is not shown because it is exactly like pass 2

where t_f is the forecast time, t_{0p} is the time of the earliest comb tooth, determined from the phase and the period, and q is a factor that can be set to give a desired confidence interval to the forecast.

Considering $q = 2$, the forecast time of the Parkfield example is $t_f = 2005.63 \pm 9.10$ year (Paper I).

The sequence probability of non-randomness.

How significant an identified semi-periodic sequence is can be measured against the probability, P_0, of the null hypothesis that the observed sequence is a random occurrence, i.e., that earthquakes occurring with uniform probability over the observed interval $[0, T]$ could result in a sequence having K elements with RMS error $\pm q\,\sigma$, where a factor $q \geq 1$ is introduced to ensure that P_0 is not underestimated. We will now describe how this probability is estimated.

For a random occurrence with uniform probability, the distribution with the largest entropy, the probability of occurrence of n events within an interval θ is Poissonian (e.g., Lomnitz 1994; Daley and Vere-Jones 2002), given by

$$\Pr\,(n) = \frac{(\lambda\theta)^n\,\mathrm{e}^{-\lambda\theta}}{n!},$$

where $\lambda = N/T$ is the occurrence ratio of earthquakes in the region.

In order to have exactly K elements over interval $[0, T]$, a sequence may have a period between $T/(K-1)$ and $T/K + \varepsilon$, where ε is a very small quantity introduced to ensure that no more than K elements fit within time T. We use the worst case that results in the largest random probability by considering the shortest period for which at least one event should occur, taking into account the uncertainty, within an interval of length $\Theta = T/K - \varepsilon + q\sigma$; at least one event should occur in each of $K-1$ intervals of length $\theta = 2q\sigma$. Thus, the worst-case random occurrence probability is

$$P_0 = \left(1 - \mathrm{e}^{-\lambda\Theta}\right)\left(1 - \mathrm{e}^{-\lambda\theta}\right)^{K-1}. \quad (3)$$

Hence, the probability that the sequence did not occur randomly is

$$P_c = 1 - P_0. \quad (4)$$

It should be mentioned that the probabilities presented here, calculated according to (3) and (4), differ slightly from those in our previous papers because, in them, P_0 was estimated approximately by a Monte-Carlo scheme.

Considering that the actual occurrence times should be distributed about the forecast time as some pdf $p(t - t_f)$ with unit area, the probability of occurrence is given by $P_c\,p(t - t_f)$, and the forecast can be represented as in Fig. 2. In this forecast, we have assumed a normal distribution $p(t) = N(t_f, \sigma)$ (see Paper I for discussion).

Actually, Fig. 2 shows an aftcast, i.e., a forecast for an event that has already occurred, based on information previous to it. The first four arrows show the identified sequence, and the curve is the forecast probability density function, while the fifth arrow indicates the occurrence of the forecast earthquake. It is evident that the actual occurrence agrees very well with the forecast, which had $P_c = 0.858$, but how exactly does the occurrence (or non-occurrence) of earthquakes after the forecast was made support or contradict the forecast and, hence, the hypothesis of semi-periodicity?

2. Bayesian Probability of Semi-Periodicity

We will now apply the Bayesian formalism to derive quantitative information from the occurrence or non-occurrence of the forecast earthquake, as well as from the accuracy of the forecast, i.e., the difference between forecast and occurrence time Δt.

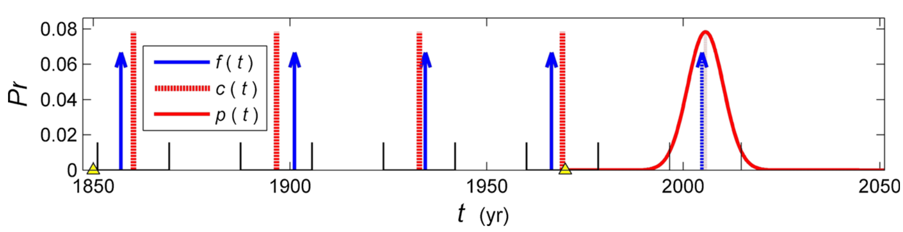

Figure 2
Forecast for the Parkfield data (Paper I). *Arrows* indicate best fitting earthquake occurrences, and *dashed vertical lines* are the comb teeth series; the curve is the forecast pdf $p(t)$. The *latest arrow* (*dotted*) marks the forecast event occurrence on $t_o = 2004.742$

2.1. Estimation of the Bayesian Probability

Let A be a semi-periodic earthquake sequence in the study region (evidence of a semi-periodic process), and let Pr(A) be the prior estimation of the probability of A.

Let B be an earthquake that occurs after the forecast has been made, at time t_0. Event B can be used to revise the probability of A by applying Bayes' formula:

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B|A)\Pr(A) + \Pr(B|\bar{A})\Pr(\bar{A})}, \quad (5)$$

(e.g., Parzen 1960; Winkler 2003). We will now calculate the various probabilities needed to apply (5).

The total forecast probability, P_c, is distributed in time according to some pdf $p(t)$; thus, in order to have a finite probability for the occurrence of B at a given time, P_w, it is necessary to consider the probability of occurrence within a window of finite length w centered on the occurrence time t_0:

$$P_w = P_c \int_{t_o - w/2}^{t_o + w/2} p(t)\, \mathrm{d}t. \quad (6)$$

The effect of the length of the time window will be discussed below. For the case $p(t) = N(t_f, \sigma)$, illustrated in Fig. 3,

$$\begin{aligned} P_w &= \frac{P_c}{\sqrt{2\pi}} \int_{t_1}^{t_2} \mathrm{e}^{-x^2/2} \mathrm{d}x \\ &= P_c \left\{ \operatorname{sgn}(t_2) \operatorname{erf}\left(|t_2|/\sqrt{2} \right) - \operatorname{sgn}(t_1) \operatorname{erf}\left(|t_1|/\sqrt{2} \right) \right\} \end{aligned} \quad (7)$$

where $t_1 = (t_f - t_o - w/2)/\sigma$ and $t_2 = (t_f - t_o + w/2)/\sigma$.

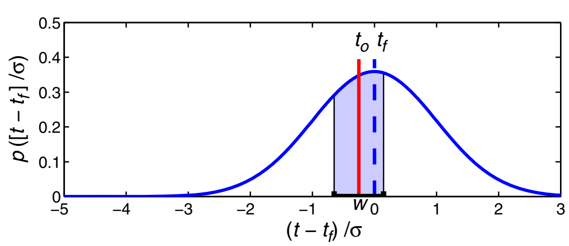

Figure 3
Forecast probability pdf $p(t)$ for normalized time

Hence, the probability of B, given A, the occurrence of at least one earthquake within window w, is the abovementioned probability of having an earthquake from the sequence, plus the Poissonian probability of having at least one earthquake that does not belong to it $\pi^*_{1+} = 1 - \mathrm{e}^{-w\,(N-K)\,/T}$:

$$\Pr(B|A) = P_w + \pi^*_{1+} - P_w \pi^*_{1+}, \quad (8)$$

since $(N - K)/T$ is the occurrence ratio of earthquakes not belonging to the given sequence.

If $\bar{A}$, then the probability of B is strictly Poissonian:

$$\Pr(B|\bar{A}) = 1 - \mathrm{e}^{-wN/T} \quad (9)$$

and, obviously,

$$\Pr(\bar{A}) = 1 - \Pr(A). \quad (10)$$

It only remains to assign a value to Pr(A), the prior probability of there being a semi-periodic sequence. Since P_0 is the (Poissonian) probability of the observed semi-periodic sequence being due to chance, and the only alternative to this is that the semi-periodic sequence is not due to chance, it follows that

$$\Pr(A) = 1 - P_0 = P_c. \quad (11)$$

Assuming $w = \sigma/40$ (see discussion about w below), the quantities estimated from (6) to (11) can be applied in (5).

The Parkfield earthquake series used for the semi-periodicity analysis in Paper I consists of six earthquakes with magnitudes between 6.0 and 7.9, among which a sequence of four events is identified (Fig. 2). The parameter values for this example and the results of the Bayesian appraisal are shown in the first row of Table 1; for $\Pr(A) = P_c = 0.858$ and $t_f = 2005.63$, the occurrence of the forecast event on $t_o = 2004.742$, so that $\Delta t = |t_f - t_o| = 0.888$ year yields $\Pr(A|B) = 0.952$, i.e., a probability gain (Daley and Vere-Jones 2003) $\Pr(A|B) = 0.952$.

Table 1 also shows values and results for aftcasts from Papers II and III. The aftcast of the last event in the J1 sequence in Japan is based on a sequence of $K = 3$ events among a series of $N = 8$ events with magnitudes ranging from 8.0 to 8.7; that of the last event in the R2 $K = 4$ sequence in Venezuela involves a series of $N = 13$ with magnitudes in the

Table 1

Quantities used in the Bayesian appraisal and results of the same for all examples mentioned in the text

Example	T (year)	N	K	σ (year)	t_f	P_c	t_o	Pr $(A\|B)$ P_G			P_c^U
								Pr $(A) = P_c$	Pr $(A) = 0.5$	Pr $(A) = 0.1$	
Parkfield	120	6	4	4.55	2005.630	0.858	2004.742	0.952 1.110	0.767 1.534	0.268 2.676	0.952
Japan J1	90	8	3	0.228	2003.090	0.994	2003.734	0.996 1.002	0.575 1.150	0.131 1.306	0.997
Venezuela R2	196	13	4	1.41	2010.230	0.967	2009.951	0.996 1.031	0.893 1.785	0.480 4.800	0.994
Venezuela 2013	91	20	6	1.525	2014.670	0.785	2013.778	0.892 1.137	0.694 1.389	0.202 2.015	0.861

The updated comb probability P_c^U is also shown for comparison.

5.6–7.4 range. Finally, the aftcast of the recent 2013 Venezuela $M = 6.5$ earthquake is based on a $K = 6$ event sequence among an $N = 20$ event series with magnitudes ranging from 5.9 to 6.9.

Bayes formalism allows testing other prior estimates or suppositions about possible semi-periodicity. Someone who does not like condition (11) and says that semi-periodicity may or may not occur would use Pr $(A) = 0.5$, and, for our Parkfield example, would obtain Pr $(A|B) = 0.767$ ($P_G = 1.534$). Someone who is skeptical and believes there is only a small probability of there being semi-periodicity could use, say, Pr $(A) = 0.1$, and would obtain Pr $(A|B) = 0.268$ ($P_G = 2.676$), a probability not large enough to be conclusive in favor of semi-periodicity, but certainly suggestive of it. Results for these hypothetical choices for the other aftcasts are shown in Table 1; note that, for our examples, the smaller prior probabilities result in larger probability gains. Of course Bayesian reasoning is useless for someone who firmly believes that semi-periodicity cannot exist, so that Pr $(A) = 0$, because the question becomes a matter of faith.

2.2. *Bayesian Probability and the Length of w*

Since forecasts are given as probability distribution functions in time, in order to work with finite probabilities, it is necessary to consider probabilities over some finite time interval, which we have called w. Probabilities (6–11) all increase with w, but since both the numerator and the denominator increase in (5), for $w \leq \sigma/4$, the results change only in the fourth decimal place, and tend to a limit for small w. Thus, for practical purposes, results for $w \leq \sigma/10$ can be considered independent of w. We used $w = \sigma/40$ in the results presented here.

2.3. *Bayesian Probability and Forecast Accuracy*

Among the probabilities used to calculate Pr $(A|B)$, only P_w depends on the time difference between the forecast time t_f and the actual occurrence time t_0, and attains its maximum when both times coincide, i.e., when the forecast is exact.

All aftcasts we have made so far are good enough such that the Bayesian estimates give enhanced semi-periodicity probabilities. However, this might not have been the case had the forecast earthquake occurred at a time very different from that of the forecast.

Figure 4 illustrates the behavior of Pr $(A|B)$ as a function of $\Delta t = |t_f - t_o|$ for the Parkfield example with Pr $(A) = P_c$ (top) and Pr $(A) = 0.5$ (bottom). In both cases, the Bayesian probability is maximal for $\Delta t = 0$, but there also is a Δt above which the Bayesian probability is smaller than the prior one, and clearly tells that, above this value, event B is evidence against semi-periodicity. Note that the smaller the prior estimate, the better the coincidence between actual and forecast times has to be in order to be convincing.

A common problem in evaluating forecasting performance is that it is sometimes difficult to decide whether a given event fulfills a forecast or not, particularly if it occurs with relatively low

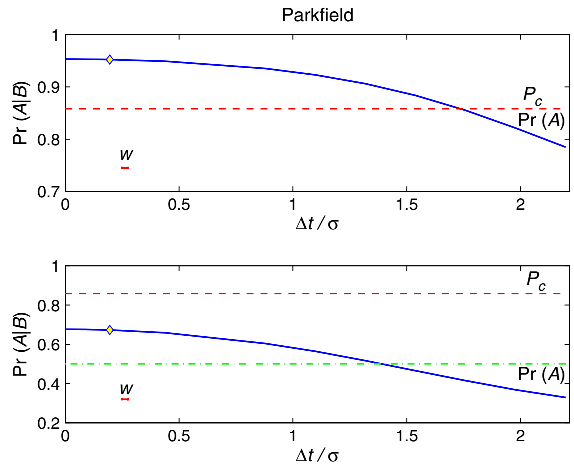

Figure 4
Dependence of the Bayesian probability on the difference between forecast and occurrence times Δt, for $\Pr(A) = P_c$ (*top*) and $\Pr(A) = 0.5$ (*bottom*). The *diamond indicates* the (normalized) Δt for the Parkfield example. The *bar* labeled w shows the window length used to compute probabilities

probability. Everyone knows of cases where forecasters claim that some event remotely resembling their forecast is an actual fulfillment. This problem need not arise for semi-periodicity forecasts when viewing each new occurrence in the light of Bayesian estimates, because any event resulting in $P_G < 1$ is clearly not fulfilling the forecast, so that P_G is a good estimator of how well an event fulfills a forecast.

2.4. *Bayesian vs. Updated Probabilities*

After a forecast (or aftcast) has been made, the occurrence of an event allows reevaluation of the prior probability, but it also results in a new (or updated) sequence of $K + 1$ earthquakes. Conditions are not the same for the new sequence: the number of earthquakes in the sequence has increased by one, but other "unrelated" events may have also occurred, i.e., N may be larger by more than one, and the total time T has increased by about one sequence period. Based on the new conditions, a new, updated, sequence probability P_c^U can be evaluated (the probability that would be used for a new forecast, and it is interesting to compare the updated probabilities for our examples with the corresponding Bayesian probabilities.

Table 1 shows that, for our aftcast examples, the Bayesian $\Pr(A|B)$ estimates for $\Pr(A) = P_c$ agree extremely well with the corresponding updated P_c^U values. We believe that two conclusions are derivable from this agreement: first, that the choice of P_c as the prior probability is correct; and second, since longer sequences involving larger numbers of earthquakes that occurred semi-periodically are, naturally, better and more convincing evidence of semi-periodic behavior; the Bayesian formalism does a very good job of estimating corrected semi-periodicity probabilities, using slightly less data than that used for updating.

3. *Discussion and Conclusions*

The real measure of the goodness of a forecast is whether an earthquake occurs around the forecast time and, given the occurrence, how small the absolute difference is between the forecast and the actual occurrence times (judged in terms of the period and the standard deviation of the semi-periodic sequence). Hence, the Bayesian estimation of the probability of a sequence being semi-periodic, given the occurrence of a given earthquake, is a measure of both whether the earthquake may be considered to fulfill the forecast and, if so, of the forecast goodness.

For the examples shown here, the agreement of the Bayesian estimates based on the non-randomness probability of the sequence used for the forecast, P_c, with the updated non-randomness probability of the new, longer sequence suggests that P_c is a good estimator of the probability of existence of a semi-periodic sequence.

The Bayesian estimates, together with the updated non-randomness probabilities, are a good basis upon which to support or reject the existence of semi-periodic sequences in earthquake occurrence time series.

Acknowledgments

Our sincere thanks to two anonymous reviewers and to Pageoph guest editor Matthew Gerstenberger. This project was partially funded by CONACYT scholarship 242919 (C. Quinteros) and CONACYT grant 222795, and was partially carried out within project CGL2011-29474-C01-01.

REFERENCES

BAK, P., TANG, C., and WIESENFELD, K. (1988) *Self-organized criticality*. Phys. Rev. A *38*(1), 364–374.

BAK, P., and TANG, C. (1989) *Earthquakes as a self-organized critical phenomenon*. J. Geophys. Res. *94*(B1), 15635–15637.

BAK, P., and CHEN, K. (1991) *Self-organized criticality*. Sci. Am. 26–33.

BAKUN, W., and LINDH A. (1985) *The Parkfield, California earthquake prediction experiment*. Science *229*, 4714, 619–624.

BRACEWELL, R., The Fourier transform and its applications. (McGraw-Hill Book Co., USA, 381 pp., 1965).

COX, A. (1973) Plate tectonics and geomagnetic reversals. (W. H. Freeman and Co.).

DALEY, D., VERE-JONES, D. (2002) An introduction to the theory of point processes. Vol. I. Springer, USA, 469 pp.

LOMNITZ, C. (1966) *Statistical prediction of earthquakes*. Reviews of Geophysics *4*, 377–393.

LOMNITZ, C., Fundamentals of earthquake prediction. (JohnWiley & 673 Sons Inc., USA, 326 pp. 1994).

MÁRQUEZ, V. (2012) Multifractal analysis of the spatial distribution of seismicity and its possible prmonitory application. Exploration of a possible mechanism for fractality through semi-stochastic modeling. D.Sc. Thesis, CICESE, 31/08/2012 (in Spanish).

MORGAN, W. (1968). *Rises, trenches, great faults, and crustal blocks*. J. Geophys. Res. *73*(6), 1959–1982.

NAVA, F., QUINTEROS, C., GLOWACKA, E., FREZ, J. (2014) *Semi-periodic sequences and extraneous events in earthquake forecasting: I. Theory and method, Parkfield*. Pageoph *171*(7), 1355–1366, doi:10.1007/s00024-013-0679-5.

PARZEN, E. (1960) *Modern probability theory and its applications.* John Wiley & Sons, Inc., Japan, 464 pp.

QUINTEROS, C., NAVA, F., GLOWACKA, E., FREZ, J. (2014) *Semi-periodic sequences and extraneous events in earthquake forecasting: II Application, forecasts for Japan and Venezuela*. Pageoph *171*(7), 1367–1383, doi:10.1007/s00024-013-0678-6.

QUINTEROS, C., NAVA, F. (2013) *Postnóstico (pronóstico hecho a posteriori) del sismo del 11 de octubre de 2013 en Venezuela, mediante análisis de semiperiodicidad*. GEOS ***33*** (2), 350-355..

REID, H. (1910) The mechanics of the earthquake, the California earthquake of April 18, 1906. Report of the State Investigation Commission, Vol. 2, Carnegie Institution of Washington, Washington, D.C.

RICHTER, C., Elementary Seismology, (W. H. Freeman and Co., USA, 768 pp. 1958).

TURCOTTE, D., Fractals and Chaos in Geology and Geophysics. (Cambridge University Press, Second edition, New York. 221 pp, 1992).

WINKLER, R. An introduction to Bayesian inference and decision. 2nd Edition, (Probabilistic Publishing, USA, 452 pp. 2003).

(Received July 29, 2014, revised January 15, 2015, accepted February 11, 2015, Published online February 25, 2015)