Bulletin of the Seismological Society of America, Vol. 87, No. 4, pp. 799–808, August 1997

Evaluation of the Global Applicability of the Regional Time- and Magnitude-Predictable Seismicity Model

by Constantinos B. Papazachos and Eleftheria E. Papadimitriou

Abstract Two main issues are examined that concern the global applicability of two models that associate earthquake recurrence intervals and the magnitude of the preceding or following event of each interval, namely, the regional time- and magnitude-predictable models. Specifically, the statistical significance of the results that are obtained by application of these models to a high-quality data set is tested, and then the effect of the seismic zonation process on these results is examined. A simple application of the time-predictable model to different seismogenic regions using data from a high-quality global data set leads to results that satisfy the 95% confidence limit for different obtained parameters in only 25% of the cases. Using simple Monte Carlo simulation, it is shown that these statistical significance tests often fail simply because of the errors in the magnitude determination and the small magnitude range spanned by the available data. Moreover, these tests examine each seismogenic region separately and ignore the global applicability of such a model. For this reason, a procedure that incorporates the whole global data set is applied. The results and the detailed statistical analysis demonstrate the consistency of the behavior of this recurrence model, as this was established from earlier regional studies. Application of the model to an independent data set shows that the results are robust when different seismic zonation techniques are used. On the contrary, the slip-predictable model is rejected using the same procedure. These results suggest that the time- and magnitude-predictable models can generally be used for practical purposes and hazard estimates in active seismogenic regions, provided that an appropriate data sample, as defined in the present article, is available for each region.

Introduction

Recurrence time-distribution models are commonly used to estimate the hazard due to large earthquakes. The most well studied and widely discussed such models are the time- and slip-predictable models (Shimazaki and Nakata, 1980). Many relevant studies have been performed, dealing mainly with large earthquakes in simple plate boundaries or in single faults, indicating that repeat times are not randomly distributed but follow the time-predictable model (Bufe *et al.,* 1977; Sykes and Quittmeyer, 1981; among others). According to this model, the time between two large earthquakes is proportional to the displacement of the preceding large event. Laboratory experiments on stick-slip behavior on pre-existing faults also favor this model (Sykes, 1983). There are, however, some other studies that do not favor such kinds of models (Davis *et al.,* 1989; Kagan and Jackson, 1991).

In most parts of the world, the historical record of seismicity is too short to enable firm conclusions to be made about the repeat time of large earthquakes or possible clustering of large events both in space and time. Thus, due to the lack of data for large earthquakes on any individual fault or segment, a generic distribution is estimated by combining data from a number of faults (Jacob, 1984; Nishenko and Buland, 1987), or a more elaborate stochastic model is assumed (Kiremidjan and Anagnos, 1984).

More recently (Papazachos, 1989, 1992; Papazachos and Papaioannou, 1993), the regional time- and magnitude-predictable model has been proposed. This model differs from the time-predictable model proposed by Shimazaki and Nakata (1980) in that it holds for seismogenic regions (or sources) that include the main fault, where the largest earthquakes occur, as well as other smaller faults, where smaller earthquakes occur. It is expressed by two relations that give the repeat time, T_t, and the surface-wave magnitude, M_f, of the next expected mainshock in that source, as a function of the minimum magnitude considered in the data set, $M_{\min}$; the magnitude of the last mainshock, M_p, which is larger than $M_{\min}$; and the annual moment rate released in that source, m_0. These two relations have the form:

$$\log T_t = bM_{\min} + cM_p + d \log m_0 + q \quad (1)$$

$$M_f = BM_{\min} + CM_p + D \log m_0 + m \quad (2)$$

where b, c, d, q, B, C, D, and m are parameters that must be determined. The model has been tested in several parts of the continental fracture system (Papazachos and Papaioannou, 1993; Papazachos *et al.*, 1994; Papadimitriou, 1993, 1994a, 1994b; Karakaisis, 1993, 1994a, 1994b; Panagiotopoulos, 1993, 1995a, 1995b).

The application of this methodology overcomes the problem related to the limited number of interevent times available for a single fault or segment. The basic features of the model (positive values for b, c, q, B, and D parameters, and negative values for d, C, and m parameters), as well as its superiority in comparison with time-independent models, have been discussed extensively (Papazachos *et al.*, 1995, 1996).

The approach taken here attempts to quantify the uncertainties and check the assumptions involved in the calculations of the model parameters. The main assumptions and ideas of the model are discussed. Emphasis will be given to the parameter uncertainties, on the basis of a statistical treatment, in order to demonstrate the robustness of the fundamental ideas underlying the proposed model.

Data Used and Problem Formulation

Papazachos and colleagues (1995, 1996) separated the continental fracture system (Circum–Pacific and Alpine–Himalayan Belt) into 274 seismogenic regions on the basis of several seismotectonic and geomorphological criteria (spatial clustering of seismicity, dimensions of rupture zones of large earthquakes, evidence for interaction between seismic events, topography variations, etc.). These regions belong to broader seismic areas (e.g. Japan, Greece, etc.). For each one of these seismogenic regions, the more complete and accurate earthquake catalogs were chosen and used in estimating the interevent times of the mainshocks. The largest earthquakes that occurred during a clustering period in a seismogenic region were considered as mainshocks. Each period consists of a period of accelerating seismic release, a large earthquake, and a period of decelerating seismic release. The duration of the clustering period is the sum of the "preshock," t_p, and "postshock," t_a, seismic activity, which are

$$t_p = 3 \text{ years} \quad \text{and} \quad \log t_a = 0.06 + 0.13M_p,$$

where M_p is the magnitude of the preceding mainshock. The time that elapsed between two successive mainshocks is taken to be the interevent time. The data sample obtained after the declustering of the data consisted of 1811 sets (T_t, $M_{\min}$, M_p, and M_f). For each individual seismogenic region, one or more data sets, which will be called seismic groups hereafter, were derived as follows: The smallest mainshock (e.g., $M = 6.0$), according to the completeness, was considered as the first value for $M_{\min}$, and then all the possible interevent times were estimated. Afterward, the next smallest mainshock ($M > 6.0$) was considered as the new larger $M_{\min}$ value, and the corresponding interevent times were recalculated since the earthquake data set becomes smaller. In that way, all the possible interevent times concerning mainshocks were determined for different lower magnitude ($M_{\min}$) levels. Since the interevent times were obtained for a broad range of $M_{\min}$ values for each seismogenic region, this procedure ensures the robustness of our results to different choices of $M_{\min}$ for this region.

Using this data set, Papazachos *et al.* (1995) obtained the following parameter values for the regional time- and magnitude-predictable model: $b = 0.19$, $c = 0.33$, $d = -0.39$, $B = 0.73$, $C = -0.28$, and $D = 0.40$. The values of the parameters q and m vary from area to area. Most of the parameters in equations (1) and (2) and their values are self-explanatory. For instance, the inclusion of the $M_{\min}$ term in equation (1) is essentially the Gutenberg–Richter law, since for larger cutoff magnitudes, $M_{\min}$, we have smaller $\log N$ (N = number of events with $M > M_{\min}$) and therefore larger $\log T \sim \log N^{-1}$. The crucial point of equations (1) and (2) is the inclusion of the linear M_p term and the values of constants c and C, which represent the heart of the time- and magnitude-predictable model. Hence, the positive global ($c = 0.33$) value indicates that large events lead to large interevent times, whereas the negative ($C = -0.28$) value indicates that large events tend to be followed by smaller events and vice versa.

There are three main issues associated with these models: (1) the theoretical justification behind the extension of the Shimazaki and Nakata (1980) model to a zone of faults rather than a single fault or plate boundary, (2) the statistical significance of the estimation of parameters c and C, and (3) the problem of seismic zonation and determination of seismic regions within which equations (1) and (2) apply and the sensitivity of the results on the specific zonation technique applied. As far as the first problem is concerned, we refer to the model recently proposed by Mulargia and Gasperini (1995). The purpose of the present study is to address the second and third issue and, therefore, check the statistical significance of the observation that the value of the parameters c and C are positive and negative, respectively.

Estimation of the Model Parameters

The main problem of equations (1) and (2) is the strong correlation between $M_{\min}$, M_p, and m_0, since $M_p \geqq M_{\min}$ and we usually have large M_p and $M_{\min}$ only for large m_0. Although methods exist for regression analysis in such cases, one cannot avoid the strong correlation of the estimated parameters b, c, and d of equation (1), especially between the first two. This makes almost impossible the unambiguous evaluation of the statistical significance of the $c\,M_p$ term in this equation. For this reason, we decided to study this term

separately, even if it is estimated by only a limited number of data concerning only one seismogenic source. Since for each seismic group both terms including $M_{\min}$ and m_0 are constant, we examined the following relation:

$$\log T_t = cM_p + a, \tag{3}$$

where a depends on $M_{\min}$, m_0, and the seismogenic region studied. It is clear that the number of observations for this determination in each case is very important for the accuracy of our results. Numerical simulations, which are shown later, indicated that the minimum number of data that should be considered was equal to 5. Therefore, from the whole data sample, only the seismic groups with five or more observations were taken into account. Another important factor in these calculations is the range of magnitudes spanned by the available mainshock data. In order to obtain a reliable estimation of the parameter c, the range of the values of M_p was taken to be equal to or larger than 0.6, as indicated from the same numerical simulations. This subsetting was necessary, since the accuracy of our calculations strongly depends on the number of interevent times and the magnitude range. A total of 48 seismic groups satisfying these conditions was found. For each seismic group, parameter c and the linear correlation coefficient, r, were estimated, as well as the value of the t-test for the hypothesis H:{$c = 0$} and the corresponding probability for this hypothesis. All the results are shown in Table 1, along with the name of the seismogenic region and a reference from which the original data were drawn.

Statistical Evaluation of the Results for a Single Seismic Group

If we want to assess the statistical importance of the previous results, we can check the significance of the linear correlation coefficient, r, and of the slope, c, of equation (3). Regarding the linear correlation coefficient, we tested the hypothesis H_0:{$r = 0$}. We rejected H_0 using the same confidence limit of 95% ($r^2 = 0.7$) as other researchers have previously used (Mulargia and Gasperini, 1995). We notice that for our final results, the estimated correlation coefficient exceeded the value of $r^2 = 0.7$ ($r = 0.837$) in only 25% of the seismic groups. Therefore, although in almost all cases c is nonzero and positive, we cannot prove the linear dependence of $\log T_t$ on M_p (equation 3), for each group separately, with the necessary statistical confidence. In order to explore the source of this poor statistical confidence, we examined the behavior of r as a function of the magnitude range that the data span in each case, for each of three different data groups (Figs. 1a through 1c). Each of these groups was defined on the basis of the number of interval times (5, 6, or 7) that was available for the calculations. It is observed that, as expected, the average value of r increases, while its corresponding variance decreases with the number of data that was used in the regression. Moreover, it is interesting to notice that the r values become more robust as the magnitude range covered by the mainshock data increases. This observation indicates that the relative small number of seismic groups that fulfill the 95% confidence limit criterion is probably due to the small magnitude range that the data cover for each seismic group.

In order to further demonstrate the previous conclusion, we performed a Monte Carlo simulation for the following relation:

$$\log T = \bar{c}M_p + a, \tag{4}$$

where $\bar{c}$ is the average value ($=0.33$) of equation (1), calculated in a previous work (Papazachos *et al.*, 1995). For the simulation, we assumed no errors for relation (4) ($\sigma_{\log T} = 0$) and a typical value of 0.3 for the error of M_p. For each group (5, 6, or 7 data points), we used randomly distributed events that follow equation (4) and cover different magnitude ranges (0.2 to 3.0), and we calculated the linear correlation coefficient. Using the Fisher's transformation

$$z = \frac{1}{2}\ln\frac{1 + r}{1 - r}, \tag{5}$$

which follows an approximately normal distribution, we can calculate the average r values and the upper and lower limits that correspond to $\pm\sigma_z$ (68% confidence limit). These curves are also plotted in Figures 1a, 1b, and 1c. We see that these curves predict approximately the same behavior observed for the source data. Of course, this fit between prediction and data does not suggest that our assumption about the validity of the time-predictable model (equation 4) is correct. However, these curves do imply that the intrinsic limitations of our data (magnitude range spanned by the data, magnitude errors, number of data) will not allow the verification (with 95% confidence) of equation (4). In order to overcome this problem, we might attempt to increase the number of data and magnitude range by incorporating older historical data (when they are available). However, in such a case, the corresponding magnitude error is much larger than the value used here. Hence, we probably will not improve the fits. On the other hand, Figure 1 suggests that we should be skeptical about statistical rejection of equation (4), given the problems related to the magnitude range and errors of the available seismological data.

Similar results are deduced for the slope, c, of equation (4). In Figure 2, we plot the average confidence for the estimated c values, calculated using the same approach for 5 and 6 data points (interval times). We also observe that we would need a magnitude range larger than 1.3 and 1.7 for 6 and 5 data points, respectively, in order to obtain an average 95% confidence estimation for c. This translates to the conclusion that for cases with 5 or 6 data points that span the previous magnitude ranges, only for 50% of the seismic groups would we find a confidence limit greater or equal to 95%. Again, this observation indicates that statistical tests

Table 1

Information and results for the 48 seismic groups of data from 31 regions for which data were used: M_{min} is the minimum magnitude for the data set of each seismic group; ΔM_p is the magnitude range spanned by the data; n is the number of data for each seismic group; c is the slope estimated for each seismic group for the time-predictable model applied to each seismic group separately; r is the corresponding linear correlation coefficient; and T_c and Q are the t-test value and corresponding probability for this estimation. The values of the constants that correspond to each seismic group for the global time-predictable model (equation 8), a_j, and the magnitude-predictable model (equation 20), A_j, as well as their corresponding errors, σ_{a_j} and σ_{A_j}, are given in the four following columns. In the final two columns, the region for each data source and the corresponding reference from which data were taken are given.

Source Parameters			Time-Predictable Model for Each Source				Source Parameters for Global Time and Magnitude-Predictable Models					
M_{min}	ΔM_p	n	c	r	T_c	Q	a_j	σ_{a_j}	A_j	σ_{A_j}	Region	Ref.
5.5	0.9	7	0.47	0.60	1.68	0.154	−1.19	0.22	7.64	0.40	WM-12 Dalmatian Coasts	1
5.6	0.8	6	0.20	0.29	0.60	0.583	−1.13	0.23	7.72	0.41	WM-12 Dalmatian Coasts	1
6.3	1.0	5	0.70	0.73	1.87	0.158	−1.34	0.26	9.06	0.46	GR-7 Cephalonia island	2
6.0	0.7	6	0.43	0.36	0.77	0.484	−1.11	0.24	8.27	0.43	GR-8 Zakinthos island	2
6.1	0.6	5	0.44	0.28	0.51	0.647	−1.12	0.24	8.32	0.43	GR-8 Zakinthos island	2
6.2	0.7	5	1.14	0.94	4.73	0.018	−1.20	0.25	8.66	0.45	GR-49 Samos island	2
5.8	0.7	5	−0.03	−0.02	−0.04	0.973	−1.28	0.24	8.19	0.42	ME-5 Frug	3
6.2	1.1	6	0.34	0.65	1.71	0.162	−1.34	0.25	9.08	0.46	ME-15 Kopet-Dag	3
6.3	1.0	5	0.62	0.91	3.84	0.031	−1.39	0.26	9.14	0.47	ME-15 Kopet-Dag	3
5.6	1.9	5	0.13	0.38	0.71	0.526	−1.41	0.24	8.65	0.43	ME-17 Dustabad	3
6.5	0.6	5	0.63	0.89	3.30	0.046	−1.68	0.26	8.88	0.46	CH-10 Alma-Ata	1
7.0	1.7	5	0.09	0.41	0.79	0.487	−1.83	0.29	10.0	0.51	SA-1 South Chile	4
6.7	1.1	5	0.30	0.59	1.26	0.297	−1.73	0.27	9.55	0.48	SA-2 Valparaiso	4
7.2	0.9	6	0.55	0.67	1.80	0.146	−1.77	0.28	9.77	0.50	SA-3 Central Chile	4
7.3	0.8	5	0.45	0.46	0.89	0.438	−1.72	0.28	9.85	0.50	SA-3 Central Chile	4
7.0	0.8	5	0.50	0.51	1.04	0.376	−1.62	0.27	9.65	0.49	SA-4 North Chile	4
7.4	0.7	5	0.89	0.92	4.12	0.026	−1.76	0.28	10.0	0.51	SA-7 South Colombia	4
6.5	1.2	6	0.38	0.69	1.90	0.131	−1.86	0.27	9.36	0.48	SA-8 North Colombia	4
6.9	0.8	5	0.43	0.52	1.04	0.374	−1.84	0.27	9.54	0.49	SA-8 North Colombia	4
7.0	0.7	7	0.70	0.71	2.26	0.073	−1.79	0.27	9.68	0.49	SA-8 North Colombia	4
7.0	0.7	5	0.57	0.86	2.98	0.058	−1.51	0.27	9.45	0.49	MA-2 Costa Rica	5
7.0	1.0	7	0.49	0.64	1.88	0.118	−1.83	0.28	9.86	0.49	MA-6 Oaxaca	4
7.4	0.6	6	0.41	0.34	0.73	0.505	−1.80	0.28	9.99	0.50	MA-6 Oaxaca	4
6.6	1.5	5	0.32	0.61	1.34	0.272	−1.71	0.27	9.31	0.48	AA-2 Shumagin	6
7.0	1.1	5	0.28	0.54	1.12	0.344	−1.59	0.28	9.63	0.49	AA-2 Shumagin	6
7.0	1.2	6	0.28	0.53	1.24	0.282	−1.72	0.28	9.90	0.49	AA-3 Andreanoff	6
7.2	1.0	5	0.00	0.01	0.01	0.990	−1.67	0.28	9.89	0.50	AA-3 Andreanoff	6
7.0	1.2	5	0.28	0.63	1.40	0.257	−1.63	0.28	9.73	0.50	AA-4 Rat Islands	6
7.0	0.7	5	−0.20	−0.28	−0.51	0.643	−1.77	0.28	9.83	0.51	KK-1 NE Kamchatka	6
6.5	1.9	5	0.28	0.74	1.89	0.155	−1.69	0.27	9.43	0.48	KK-2 SE Kamchatka	6
7.1	1.5	7	0.40	0.89	4.28	0.008	−1.86	0.28	9.68	0.50	J-3 Tohoku	7
6.7	1.7	5	0.28	0.87	3.09	0.054	−1.85	0.29	10.3	0.51	J-3 Tohoku	7
7.5	0.9	5	0.32	0.62	1.38	0.260	−1.82	0.30	10.4	0.54	J-3 Tohoku	7
7.0	0.9	5	0.72	0.74	1.88	0.157	−1.61	0.28	9.71	0.50	J-4 Fukushima	7
7.0	1.1	5	0.23	0.61	1.35	0.271	−1.65	0.28	9.97	0.51	J-5 Sagami	7
6.5	1.3	6	0.49	0.85	3.17	0.034	−1.77	0.26	9.34	0.46	PH-5 T3 Taiwan	8
6.6	1.2	5	0.56	0.87	3.02	0.057	−1.78	0.27	9.42	0.48	PH-5 T3 Taiwan	8
7.0	0.6	6	0.73	0.66	1.76	0.154	−1.69	0.27	9.70	0.49	PH-19 N. Mindanao	8
6.5	1.5	7	0.33	0.62	1.75	0.141	−1.90	0.26	9.40	0.47	G-12 New Ireland	9
6.7	1.3	6	0.67	0.87	3.56	0.024	−1.93	0.27	9.56	0.48	G-12 New Ireland	9
7.0	1.0	7	0.74	0.83	3.31	0.021	−1.90	0.27	9.68	0.49	G-12 New Ireland	9
7.1	0.9	6	0.79	0.77	2.43	0.072	−1.86	0.28	9.78	0.50	G-12 New Ireland	9
7.2	0.7	5	0.60	0.53	1.09	0.355	−1.79	0.28	9.89	0.51	G-12 New Ireland	9
7.1	0.9	5	0.72	0.97	6.43	0.008	−1.85	0.28	9.91	0.50	SH-5 San Cristobal	10
6.8	0.8	5	0.35	0.42	0.79	0.487	−1.76	0.27	9.26	0.48	SH-9 Malecula	10
6.5	1.0	5	0.11	0.25	0.46	0.680	−1.60	0.26	9.11	0.47	SH-11 Tanas	10
6.7	1.4	5	0.19	0.88	3.16	0.051	−1.84	0.29	10.0	0.50	TKZ-5 N. Kermadec	11
7.2	0.9	5	0.25	0.45	0.88	0.444	−1.80	0.29	10.1	0.51	TKZ-5 N. Kermadec	11

1. Papazachos *et al.* (1995). 2. Papazachos (1992). 3. Karakaisis (1994a). 4. Papadimitriou (1993). 5. Panagiotopoulos (1995a). 6. Papadimitriou (1994a). 7. Papazachos *et al.* (1994). 8. Panagiotopoulos (1993). 9. Karakaisis (1993). 10. Panagiotopoulos (1995b). 11. Papadimitriou (1994b).

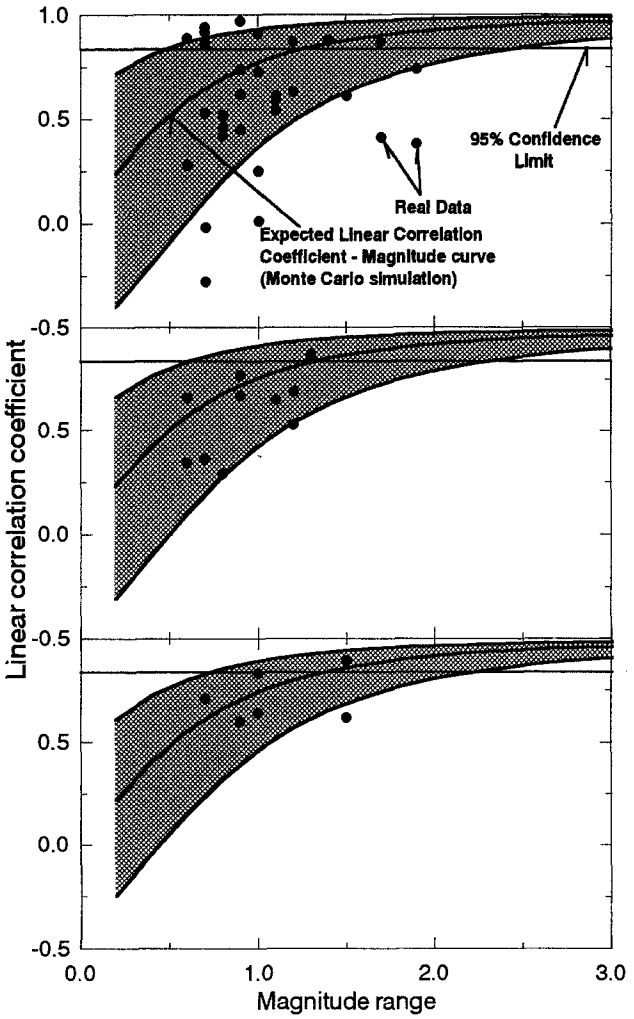

Figure 1. Plot of the calculated values of the linear correlation coefficient (solid circles) as a function of the magnitude range spanned by the available mainshock data of each group, as these were estimated for the 48 available groups. The three plots correspond to data from sources containing (a) five, (b) six, and (c) seven data points (intervent times). The expected variation of the linear correlation coefficient (solid line) and the corresponding standard deviation zone (shaded area) as calculated from a simple Monte Carlo simulation are also superimposed. In each plot, the $r = 0.837$ (95% confidence) line is also drawn. In all plots, the improvement of the r values as the magnitude range and the number of used data increase is clearly observed. A similar behavior is noticed for the results of the numerical simulations.

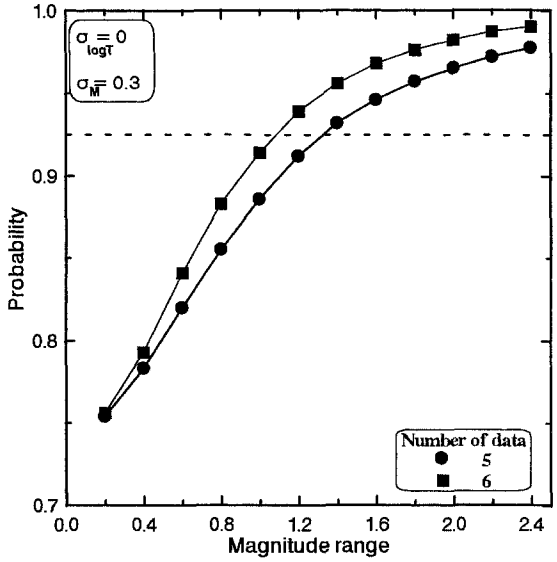

Figure 2. Plot of the confidence probability calculated from a simple Monte Carlo simulation for the slope, $\bar{c}$, of equation (4) for a test source with five or six interevent times, as a function of the magnitude range spanned by the data. A quite large range is necessary in order to obtain an average confidence probability of 95%.

for the slope of equation (4) will often fail, not because of the invalidity of the proposed relation but because of the inaccuracy of magnitudes and the relatively small magnitude range spanned by the data.

Statistical Evaluation of the Obtained c Values

The results presented in the previous section test the statistical significance of equation (4) for each seismic group. Hence, when we perform this test, we only use data from that seismic group, and we do not take into account the fact that we observe that a similar relation holds for all the other groups. In other words, the results of the previous section do not test the significance of a global applicability of equation (4) or the significance of the calculation of positive c values for each group. This is shown in Figure 3 where we plot the probability density function of c for all the seismic groups of Table 1. The c values are plotted on the horizontal axis, while the seismic group number is plotted on the vertical axis. The range $(-1,2)$ for c values is binned in intervals of 0.1, and the probability is plotted for each interval and seismic group using the c and σ_c values calculated for each group. Light and dark colors denote high and low probabilities, respectively. Seismic groups are sorted according to the estimated c values. It is clear that the $c = 0$ line falls quite off the high probability zone for most of the seismic groups. On the other hand, the weighted average of c,

$$\bar{c} = \frac{\sum_{i=1}^{n} c/\sigma_c^2}{\sum_{i=1}^{n} 1/\sigma_c^2} = 0.36, \qquad (6)$$

is well within the 95% limit of most groups.

Figure 4 shows the frequency histogram of the obtained c values. The observed distribution satisfies quite well most of the normality tests: Using a typical binning step of 0.1 results in a $\chi^2 = 3.47$ (4 degrees of freedom), which is

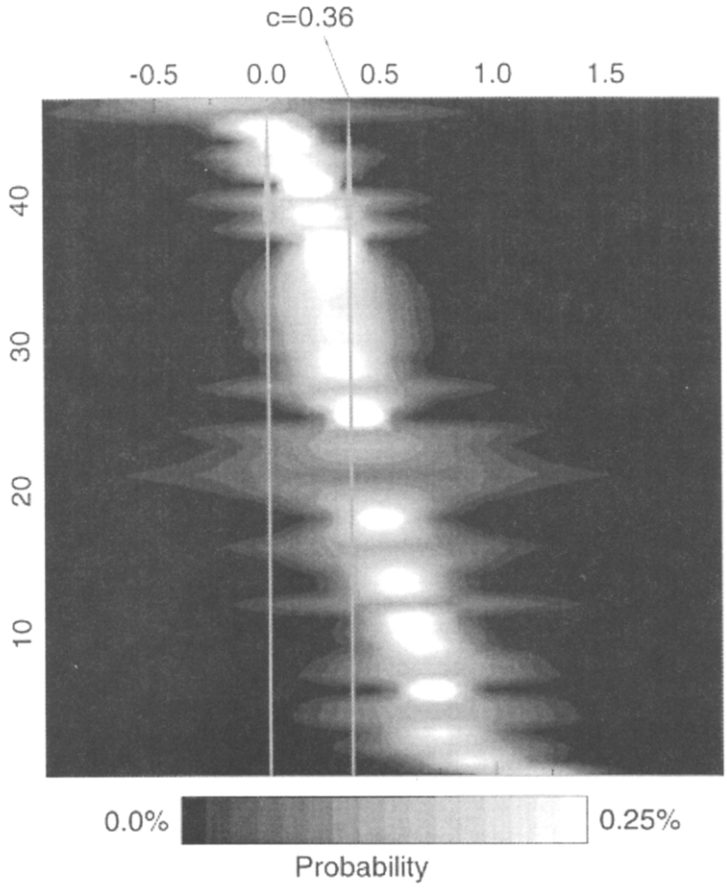

Figure 3. Plot of the probability density for the values of the constant c as calculated for each one of our 48 groups (sorted with increasing c value). The $c = 0$ and the average $c = 0.36$ lines are also plotted. The average $c = 0.36$ line intersects the high probability area of the majority of the sources that does not occur with the $c = 0$ line.

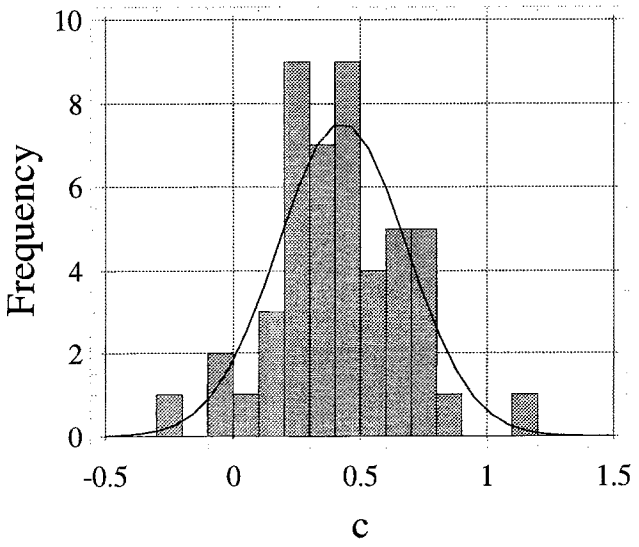

Figure 4. Frequency histogram of the obtained c values for the 48 available groups. Superimposed is the corresponding normal distribution curve.

equivalent to a probability of 48%, while other parametric (e.g., skewness) and nonparametric (e.g., Kolmogorov–Smirnov) tests return probabilities larger than 50%. Therefore, it is not unreasonable to assume that the c values follow a normal distribution (also shown in Fig. 4) with a weighted standard deviation of $s_{\bar{c}} = 0.228$. Taking this observation into account, we can check the $c = 0$ hypothesis for our $n = 48$ seismic groups with the t-test. We find that

$$t = \frac{\bar{c} - c_0}{s_{\bar{c}}/\sqrt{n}} = 10.97, \tag{7}$$

which rejects the $c = 0$ hypothesis for the 99% confidence limit ($t_{0.995} = 2.576$).

Calculation of a Global c Value

Based on the results of the previous section, we can test the applicability of equation (4) under the assumption that the slope, $\bar{c}$, is the same for all groups. Therefore, we want to fit the data of all seismic groups using a common slope. For this reason, equation (1) was modified to the following relation:

$$\log T_i = c\, Mp_i + \sum_{j=1}^{n} \delta_{ij}\, a_j = c\, Mp_i + a_i, \tag{8}$$

where the T_i and M_{p_i} is a time interval–magnitude pair that belongs to seismic group i, c is the common slope for all seismic groups, a_j is the constant that corresponds to each seismic group j, and n is the number of the available groups of data. It is clear that using equation (8) we can treat all data as a unified data set. Each datum, T_i, is now controlled by two different factors, M_{p_i} and the seismic group where it belongs, i. If we write equation (8) for all the m available data, after sorting them for their seismic group number, we form the following linear system:

$$\begin{bmatrix} \log T_1 \\ \log T_2 \\ \vdots \\ \log T_k \\ \vdots \\ \log T_m \end{bmatrix} = \begin{bmatrix} Mp_1 & 1 & 0 & \cdots & 0 \\ Mp_2 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Mp_k & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Mp_m & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} c \\ a_1 \\ a_2 \\ \vdots \\ \vdots \\ a_n \end{bmatrix} \tag{9}$$

or using a matrix notation:

$$\mathbf{T} = \mathbf{A}\,\mathbf{a}. \tag{10}$$

In our case, the number of available data was $m = 263$ and the number of parameters to be determined was $n + 1 = 49$. Therefore, we used the least-squares solution of equation (10):

$$\mathbf{a}_{\mathrm{LSQ}} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\,\mathbf{A}^{\mathrm{T}}\mathbf{T}. \tag{11}$$

In our case, the least-squares solution was determined using singular value decomposition; therefore, $\mathbf{a}_{\mathrm{LSQ}}$ is given by

$$\mathbf{a}_{\mathrm{LSQ}} = \mathbf{V}\,\mathbf{\Lambda}^{-1}\mathbf{U}^{\mathrm{T}}\mathbf{T}. \tag{12}$$

For the available data set, matrix $\mathbf{A}$ had a condition number smaller than 10^3 (=953), and the smallest eigenvalue was around 0.2. Therefore, the least-squares solution is quite adequate and stable, and no additional damping or smoothing constraints were used. The slope c of the final solution was found equal to 0.376, which is almost identical to the weighted average value determined from the c values of each seismic group in the previous section (=0.361) and very close to the value calculated from Papazachos *et al.* (1995) (=0.33). The constants a_j of equation (8) computed for each seismic group are listed in Table 1. The generalized linear correlation coefficient was equal to 0.71. The overall misfit of equation (10) to the data was $s_{\log T} = 0.217$. The corresponding value when different constants, c, are used for each seismic group, which is equivalent to the application of equation (3) on each seismic group separately, is $s_{\log T} = 0.178$. These values are very close if we consider that $n - 1 = 47$ additional parameters were used in order to obtain this additional misfit reduction.

If we want to test the statistical significance of the fit of equation (10) on the data, we can perform the χ^2 test on the quantity

$$\chi^2 = (\mathbf{A}\,\mathbf{a} - \mathbf{T})^{\mathrm{T}}\mathbf{C}_{\mathbf{T}}^{-1}(\mathbf{A}\,\mathbf{a} - \mathbf{T}). \tag{13}$$

In the absence of any specific information about the *a priori* data errors, we assumed a diagonal covariance matrix for the data, $\mathbf{C}_{\mathbf{T}} = \sigma^2_{\log T}\,\mathbf{I}$, where $\sigma_{\log T}$ is the typical error for the logarithm of the interval times. The main problem is that the typical error in $\log \mathbf{T}$ is not a result of measurement, since there are no direct errors involved in the definition of interval times. Therefore, $\sigma_{\log T}$ here expresses the different sources of deviation from equation (10), which are either errors associated with M_p (as indicated earlier) or intrinsic errors associated with the applicability of equation (10) (possible limited variability of c in different areas, limitations of the model, etc.). Previous estimates (Nishenko and Buland, 1987; Papazachos *et al.*, 1995) suggest for $\sigma_{\log T}$ values in the range (0.2 to 0.3). Because of the intolerance of the normal distribution to large outliers, this range of $\sigma_{\log T}$ values returns probability significance levels for the χ^2 test that vary from 4% to 100%. For this reason, we decided to disregard the χ^2 test and adopt the $\sigma_{\log T} = s_{\log T} = 0.217$ value for all further calculations. In any case, the calculated value is quite close to the lower bound of the proposed $\sigma_{\log T}$ range; therefore, the fit is essentially acceptable.

In order to compute confidence limits for the estimated parameters, we calculated the *a posteriori* covariance matrix of the solution, which is given by

$$\mathbf{C}_{\mathbf{a}} = \sigma^2_{\log T}\,\mathbf{V}\,\mathbf{\Lambda}^{-2}\mathbf{V}^{\mathbf{T}}. \tag{14}$$

Using the previous value for $\sigma_{\log T}$, we computed the covariance matrix for the solution and the corresponding 95% ($\pm 2\sigma$) confidence limits for each parameter. The final error $\sigma_j = \sqrt{\mathbf{C}_{(\mathbf{j}+1)(\mathbf{j}+1)}}$ for each parameter a_j is also listed in Table 1. For the slope, c, the estimated standard error was very small ($\sigma_c = 0.035$), since this slope is now determined by the total number of available data. The 95% confidence interval for c is (0.31, 0.44); therefore, it is quite clear that we can reject the $c \leqq 0$ hypothesis at any reasonable significance level.

However, the previously calculated errors are only conditional; that is, they are valid for one parameter only if all the others take the value computed by the least-squares regression. If we want to compute the joint confidence limits for all the parameters, we should use the $(n + 1)$-dimensional ellipsoid of confidence:

$$\Delta\chi^2 = \mathbf{da}^{\mathrm{T}}\mathbf{C}_{\mathbf{a}}^{-1}\mathbf{da}, \tag{15}$$

where $\Delta\chi^2$ is the allowed increase of χ^2 for $n + 1$ degrees of freedom for a given confidence limit and **da** is a perturbation of the **a** vector. In our case, for $n + 1 = 49$ parameters (degrees of freedom), the 95% confidence ellipsoid has

$\Delta\chi^2 \cong 66.34$. Hence, we would like to test if we have any solutions within this ellipsoid that have $c = 0$. For this reason, we computed the solution of equation (15) that has the minimum c value. This translates to

$$\frac{\partial c}{\partial a_j} = 0, \quad j = 1, \ldots, n, \tag{16}$$

which, in combination with equation (15), results in

$$(\mathbf{C_a}^{-1})_i \, \mathbf{da} = 0, \quad i = 2, \ldots, n + 1 \tag{17}$$

or

$$\sum_{j=2}^{n+1} (\mathbf{C_a}^{-1})_{ij} \, \mathbf{da}_j = -(\mathbf{C_a}^{-1})_{i1} \, dc, \quad i = 2, \ldots, n + 1, \tag{18}$$

where $(\mathbf{C_a}^{-1})_\mathbf{I}$ is the ith row of the inverse covariance matrix and dc is the slope perturbation. Equation (18) is an $(n \times n)$ linear system, which can be solved for $\mathbf{da}_j$ $(j = 2, \ldots, n + 1)$, given dc. In our case, since we are interested in the $c = 0$ solution, we have $dc = -c = -0.376$. After solving equation (18), we computed $\Delta\chi^2$ using equation (15) and found a $\Delta\chi^2 \cong 116.2 > 66.34$. Therefore, there are no acceptable solutions in the 95% confidence ellipsoid with $c = 0$, nor with $c < 0$. On the other hand, it would be interesting to find this minimum c value for a given confidence limit. Substituting equation (17) in equation (15) results in

$$(\mathbf{C_a}^{-1})_1 \, \mathbf{da} = \Delta\chi^2/dc. \tag{19}$$

Equation (19) in combination with equation (17) forms a $(n + 1) \times (n + 1)$ linear system, which can be solved iteratively, using an initial approximate value for dc. In our case, the solution with the minimum c value, c_{low}, calculated for the 95% ellipsoid was approximately equal to 0.1. This conclusion is useful because it sets an absolute lower bound, c_{low}, on the possible values of c. If we would like to compute the probability of solutions with $c \leqq c_{\text{crit}}$ $(> c_{\text{low}})$, where c_{crit} is a critical c value in which we are interested, we need to calculate the ratio of the volume bounded by the $(n + 1)$-dimensional ellipsoid and the generalized plane $c = c_{\text{crit}}$ to the total volume of the ellipsoid. Numerical computation of this integral (e.g., using a Monte Carlo technique) is very inefficient in our case, since we have to sample a 49-dimensional space. However, we can take advantage of some special symmetry of our problem and approximately compute this probability. Therefore, for the adopted standard error, $\sigma_{\log T} = 0.217$, and the calculated c value $(= 0.376)$, we find that 99.95% of the possible solutions within the 95% confidence ellipsoid have $c > 0.25$. If we choose a different *a priori* error for $\sigma_{\log T}$, e.g., 0.3, we find slightly smaller probabilities ($\cong$99%). Therefore, the probability of even small positive solutions for c (< 0.25) within the 95% confidence ellipsoid must be rejected for any reasonable probability level.

We also performed a similar analysis for the corresponding magnitude predictable equation:

$$M_{fi} = CM_{pi} + \sum_{i=1}^{n} \delta_{ij} A_j, \tag{20}$$

which connects the magnitude of each earthquake with the magnitude of the preceding event. The slope C was found equal to -0.30, which indicates that large events tend to be followed by small events and vice versa. This value is also very close to the value $(= -0.28)$ calculated by Papazachos *et al.* (1995). An appropriate model for this behavior observed in various seismic regions has been proposed by Scholz (1990). The overall standard error was $s_{M_a} = 0.387$, and the generalized linear correlation coefficient was 0.818. The constants A_j $(j = 1, \ldots, n)$ and their corresponding errors are presented in Table 1. If we use the calculated standard error for M_f, we find that the corresponding 95% confidence interval for C is $(-0.177, -0.425)$. From equation (18), we compute $\Delta\chi^2 = 23.42 < 66.34$; therefore, some solutions with $C \geqq 0$ exist within the 95% confidence ellipsoid. The maximum non-negative C value computed from equation (19) is 0.205. On the other hand, the percentage of solutions within the 95% confidence error ellipsoid, having $C \geqq -0.15$, is only 1.61%, which drops to 0.2% for $C \geqq -0.1$. Therefore, although the errors are larger, the alternation of large and small events (imposed by $C < 0$) is quite robustly demonstrated by the data. Of course, this does not exclude the regions where "characteristic" events occur, as can be seen by setting $M_f = M_p$ in equation (20). Then for each seismic group, the magnitude of the characteristic event is equal to $M_{\text{char}} = A_j/(1 - C)$. Events larger than M_{char} will be followed by events smaller than M_{char} and vice versa, whereas events close to M_{char} will tend to be followed by events of the same magnitude.

We see that both the errors and the confidence intervals are larger for relation (20) than relation (8). This difference is also observed in the partial correlation coefficients, R_c and R_C, equal to 0.77 and 0.47, respectively, which give a good idea about the robustness of the physical laws presented through equations (8) and (20). The larger R_c value indicates that prediction of interevent times is more robust than the corresponding magnitude estimation for the following event.

Equation (20) should not be confused with a slip-predictable relation. Of course, a combination of equation (20) with equation (8) would result in an equation that resembles the slip-predictable equation:

$$\log T_i = EMa_i + \sum_{i=1}^{n} \delta_{ij} F_j. \tag{21}$$

We can process the data in a similar manner and calculate a slope E as well as the constants F_j. The important point of the slip-predictable model is that constant E is positive; hence large interevent times lead to large events. However, application of equation (21) on the data yields a neg-

ative value (-0.249), which could also be predicted by the combination of equations (8) and (20). The probability of a positive ($E > 0$) value within the 95% confidence ellipsoid is essentially null; therefore, the slip-predictable model (large events follow large interevent times) is rejected by the data for any reasonable confidence limit.

Model Dependence on Zonation

A critical point for the previous computations is the problem of seismic zonation. The seismogenic regions previously used were determined on the basis of various criteria, which might be considered subjective. Since the problem of seismic zonation is much broader than the problem studied here, we chose an indirect way to test the robustness of our results by using a different data set from Italy, recently published by Mulargia and Gasperini (1995). The zonation presented in their article was based on a different technique and information than that used in the present article. From this set, we used data with a magnitude range larger than 0.6, but we accepted seismic groups with $n \geqq 3$ number of observations (interevent times), since the available amount of data was limited. Moreover, we considered that data with $M \geqq M_{\min}$ were complete since the year when the $M_{\min}$ appeared for the first time in the Mulargia and Casperini catalog. For instance, for the Naso region, the data are considered complete since 1613 for $M \geqq 5.9$ and since 1783 for $M \geqq 5.6$ because these two magnitudes appeared for the first time in the corresponding years. The data for this region were not used since their magnitude range was $\Delta M = 5.9 - 5.6 = 0.3$.

The final data set consisted of eight seismogenic regions (or 12 seismic groups) listed in Table 2. Application of equation (4) gave a c_{Italy} value equal to 0.346, which is very close to the global value of 0.376. The equation linear correlation coefficient is 0.73, and the misfit is $s_{\log T} = 0.334$, which is larger than the global misfit, as expected for the smaller data set used and since we included seismic groups with three and four observations. Moreover, the probability that $c_{\text{Italy}} > 0.15$ is found to be approximately 98%. As far as the magnitude predictable model is concerned, we found $C_{\text{Italy}} = -0.436$, which is larger than the global value. However, the misfit is equal to 0.306, and linear correlation coefficient is 0.919, which are better than the corresponding values for the global model. On the other hand, this is not as important for equation (4), since the error induced to log T from magnitude errors is much smaller ($= c \cdot \sigma_{M_s} \cong 0.1$ to 0.15). All the calculated parameters are shown in Table 2. For the magnitude-predictable model, 99.9% of the solutions in the 95% confidence ellipsoid have $C < -0.2$. Using a similar procedure, we can reject the slip-predictable model for Italy for any reasonable confidence limit. Therefore, we may conclude that these additional results from the application of the regional time- and magnitude-predictable model on a different data set demonstrate the robustness of the model and of the processing procedure.

Conclusions

In the present work, we have studied in detail the statistical significance of the regional time- and magnitude-predictable model. We focused on the crucial points of this model, which is the linear dependence of the logarithm of interevent times and of the magnitude of the following earthquake on the magnitude of the preceding mainshocks for each pair of events belonging to a seismogenic region. Simulation of the errors involved in the calculations show that we cannot rely on statistical tests for data from a single region, due mainly to errors in the magnitude estimation and the small magnitude range spanned by the data. Moreover, these tests do not take into account the global character of the estimated parameters.

Table 2

Same as Table 1 for Italian seismogenic regions, except for second column where the year since when the data are considered complete is presented. Data are taken from Mulargia and Gasperini (1995).

Source Parameters				Time-Predictable Model for Each Source				Source Parameters for Italian Time- and Magnitude-Predictable Models				
$M_{\min}$	Year of Compl.	ΔM_p	n	c	r	T_c	Q	a_j	σ_{aj}	A_i	σ_{Aj}	Region
5.9	1626	1.2	10	0.47	0.60	1.68	0.154	−0.79	0.90	9.09	0.82	Calabria
6.5	1626	0.6	5	0.20	0.29	0.60	0.583	−0.64	0.95	9.52	0.87	Calabria
5.9	1694	0.7	6	0.70	0.73	1.87	0.158	−0.62	0.89	8.91	0.82	Irpinia—Basilicata
6.0	1694	0.6	5	0.43	0.36	0.77	0.484	−0.60	0.91	8.99	0.83	Irpinia—Basilicata
5.9	1741	0.6	5	0.44	0.28	0.51	0.647	−0.80	0.87	8.84	0.80	Umbro—Marchigniano
5.6	1767	0.9	5	1.14	0.94	4.73	0.018	−0.75	0.85	8.72	0.78	Region 29
5.9	1730	0.6	5	−0.03	−0.02	−0.04	0.973	−0.80	0.87	8.84	0.80	Region 29
5.9	1835	0.6	4	0.34	0.65	1.71	0.162	−0.64	0.89	8.76	0.82	Region 44
5.9	1626	0.6	4	0.62	0.91	3.84	0.031	−0.36	0.91	9.13	0.84	Region 46
4.5	1908	0.9	6	0.13	0.38	0.71	0.526	−0.76	0.70	6.97	0.64	Region 37
4.8	1908	0.6	3	0.63	0.89	3.30	0.046	−0.83	0.75	7.23	0.68	Region 37
4.2	1968	0.7	4	0.09	0.41	0.79	0.487	−0.91	0.68	6.62	0.62	Region 20

It is shown that we can use a different formulation that demonstrates the global applicability of the regional time- and magnitude-predictable model. Using the same formulation, we can reject the regional slip-predictable model. On the other hand, it is demonstrated that these results are independent to the specific seismic zonation technique that is employed, as is shown from an application to a different data set.

The results presented in this work should not be interpreted as an accurate short-term prediction method. The main reason is the intrinsic errors of the corresponding regional time- and magnitude-predictable models. Moreover, these models should be applied only in connection with the declustering procedure, which is used to scale the energy released in a seismic cycle to a single event. Therefore, it poses certain limitations to the recurrence models, especially the time-predictable model, since it cannot be applied directly after a large event, as long as we have a significant postshock activity. Therefore, equations (4) and (20), or the corresponding equations (1) and (2), should be exclusively considered as long-term prediction relations. However, we might use the deviations from these relations or incorporate additional factors that account for other effects such as triggering of events from neighboring seismic regions, variability of the model parameters, etc. in order to improve the accuracy of the examined recurrence models.

Acknowledgments

The authors are grateful to Prof. B.C. Papazachos for his creative suggestions and his advice during the implementation of this work. This work has been financially supported by the EEC Environment Research Project EV5V-CT94-0513.

References

Bufe, C. G., P. W. Harsch, and R. O. Burford (1977). Steady-state seismic slip: a precise recurrence model, *Geophys. Res. Lett.* **4,** 91–94.

Davis, P. M., D. D. Jackson, and Y. Y. Kagan (1989). The longer it has been since the last earthquake, the longer the expected time till the next?, *Bull. Seism. Soc. Am.* **79,** 1439–1456.

Jacob, K. H. (1984). Estimates of long-term probabilities for future great earthquakes in the Aleutians, *Geophys. Res. Lett.* **11,** 295–298.

Kagan, Y. Y. and D. D. Jackson (1991). Long-term earthquake clustering, *Geophys. J. Int.* **104,** 117–133.

Karakaisis, G. F. (1993). Long term earthquake prediction in the New Guinea–Bismarck Sea region based on the time and magnitude predictable model, *J. Phys. Earth* **41,** 365–389.

Karakaisis, G. F. (1994a). Long term earthquake prediction in Iran based on the time and magnitude predictable model, *Phys. Earth Planet. Interiors* **83,** 129–145.

Karakaisis, G. F. (1994b). Long-term earthquake prediction along the North and East Anatolian fault zones based on the time- and magnitude-predictable model, *Geophys. J. Int.* **116,** 198–204.

Kiremidjan, A. S. and T. Anagnos (1984). Stochastic slip-predictable model for earthquake occurrence, *Bull. Seism. Soc. Am.* **74,** 739–755.

Mulargia, F. and P. Gasperini (1995). Evaluation of the applicability of the time- and slip-predictable earthquake recurrence models to Italian seismicity, *Geophys. J. Int.* **120,** 453–473.

Nishenko, S. P. and R. Buland (1987). A generic recurrence interval distribution for earthquake forecasting, *Bull. Seism. Soc. Am.* **77,** 1382–1399.

Panagiotopoulos, D. G. (1993). Long term earthquake prediction in the Philippines region based on the time and magnitude predictable model, *Proc. of the 2nd Congr. Hellenic Geophysical Union,* Florina, Greece, 5–8 May 1993, 482–491.

Panagiotopoulos, D. G. (1995a). Long-term earthquake prediction in central America and Carribean sea based on the time- and magnitude-predictable model, *Bull. Seism. Soc. Am.* **85,** 1190–1201.

Panagiotopoulos, D. G. (1995b). Long term earthquake prediction along the seismic zone of Solomon Islands and new Hebrides based on the time and magnitude predictable model, *Natural Hazards* **11,** 17–43.

Papadimitriou, E. E. (1993). Long term earthquake prediction along the western coasts of south and central America based on a time predictable model, *Pure Appl. Geophys.* **140,** 301–316.

Papadimitriou, E. E. (1994a). Long term prediction in North Pacific seismic zone based on the time and magnitude predictable model, *Natural Hazards* **9,** 303–321.

Papadimitriou, E. E. (1994b). Long term prediction of large shallow mainshocks along the Tonga–Kermadec–New Zealand seismic zone based on the time and magnitude predictable model, *Tectonophysics* **235,** 347–360.

Papazachos, B. C. (1989). A time-predictable model for earthquakes in Greece, *Bull. Seism. Soc. Am.* **79,** 77–84.

Papazachos, B. C. (1992). A time and magnitude predictable model for generation of shallow earthquakes in the Aegean area, *Pure Appl. Geophys.* **138,** 287–308.

Papazachos, B. C. and Ch. A. Papaioannou (1993). Long-term earthquake prediction in the Aegean area based on a time and magnitude predictable model, *Pageoph* **140,** 593–612.

Papazachos, B. C., E. E. Papadimitriou, G. F. Karakaisis, and T. M. Tsapanos (1994). An application of the time and magnitude predictable model for the long term prediction of strong shallow earthquakes in the Japan area, *Bull. Seism. Soc. Am.* **84,** 426–437.

Papazachos, B. C., E. E. Papadimitriou, G. F. Karakaisis, and D. G. Panagiotopoulos (1995). Long term earthquake prediction in the Circum–Pacific convergent belt, *Pure Appl. Geophys.* **145** (in press).

Papazachos, B. C., G. F. Karakaisis, E. E. Papadimitriou, and Ch. A. Papaioannou (1996). Time dependent seismicity in the Alpine–Himalayan belt, *Tectonophysics* (in press).

Scholz, C. H. (1990). *The Mechanics of Earthquakes and Faulting,* Cambridge Univ. Press, New York, 439 pp.

Shimazaki, K. and T. Nakata (1980). Time-predictable recurrence model for large earthquakes, *Geophys. Res. Lett.* **7,** 279–282.

Sykes, L. R. (1983). Predicting great earthquakes, in *Earthquakes: Observation, Theory and Interpretation,* H. Kanamori and E. Boschi (Editors), North-Holland Publishing Co., Amsterdam, 398–435.

Sykes, L. R. and R. C. Quittmeyer (1981). Repeat times of great earthquakes along simple plate boundaries, in *Earthquake Prediction, An International Review, Maurice Ewing Series 4,* D. W. Simpson and P. G. Richards (Editors), American Geophysical Union, Washington, D.C., 217–247.

Institute of Engineering Seismology and Earthquake Engineering
P.O. Box 53, Foinikas
Thessaloniki GR-55102, Greece
(C.B.P.)

Geophysical Laboratory
University of Thessaloniki
P.O. Box 352-1
Thessaloniki GR-54006, Greece
(E.E.P.)

Manuscript received 26 February 1996.