

## AN ALTERNATIVE BAYESIAN STATISTICS FOR PROBABILISTIC EARTHQUAKE PREDICTION IN MEXICO, CENTRAL AND SOUTH AMERICA

O.CH. GALANIS<sup>1</sup>, T.M. TSAPANOS<sup>1</sup>, G.A. PAPAPOULOS<sup>2</sup> AND A.A. KIRATZI<sup>1</sup>

### ABSTRACT

The probabilities of occurrence of strong ( $M \geq 6.5$ ) earthquakes, in the seismically active regions of Mexico, central and south America, are estimated. The straightforward approach of Bayes statistics is applied in order to search for the inter-arrival times of strong earthquakes in predefined seismic zones of the above referred regions. The method introduced allows to determine the uncertainties involved, which are expressed as percentages of the earthquake mean return period. The determination in this way is very efficient because one may calculate uncertainties on the same time scale. It is also shown that the final maximum Bayesian probabilities of the inter-arrival times in the several seismic zones are dependent on the data set used and particularly on its time length. Comparisons between the predicted and the real time of earthquake occurrences are finally made in order to evaluate the correlation between them.

### ΣΥΝΟΨΗ

Εκτιμώνται οι πιθανότητες γένεσης ισχυρών ( $M \geq 6.5$ ) σεισμών στις σεισμικά ενεργές περιοχές του Μεξικού, της κεντρικής και της νότιας Αμερικής. Η ευθεία προσέγγιση της στατιστικής Bayes εφαρμόζεται με σκοπό την εύρεση των ενδιάμεσων χρόνων ισχυρών σεισμών σε προκαθορισμένες σεισμικές ζώνες των προαναφερθεισών περιοχών. Η μέθοδος που εισάγεται επιτρέπει τον προσδιορισμό των εμπλεκόμενων αβεβαιοτήτων, οι οποίες εκφράζονται ως ποσοστά της μέσης περιόδου επανάληψης των σεισμών. Ο προσδιορισμός με αυτό τον τρόπο είναι πολύ αποτελεσματικός επειδή οι αβεβαιότητες μπορούν να υπολογιστούν στην ίδια χρονική κλίμακα. Επίσης δείχνεται ότι οι τελικές μέγιστες πιθανότητες των ενδιάμεσων χρόνων, κατά Bayes, εξαρτώνται από τα δεδομένα και ειδικότερα από τη χρονική τους διάρκεια. Τελικά, γίνονται συγκρίσεις μεταξύ των προβλεπομένων και των πραγματικών χρόνων γένεσης, με σκοπό να εκτιμηθεί η συσχέτιση μεταξύ τους.

**KEY WORDS:** Bayes approach, probabilistic prediction, inter-arrival times, Mexico, south and north America.

### 1. INTRODUCTION AND DATA USED

The theory of Bayesian probability has been extensively used in the past to approach seismological problems mainly to determine conditional probabilities of earthquake occurrences. Campbell (1982, 1983) proposed a Bayesian extreme value distribution of earthquake occurrence to evaluate the seismic hazard along San Jacinto fault. A similar procedure has been applied by Stavrakakis and Tselentis (1987), for a probabilistic prediction of strong earthquakes in Greece.

Ferraes (1985; 1986) used a Bayesian analysis to predict the inter-arrival times for strong earthquakes along the Hellenic arc, as well as for Mexico. An alternative view of Ferraes research was presented by Papadopoulos (1987), for the occurrence of large shocks in the east and west segments of the Hellenic arc. Stavrakakis and Drakopoulos (1995) adopted the Bayesian extreme value distribution of earthquake occurrence in order to estimate the seismic hazard in some seismogenic zones in Greece and the surrounding area.

The examined regions were divided in thirteen seismic zones or seismic sources according to zoning proposed primarily by Papadimitriou (1993), Papazachos et al. (1997) and Cernadas et al. (1998). Given that a large number of events is needed to avoid instability in the applied method, we modified slightly the seismic zoning proposed by the above mentioned authors by using as criterion the spatial clusters of the earthquakes epicenters. The zones and epicenters of shallow events finally adopted are depicted in figure (1).

For the purpose of the present research, shallow ( $h \geq 60$  Km), mainshocks of magnitude  $M \geq 6.5$  that occurred

1. Aristotle University of Thessaloniki, School of Geology, Geophysical Laboratory, 54006 Thessaloniki, Greece.

2. National Observatory of Athens, Institute of Geodynamics, 11810 Athens, Greece.

in the time interval from 1900 to 1996 inclusive are considered. This earthquakes are extracted from the catalog constructed by Tsapanos et al. (1990). This catalog was improved by considering: a) the magnitudes listed by Pacheco and Sykes (1992) and b) the revised world seismicity catalog given by Puez (1999).

The exclusion of foreshocks and aftershocks was made by the following criteria (Acharya 1979): All events that occurred 40 days before the main event (Jones and Molnar, 1979) in a radius of  $L=10^{0.5M_S-1.8}$  were considered as foreshocks, where  $M_S$  is the surface magnitude of the main event. All events that occurred 365 days after the main shock (Tajima and Kanamori, 1985) were considered as aftershocks.

**2. METHOD APPLIED**

In the method applied here, introduced by Papadopoulos (1987) the inter-arrival times for each zone were extracted from the catalogue of main shocks. Assigning a prior probability to each inter-arrival time and assuming an exponential distribution of inter-arrival times, as one would expect from the random (Poissonian) model, we get that the prior probability is:

$$P' (T_j) = 1 - \exp (-\lambda T_j) \tag{1}$$

Where  $\lambda$ , the mean rate of earthquake occurrences, calculated directly by:

$$\lambda = \frac{1}{T} = \frac{1}{\sum_{j=1}^n T_j / n} \tag{2}$$

As the Poisson distribution is assumed for the main shock time distribution, the likelihood function  $f(T_j)$  is the probability of only one event occurrence in a given inter-arrival time, that is:

$$f (T_j) = \lambda T_j \exp (-\lambda T_j) \tag{3}$$

Then, the posterior probability can be calculated by the Bayes' theorem:

$$P'' (T_j) = \frac{[1 - \exp (-\lambda T_j)] [\lambda T_j \exp (-\lambda T_j)]}{\sum_{j=1}^n [1 - \exp (-\lambda T_j)] [\lambda T_j \exp (-\lambda T_j)]} \tag{4}$$

For each seismic zone the three inter-arrival times with the highest posterior probability are selected and they are added to the occurrence time of the last event,  $E_n$ , in the main shock catalogue. Thus, the three most likely times of occurrence for the next event,  $E_{n+1}$ , are estimated.

The results of this method are summarized in Table (1). In the first three columns the zone code number, the number of events used and the mean rate of event occurrence,  $\lambda$ , are listed. In the fourth column the three most probable inter-arrival times are demonstrated, while in the fifth column presents the corresponding posterior Bayesian probabilities  $P''$ . It can be shown (Ferraes, 1985; 1986) that  $P''$  is a measure of the probability that T will be included in a small range around the point  $T_j$ . In the last column there are the three estimated times of occurrence of the next event in the zone that correspond to the inter-arrival times in the fourth column.

**TABLE 1. Results of the alternative Bayesian procedure. The three most probable inter-arrival times and their**

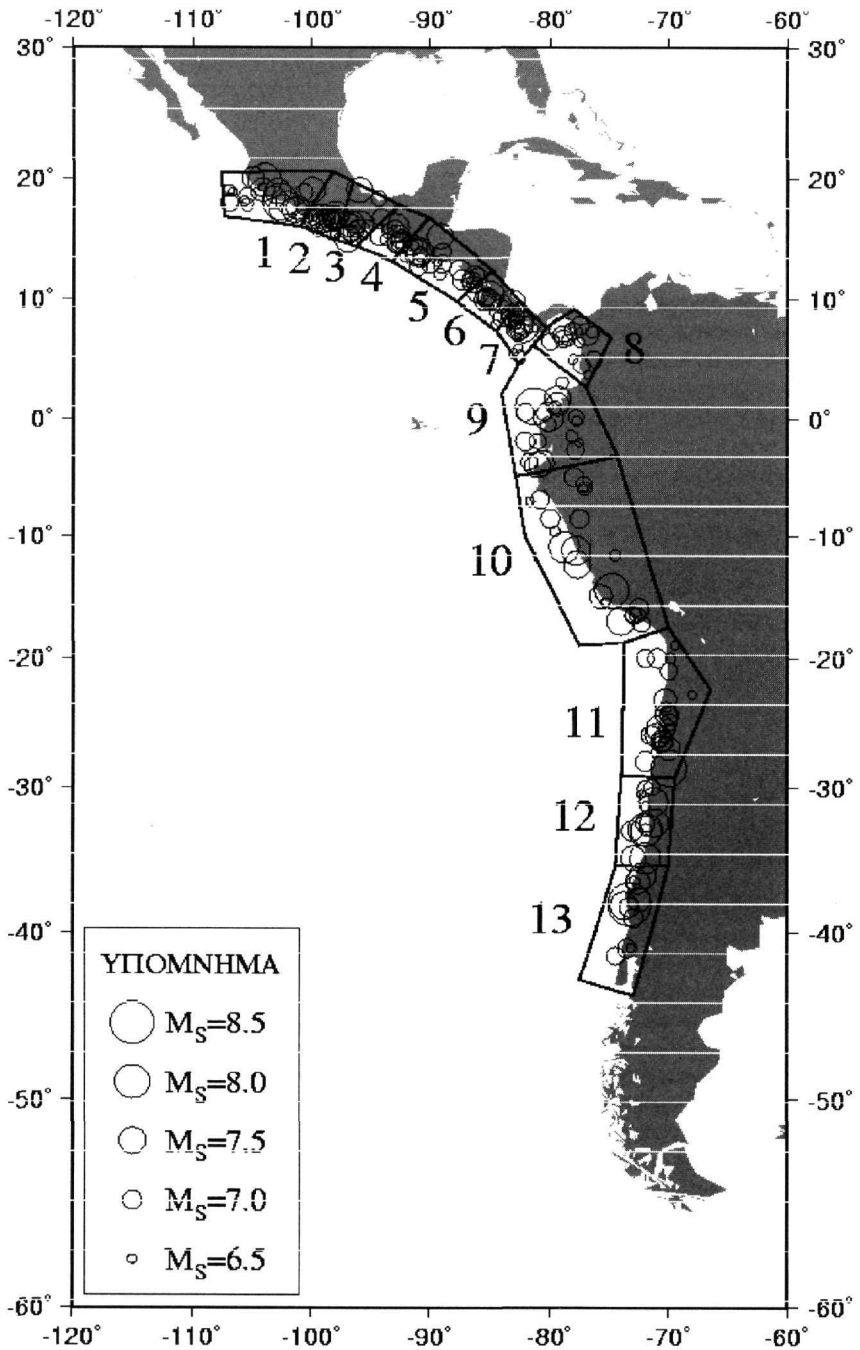


FIGURE 1. Seismic zonation adopted by the present work. Circles denote main shocks used (after Galanis, 2001).

corresponding probabilities are presented for each zone.

Zone	Number of events (n)	$\lambda$	Inter-arrival time (T)	Posterior Bayesian Probability P'	Estimated time of occurrence of $E_{n+1}$
Zone 1	20	0.1985	6.8713	0.0885	2002, Aug 24
			6.5476	0.0879	2002, Apr 28
			6.3775	0.0874	2002, Feb 25
Zone 2	19	0.1875	8.2233	0.0898	2003, Dec 5
			6.8830	0.0878	2002, Aug 3
			10.0372	0.0859	2005, Sep 28
Zone 3	14	0.1889	7.7058	0.1277	1990, Oct 8
			5.9931	0.1213	1989, Jan 21
			10.2287	0.1174	1993, Apr 17
Zone 4	13	0.1319	11.0031	0.1398	2004, Sep 12
			10.2103	0.1392	2003, Nov 26
			10.0119	0.1388	2003, Sep 15
Zone 5	25	0.2453	5.6972	0.0800	2001, Nov 14
			5.0442	0.0784	2001, Mar 21
			3.8056	0.0685	1999, Dec 24
Zone 6	13	0.1321	11.5903	0.1363	2002, Nov 25
			8.8300	0.1314	2000, Feb 21
			8.0192	0.1260	1999, May 1
Zone 7	18	0.2172	7.1004	0.1289	1990, May 10
			7.3832	0.1281	1990, Aug 22
			5.7395	0.1269	1988, Dec 30
Zone 8	13	0.1352	9.6995	0.1705	2002, Jul 1
			8.1548	0.1615	2000, Dec 14
			15.3555	0.1504	2008, Feb 26
Zone 9	19	0.1900	7.5813	0.0862	2003, May 4
			7.2331	0.0860	2002, Dec 27
			8.5748	0.0850	2004, May 1
Zone 10	21	0.2402	6.2902	0.0848	2003, Feb 26
			5.5924	0.0846	2002, Jun 17
			4.8781	0.0818	2001, Sep 29
Zone 11	22	0.2291	6.9591	0.0753	2002, Jul 16
			7.2197	0.0747	2002, Oct 19
			5.1695	0.0734	2000, Oct 1
Zone 12	13	0.1391	11.0361	0.1342	2003, Dec 11
			12.0514	0.1318	2004, Dec 17
			7.7329	0.1250	2000, Aug 22
Zone 13	9	0.4104	12.5114	0.2675	1987, Nov 14
			13.7913	0.2541	1989, Feb 25
			5.8223	0.2186	1981, Mar 8

### 3. RESULTS AND CONCLUSIONS

In order to test the reliability of the method it was used to reproduce the occurrence time of the last event in each zone. For this purpose, in each zone all the main events but the last one,  $E_n$ , were used as data for the application of the two methods.

The results of the test are summarized in Table (2). In the first two columns there are the names of the zones and the average inter-arrival time. In the next three columns one may find the three most likely estimations obtained. The sixth column shows the date obtained by adding the average inter-arrival time to the time of event  $E_{n-1}$ , which is an independent non-probabilistic estimation that can be compared to the estimations made by the method. The last column presents the actual time of occurrence, that is the date of the last event,  $E_n$ , introduced in the catalogue.

**TABLE 2. Estimated times of occurrence according to the alternative Bayesian procedure. Three times of occurrence are estimated for each zone. These times are compared to the time that corresponds to the mean return period as well as the actual time of occurrence.**

Zone	Average inter-arrival time (T)	Estimated time of occurrence 1	Estimated time of occurrence 2	Estimated time of occurrence 3	Time of $E_{n-1}$ event plus average T	Actual time of occurrence of $E_n$ event
Zone 1	4.98	1996.53	1996.21	1996.04	1994.64	1995.77
Zone 2	5.82	2002.04	2003.85	2000.70	1999.64	1995.70
Zone 3	5.39	1986.62	1984.91	1989.14	1984.30	1983.93
Zone 4	6.15	1980.34	1980.54	1981.33	1976.47	1993.69
Zone 5	4.10	1998.37	1997.71	1996.48	1996.77	1996.17
Zone 6	8.16	2001.82	2005.09	1999.06	1998.39	1991.31
Zone 7	4.85	1989.73	1990.02	1988.37	1987.49	1983.26
Zone 8	7.98	2001.58	2007.24	2000.04	1999.87	1992.80
Zone 9	5.07	1994.41	1994.76	1993.57	1992.25	1995.76
Zone 10	4.34	2002.43	2001.73	2001.02	2000.48	1996.86
Zone 11	4.23	1993.79	1995.84	1996.39	1992.86	1995.58
Zone 12	7.14	1996.21	1997.23	1992.63	1992.32	1992.91
Zone 13	7.92	1987.14	1988.42	1994.47	1982.55	1975.36

The four predicted occurrence times and the occurrence times of the actual events,  $E_n$ , are compared. This is illustrated in Table (3). The inconsistency is expressed as a percentage of the average inter-arrival time,  $\bar{T}$ , for each zone. This allows comparisons between the accuracy of the estimations in different zones with different occurrence rates.

The average errors involved are slightly smaller than the mean inter-arrival times. However, the error standard deviations (SD's) are rather high being as a rule on the same order with the corresponding errors. It must also be noted that in most cases the actual events occur before the time predicted by the first method

Finally, the errors are plotted against the size of the sample of each zone. This is demonstrated in figure (2). The errors are plotted as data points, with error on the vertical axis and number of observations on the horizontal axis. Linear regression is applied in each of the four different data sets of the figure. Numbers 1, 2 and 3 denote the first, second and third prediction of the alternative Bayesian approach. Number 4 denotes the prediction of the mean return period. The coefficients of determination are numbered similarly and they appear in the top right hand corner. The coefficient of determination is the square of the correlation coefficient.

The slopes of the least squares lines are shown in Table (4). Because it is necessary for the lines' slopes (in fig. 2) to be negative, three upper confidence limits for the slopes were also computed. It is observed that the third estimated time of occurrence is more consistent than the other two. This means that at least three estimations are needed when applying this method.

The upper confidence limits for the slope are negative for all confidence levels, with one exception. The high significance of the slope of the lines manifests that this is negative.

Comparisons made to show how significant are the obtained results, that is the predicted times of the next earthquake occurrences, lead to the following conclusions.

TABLE 3. Errors of the method. The errors are expressed as percentages of the mean return period of each zone.

Zone	Error in estimated time of occurrence 1	Error in estimated time of occurrence 2	Error in estimated time of occurrence 3	Error in last event plus average T
Zone 1	15.26	8.76	5.35	22.76
Zone 2	108.81	139.96	85.79	67.57
Zone 3	49.93	18.14	96.74	6.93
Zone 4	217.30	214.07	201.17	280.22
Zone 5	53.43	37.51	7.32	14.54
Zone 6	128.88	168.89	95.05	86.84
Zone 7	133.48	139.31	105.45	87.21
Zone 8	110.06	180.90	90.72	88.59
Zone 9	26.47	19.60	43.14	69.18
Zone 10	128.13	112.07	95.63	83.34
Zone 11	42.25	6.17	19.20	64.34
Zone 12	46.26	60.48	3.92	8.29
Zone 13	148.74	164.89	241.21	90.83
Average	86.36	90.77	77.91	69.33
St Dev.	59.10	75.00	72.87	69.87

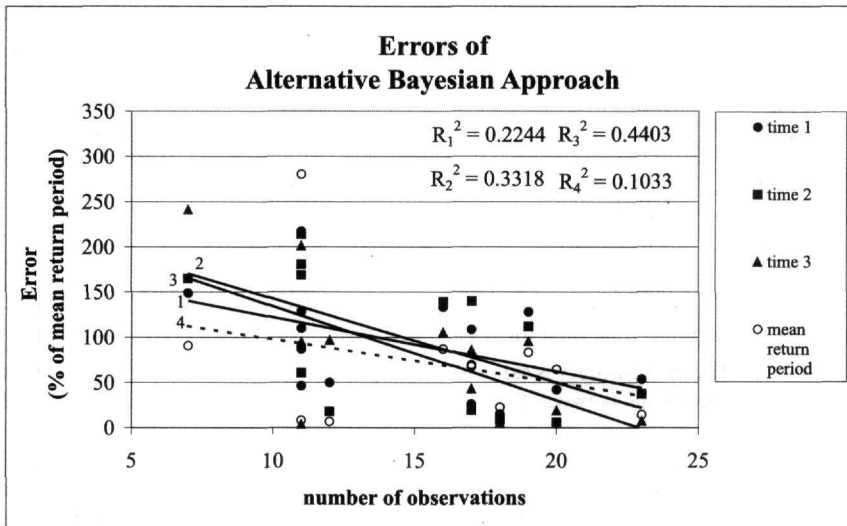


FIGURE 2. Errors of the alternative Bayesian approach. Lines 1, 2 and 3 are the least squares lines of the errors of the first, second and third most probable times (denoted by closed circles, squares and triangles respectively). Dotted line 4 is the least square line of the errors of the non-probabilistic calculation described (denoted by open circles).

**TABLE 4. Slopes of the lines of figure (2) and their upper confidence limits, for confidence levels of 80%, 90% and 95%.**

Least squares line	Slope of least squares line	Upper confidence limits		
		80%	90%	95%
1	-6.0173	-3.0639	-1.4181	0.0407
2	-9.2859	-5.8070	-3.8683	-2.1500
3	-10.3936	-7.3000	-5.5760	-4.0480
4	-4.8269	-1.0726	1.0195	2.8739

Linear regression (fig. 2) verifies the hypothesis that in the application of the method, the errors depend on the sample size. Namely, the larger the sample, the smaller the error. This is also true of the simplistic approach of adding the average inter-arrival time to the occurrence time of the last event (number 4 in fig. 2).

The method is consistent. The dependence of the error on the size of the sample is strong, as can be seen by the slopes in the trendlines. The slopes obtained by the method are clearly negative. The consistency of the simplistic, non-probabilistic calculation is lower than the consistency of the alternative Bayesian approach.

It was proved that in the application of the method it is important to examine not only one but at least three values of probable inter-arrival time. The results implied that even the third most likely inter-arrival time is as consistent as the first and the second and it can not be ignored.

Finally it is concluded that the method (Papadopoulos 1987) is both consistent and significant.

## REFERENCES

- ACHARYA, H.K. (1979). Regional variations in the rupture-length magnitude relationships and their dynamical significance. *Bull. Seismol. Soc. Am.*, 69, 2063-2084.
- CAMPBELL, K.W. (1982). Bayesian analysis of extreme Earthquake occurrences. Part I. Probabilistic hazard model, *Bull. Seismol. Soc. Am.*, 72, 1689-1705.
- CAMPBELL, K.W. (1983). Bayesian analysis of extreme Earthquake occurrences. Part II. Application to the San Jacinto fault zone of Southern California, *Bull. Seismol. Soc. Am.*, 73, 1099-1115.
- CERNADAS, D., OSELLA, A. and SABBIONE, N. (1998). Self similarity in the seismicity of the South American subduction zones, *Pageoph*, 152, 57-73.
- FERRAES, S.G. (1985). The Bayesian probabilistic prediction of strong earthquakes in the Hellenic Arc, *Tectonophysics*, 11, 339-354.
- FERRAES, S.G. (1986). Bayes theorem and the probabilistic prediction of inter-arrival times from strong earthquakes felt in Mexico City, *J. Phys. Earth*, 34, 71-83.
- GALANIS, O.CH. (2001). Probabilistic estimation of seismicity of the regions of Mexico, central and south America using the Bayes statistics, Msc Dissertation (in Greek), Univ. of Thessaloniki, 97 pp.
- JONES, L.M. and MOLNAR, P. (1979). Some characteristics of foreshocks and their possible relationship to earthquake prediction and premonitory slip of faults, *J. Geophys. Res.*, 84(B7), 3596-3608.
- PACHECO, J. and SYKES, L.R. (1992). Seismic moment catalog of large shallow earthquakes, 1900 to 1989, *Bull. Seism. Soc. Am.*, 82, 1306-1349.
- PAPADIMITRIOU, E.E. (1993). Long-term Earthquake Prediction along the Western Coast of South and Central America Based on a Time Predictable Model, *Pageoph*, 140, 301-316.
- PAPADOPOULOS, G.A. (1987). An Alternative view of the Bayesian probabilistic prediction of strong shocks in the Hellenic Arc, *Tectonophysics*, 132, 311-320.
- PAPAZACHOS B.C., PAPADIMITRIOU, E.E., KARAKAISIS, G.F. and PANAGIOTOPOULOS D.G. (1997). Long-term Earthquake Prediction in the Circum-Pacific Convergent Belt, *Pageoph*, 149, 173-217.
- PEREZ, O. (1999). Revised world seismicity catalog (1950-1997) for strong ( $M_s > 6$ ) shallow ( $h < 70$  km) earthquakes, *Bull. Seismol. Soc. Am.*, 89, 335-341.
- STAVRAKAKIS, G.N. and DRAKOPOULOS, J. (1995). Bayesian Probabilities of Earthquake Occurrences in Greece and Surrounding Areas, *Pageoph*, 144, 307-319.
- STAVRAKAKIS, G.N. and TSELENTIS, G.A. (1987). Bayesian Probabilistic Prediction of Strong Earthquakes in the main seismogenic zones of Greece, *Boll. Geof. Teor. Appl.*, 29, 51-63.
- TAJIMA, F. and KANAMORI, H. (1985). Global survey of aftershock area expansion patterns, *Phys. Earth Planet. Inter.*, 40, 77-134.
- TSAPANOS, T.M., SCORDILIS, E.M. and PAPAZACHOS, B.C. (1990). A global catalogue of strong ( $M > 5.5$ ) earthquakes during the time period 1897-1985, *Public. of Geophysical Lab., Univ. Of Thessaloniki*, 90pp.