БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ*BULGARIAN ACADEMY OF SCIENCES

БЪЛГАРСКО ГЕОФИЗИЧНО СПИСАНИЕ*BULGARIAN GEOPHYSICAL JOURNAL
т. XVIII, № 1 Vol. XVIII, No 1
*София * 1992 * Sofia*

Statistical investigations of the seismic process along the San Andreas fault

B. Ranguelov, E. Rakova

Geophysical Institute, BAS, Sofia

Introduction

The ideas about the discrete seismic phenomena and processes and the attempts for their quantitative study during the last few years [1, 2] have established themselves as a modern tendency with increasing means and new methods for its investigation. On the other hand, the San Andreas fault is of a high seismic activity and has a perfect network for recording of the seismic events along it. These are two main motives that help the choice to select this region for a more detailed statistical investigation and establishment of quantitative relations which can help the better understanding of the discrete process.

Of course, the suggested investigation has no claims to be first of its kind in this well-known region. We are familiar with lots of publications concerning studies both of the background seismicity [3, 4], as well as the manifestations of strong destructive events [5].

The main purpose of this investigtion is the attempt to describe analytically the seismic process in time and space on a limited segment of the San Andreas fault. We have available relatively reliable data [6] considering the statistical distributions of the time intervals ($\Delta t = t_{i+1} - t_i$) and the distances (Δx) between each two consecutive earthquakes which we consider as the most representative characteristics of the discrete seismic process [12, 13, 14].

Data

Both ends of the San Andreas fault segment have coordinates: 120°18′W, 35°45′N, 121°°00′W, 36°25′N and the total length is about 100 km. The epicentres which are considered to be connected to the fault are spread in a band wide about 10 km. The data used are taken from a catalogue [6] which contains more than 2000 earthquakes within a magnitude interval 1.5-4.87 and for the period 1969-1982, i. e. a 14 year period. The processing covers 1883 events divided into three magnitude intervals: (2.0-2.5) —451, (2.5-3.0)—126, (3.0—3.9)—70. It was necessary to check the eventual differen-

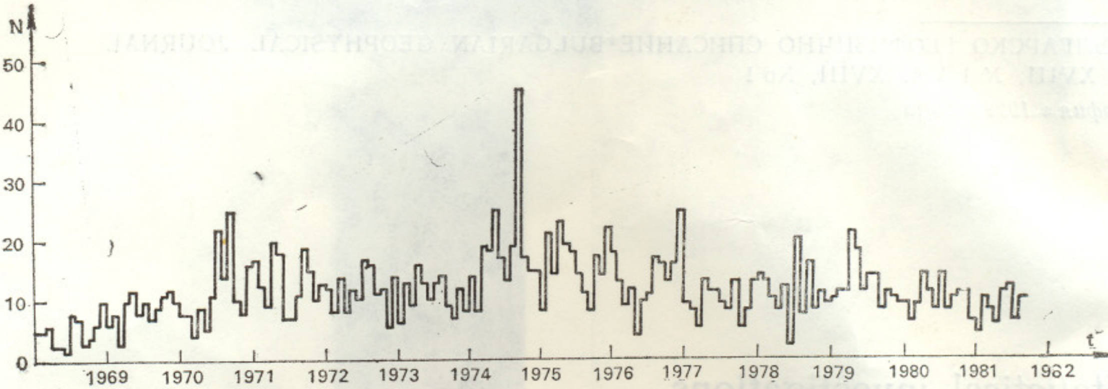

Fig. 1. Time variations of the monthly number of earthquakes $N(t)$ for the San Andreas fault region

ce in the occurrence of stronger and smaller events and this was the reason for such a division.

The accuracy in determining the parameters of the earthquakes is as follows:

quality	RMS [s]	ERH [km]	ERZ [km]
A	0.15	1.0	2.0
B	0.30	2.5	5.0

RMS — mean square error of time differences; ERH — standard error for the epicentre in [km]; ERZ — standard error for depth in [km].

The depths of the earthquakes are down to 12-15 km [6].

For the establishment of a statistical independence between the pairs of events under investigation, a preliminary filtration of aftershock events is applied using the Knopoff's space-time window [11]. It leads to a reduction of the number of events in the different magnitude intervals: (2.0-2.5) — 330, (2.5-3.0) — 93, (3.0-3.9)—54.

The preliminary analysis of initial data shows that the period of time under consideration is characterized by the absence of strong destructive earthquakes and as a whole can be evaluated as a period of background seismicity (Fig. 1). An exception is the month of September 1975 when the number of recorded events per month reaches 45, when the average number for the whole period is about 11-12. The careful consideration of the data shows that the anomaly is random. During this month no strong earthquake has occurred and if considered every 15 days with overlapping of the months, this anomalous value reaches acceptable boundaries of about 20 events. In Fig. 2 is shown a recurrence graph for the whole period under consideration which has parameters obtained by the method of the least squares:

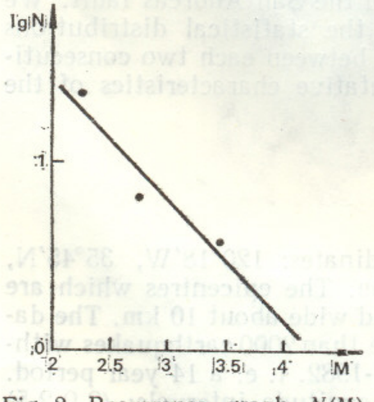

Fig. 2. Recurrence graph $N(M)$

$$\lg N = 1.44 - 0.64.M;$$

$$2.0 \leqslant M \leqslant 3.9.$$

Methods of investigation

In accordance with the main goal the following methods of data processing have been used. By means of a PC-program the distances between the epicentres and the differences in time of occurrence, as well as the azimuths of rays connecting each pair of consecutive events are calculated. The accuracy does not surpass the initial accuracy of the estimated parameters.

By the data obtained from the original processing, the corresponding histograms of distributions in distance and time differences, in km and days respectively, are constructed. The values of the relative frequencies p_i^* are calculated from $p_i^*=N_i/N$, where N_i is the number of values in the i-interval of grouping, and N is the total number.

Table 1

Values for establishment of the optimal histogram interval

Interval M	N	Δx_{max} [km]	Δx_{min} [km]	δx [km]	Δt_{max} [days]	Δt_{min} [days]	δt [days]	$\widetilde{\Delta x}$ [km]	$\widetilde{\Delta t}$ [days]
2.0-2.5	329	100	5	10.2	77	3h	8.2	10	10
2.5-3.0	92	72	5	8.9	286	2	37.8	10	30
3.0-3.9	52	70	8	9.2	277	2	40.8	10	60

The statistical characteristics are calculated too: $\bar{x}$ — mean value, S^2 — dispersion, S — mean square deviation, $V=S/x$ — coefficient of variation, A — asymmetry, E — excess.

The choice of histogram interval has been made by Sturges' estimation [7] which has the form $\Delta x=(\Delta x_{max}-\Delta x_{min})/(1+3.321.\lg N)$. The results are presented in Table 1. The accepted histogram intervals $\widetilde{\Delta x}$, $\widetilde{\Delta t}$ are selected taking into account also the convenience values.

Special attention is paid to the approximation of the obtained empirical distributions with analytical ones. Because of the strongly asymmetric character of the obtained histograms and their evidently big excess, analytical distributions of Poisson, Gauss, binomial, etc. were tested. It was seen that the best approximation could be obtained from the β-distribution with respect to the intervals by distances and exponential distribution in relation to time intervals. This is a fact established also by other researchers [8].

For approximation of the distributions by distance intervals the method of nonlinear programming has been used. The simplex method has been used for the linearization of the optimization procedure. The latter is realized through a program — Program SIMPLEX in FORTRAN is applied [9].

The general form of β-distribution is:

$$f(x)=\Gamma(p)\,.\,\Gamma(q)/\Gamma(p+q)\,.\,x^{p-1}(1-x)^{q-1},$$

where

$$\Gamma(p)=\int_0^\infty e^{-t}\,.\,t^{p-1}dt$$

is gamma-function [10]. The parameters of the distribution p, q are $p>0$, $q>0$, and the variable x—$0\leqslant x\leqslant 1$. As the coefficient containing a gamma-function is a normali-

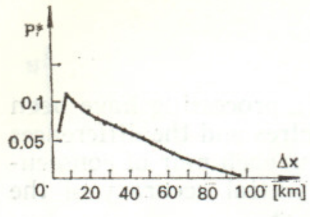

Fig. 3. Distances distribution for 100 000 pairs of events, generated randomly on a rectangular area with dimensions 100×10 km

zed multiplier, i. e. such that $\int_{-\infty}^{\infty} f(x) \, . \, dx = 1$, then the distribution function achieves the form:

$$f(x) = c \, . \, x^{p-1}(1-x)^{q-1}.$$

For the most precise description of the empiric distribution by an analytical curve, the following criteria must be followed:

$$F = \sum_{i=1}^{k} [f(\bar{x}_i) - p_i^*]^2 = \min,$$

$\bar{x}_i$ is the mean of the i-th interval from the grouping in the histogram; p_i^* is the relative frequency attributed to the same point; $f(\bar{x}_i)$ is the value of the selected analytical function. Of course, a normalization with the variable x is also required.

An analogous procedure has been also used in the approximation of the histograms for time intervals, using for the analytical function the exponent of the type:

$$f(t) = c \, . \, \theta \, . \, exp(-\theta \, . \, t),$$

The coefficient c is as follows $\int_{-\infty}^{\infty} f(t) \, . \, dt = 1.$

After determining the parameters of the corresponding analytical distributions describing the best fitting of the empiric histograms a statistical check up is made. The empiric distribution is best fitted with the necessary reliability by the obtained theoretical function. For this purpose the χ^2-criterion has been applied at a level of reliability 95% (which corresponds to 5% level of significance of the deviation between the empirical and theoretical distributions).

To avoid the boundary effects caused by the zone geometry and to estimate their influence, the following numerical experiment has been performed. On a rectangular area with dimensions 100×10 km along which the points are randomly distributed, pairs of events are generated and the distances between them are calculated. By means of the values thus obtained the expected distribution of the distances generated by chance is found, which is presented in Fig. 3.

Results

The results obtained from the processing of initial data are presented in Fig. 4, *a*, *b*, *c* | in distance, and Fig. 5*a*, *b*, *c* — in time, and the values from the calculations of the statistics and parameters of distribution are shown in the following table (2).

The rose-diagrams from Fig. 6 *a*, *b*, *c* give the azimuthal distribution of the rays connecting the pairs of successive events.

The data analysis shows comparatively high values of the variation coefficient. It also distinctly corroborates the high values of asymmetry and excess for almost all distributions. All distance distributions are of unimodal character and the curves according to time are similar exponents.

able 2

Statistical parameters of the empirical and theoretical distributions

Interval M	By distance								By time						
	p	q	c	$\bar{x}$	S_x	V_x	A_x	E_x	θ	c	$\tilde{t}$	S_t	V_t	A_t	E_t
2.0-2.5	1.46	3.56	0.7	31	20	0.6	0.7	0.02	0.065	10.2	15	14	0.9	2	4
2.5-3.0	1.44	2.87	0.7	26	16	0.6	0.5	0.36	0.017	31.6	52	42	0.8	1	2
3.0-3.9	1.40	1.71	0.4	32	18	0.6	0.3	—0.9	0.010	64.0	86	70	0.8	1	0.1

The analysis of these curves gives grounds for the following affirmations:

The predominant distances between the pairs of consecutive events are within the 10-30 km interval, this distance is practically not depending on the earthquake magnitude. The analytical curves which describe the statistical distributions sufficiently well have parameters varying within the boundaries: $p=1.4—1.46$; $q=1.7—3.56$; $c=0.38—0.7$

The time intervals between the consecutive earthquakes decrease exponentially with time, reaching up to 300 days for the pairs with bigger magnitudes. The relative

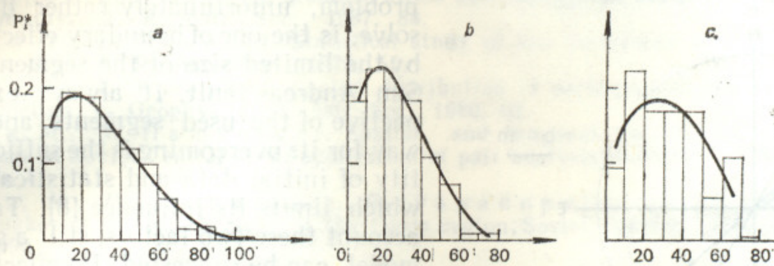

Fig. 4. Histograms and analytical curves of the distributions of distances between the consecutive earthquakes in the region of San Andreas for the selected magnitude intervals as follows:
a) M ∈ [2.0, 2.5) $f(x)=0.7x^{0.46}(1-x)^{2.56}$; b) M ∈ [2.5, 3.0) $f(x)=0.7x^{0.44}(1-x)^{1.87}$; c) M ∈ [3.0, 3.9) $f(x)=0.4x^{0.40}(1-x)^{0.71}$

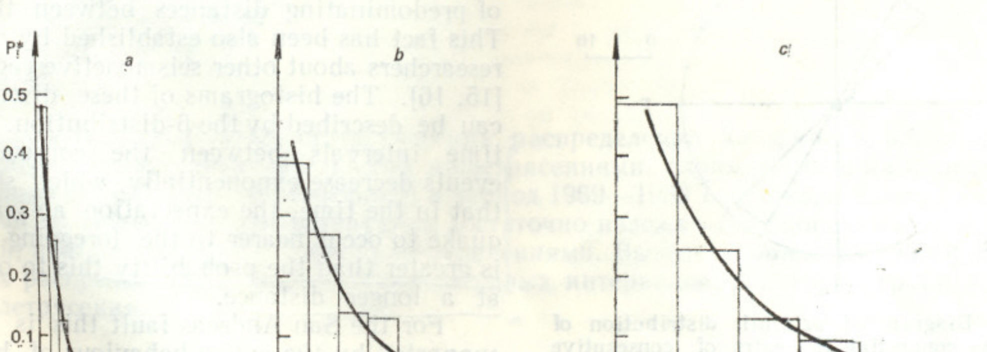

Fig. 5. Histograms and analytical curves of the time intervals distributions between the consecutive earthquakes for the following magnitude intervals:
a) M ∈ [2.0, 2.5) $f(t)=0.66.e^{-0.065t}$; b) M ∈ [2.5, 3.0) $f(t)=0.54.e^{-0017.t}$; c) M ∈ [3.0, 3.9) $f(t)=0.64.e^{-0.010t}$

share of the bigger time intervals decreases with the decrease of earthquake magnitud and reaches up to 60-80 days for the magnitude interval (2.0-2.5). The relative shar of the least time interval (up to 10 days) is over 50% This is clearly seen from the chan ge of the coefficints' values which are for (from 0.01 to 0.65 and for *c* from 64 to 110

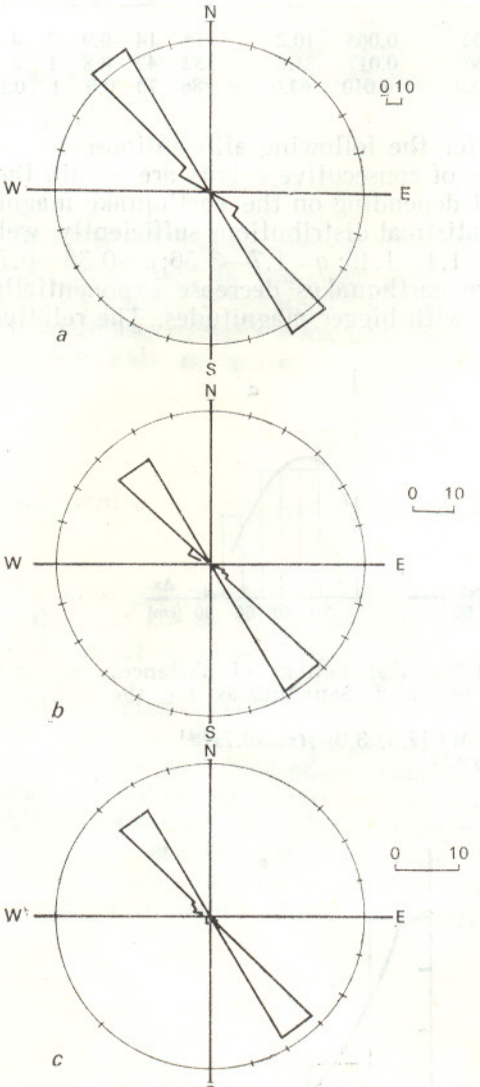

Fig. 6. Diagram of azimuth distribution of the rays connecting the pairs of consecutive events for:
a) $M \in [2.0,\ 2.5)$; b) $M \in [2.5,\ 3.0)$;
c) $M \in [3.0,\ 3.9)$

The biggest part of the direction o the rays connecting the consecutive pairs o earthquakes is within the interval 30-45°NW i. e. they are situated almost parallell to the fault line (40.9°NW).

Some unsolved problems in relatior with these investigatons refer to the accurac evaluation of the obtained parameters o the approximating distributions. They ar a special subject for future investigations — it turned out that at the non-linear appro ximations the solution of this problem i made somewhat difficult. Another objective problem, unfortunately rather difficult t solve, is the one of boundary effects imposed by the limited size of the segment from th San Andreas fault. It always exists irres pective of the used segments and the onl way for its overcoming is the sufficient quan tity of initial data and statistical approach which limits its influence [8]. Taking int account these two factors, still a generalized model can be suggested. It reflects the rea conditions for the seismic events realiza tions.

The proposed real physical model based on these data supports the thesis that th consecutive events in time are located in spa ce in a manner which shows the availabilit of predominating distances between them This fact has been also established by othe researchers about other seismoactive region [15, 16]. The histograms of these distance can be described by the β-distribution. Th time intervals between the consecutiv events decrease exponentially, which show that in the time, the expectation an earth quake to occur nearer to the foregoing on is greater than the probability this to occu at a longer distance.

For the San Andreas fault this is als supported by the entire behaviour of back ground seismicity as a discrete, but relative ly constant in time flow of events.

References

1. Садовский, М. А., Л. Г. Болховитинов, В. Ф. Писаренко. Деформирование геофизической среды и сейсмический процесс. М., Наука, 1987.
2. Дискретные свойства геофизической среды. Сборник научных трудов, М., Наука, 1989.
3. Eneva, M., G. Pavlis. Application of pair analysis statistics to aftershocks of the 1984 Morgan Hill California earthquake. — J. Geophys. Res., vol. 93, 1988, No B8.
4. Eneva, M., G. Pavlis. Spatial distribution of aftershocks and background seismicity in Central California. — PAGEOPH, 1990.
5. Chi-Yu King, Z. Ma. Migration of historical earthquake in California. — PAGEOPH, vol. 127, 1988, No 4.
6. Reasenberg, P. USGS Central California Microearthquake Catalog. USGS, Menlo Park, 1982 (Personal Communication).
7. Львовский, Е. Н. Статистические методы построения эмпирических формул. М., 1982.
8. Енева, М. Оценка на пространствено-временните и енергетични характеристики на земетресенията за целите на прогностичното райониране. Дисертация. 1984.
9. Химмельблау, В. Прикладное нелинейное программирование. М., Мир, 1975.
10. Мосталер, Ф., Дж. Тьюки. Анализ данных и регрессия. М., Финансы и статистика, 1982.
11. Gardner, J. K., L. Knopoff. Is the sequence of earthquake in Southern California with aftershocks removed, Poissonian? — Bull. Seismol. Soc. Am., vol. 64, 1974, No 5.
12. Kagan, Y., L. Knopoff. Random stress and earthquake statistics: time dependence. — Geophys. J. R. Astr. Soc., 1987, 88.
13. Kagan, Y., L. Knopoff. Statistical study of the occurrence of shallow earthquakes. — Geophys. J. R. Astr. Soc., 1978, 55.
14. Kagan, Y., L. Knopoff. Spatial distribution of earthquakes: the two-point correlation function. — Geophys. J. R. Astr. Soc., 1980, 62.
15. Eneva, M., M. Hamburger. Spatial and temporal patterns of earthquake distribution in Soviet Central Asia: application of pair analysis statistics. — BSSA, vol. 79, 1989, No 4.
16. Eneva, M., M. Hamburger, G. Papadopulo. Spatial distribution of earthquakes in aftershock zones of the Garm region, Soviet Central Asia. — Geophys. J. Int., 1990.

Статистические исследования сейсмического процесса в разломе Сан Андреас

Б. Рангелов, Е. Ракова

(Резюме)

Проведены исследования статистических распределений интервалов расстояний и времени между последовательными землетрясениями, случившимися на протяжении разлома Сан Андреас (Калифорния) в период 1969—1982 гг. Показано, что распределения по времени аппроксимируются достаточно надежно с экспоненциальной функцией, а по расстоянию — с бета-распределениями. Выведены аналитические формулы этих распределений для разных магнитудных интервалов, в которые группированы землетрясения.