Earthquake catalogs: Evaluation of data completeness

Article · January 1998

Vladimir Smirnov
Lomonosov Moscow State University

Volc. Seis., 1998, Vol. 19, pp. 497-510
Reprints available directly from the publisher
Photocopying permitted by license only

Amsterdam B.V. Published under license
in The Netherlands under the Gordon and Breach
Science Publishers imprint
Printed in India

Earthquake Catalogs: Evaluation of Data Completeness

V. B. SMIRNOV
Physics Department, Lomonosov University, Moscow, 119899 Russia

(Received December 29, 1996)

The lowest magnitude of complete reporting and its variations in space and time are estimated for several regional earthquake catalogs. The discussion focuses on questions of methodology and the space-time inhomogeneity of real earthquake catalogs.

Introduction. Studies of the physics of seismicity call for comparison between seismicity parameters derived in different regions for different periods of time. An important factor in this comparison is the homogeneity of raw data. A seismograph network has several properties that affect the completeness of data, similarly to any instrument designed to observe natural processes. The foremost among these properties are sensitivity and resolution. Sensitivity is characterized by a cutoff magnitude, defined as the lowest magnitude of earthquakes that are completely reported in an area. Resolution is determined by the uncertainties involved in the coordinates, occurrence time, and magnitudes of earthquakes.

These characteristics are largely controlled by the density and geometry of the network and the available knowledge of velocity structure and wave attenuation [1]. All of these components vary as our knowledge of velocity structure improves and as the seismic network is progressively developed or reduced to a lower rank. As a result, the cutoff magnitude of an earthquake catalog is necessarily inhomogeneous in time. Catalog completeness is inhomogeneous in space too, because sensitivity and resolution are usually different inside the network and along its periphery.

It has unfortunately not become common practice, either in Russia or abroad, to furnish seismograph networks (as contrasted with other similar instruments, e.g. radio telescopes) with certificates containing information on sensitivity and resolution, to say nothing about the variation of these parameters in time and space. Accordingly, the completeness of a catalog has to be evaluated relying only on the data contained in the

Table 1 Characteristics of catalogs used.

Region	*Time interval covered*	*Number of events*	*Percentage of aftershocks*	*Energy parameter*
Kamchatka	1962-1996	51715	18	Energy class
Kirgizia	250-1992	57500	24	Same
Caucasus	139-1990	48046	36	"
Greece	1964-1995	51658	23	Magnitude
New Zealand	1460-1986	26757	13	"
North China	1970-1996	82155	17	"

catalog itself.

Estimates of cutoff magnitude are usually based on the assumption of a power law form governing the distribution of earthquakes in terms of energy; the magnitude-frequency relation is in that case linear. When some of the events were not reported, data points for the respective magnitudes lie below the magnitude-frequency line, showing as a bend at low magnitudes. Consequently, the problem of finding the cutoff magnitude when stated in statistical terms reduces to the question whether an observed distribution of earthquake magnitudes conforms to a power law. When answered in the negative, the next step is to find out whether the discrepancy is due to the bend. The relevant statistical problem was formulated and solved in [4], [5]. In contrast to numerous previous studies in this field, these authors provided a rigorous statistical solution of the problem which allows one to use a computer-based analytical procedure prescribing a significance level for testing. The present paper reports the results of the practical application of this approach to several earthquake catalogs used in joint research conducted by the Institute of Seismology, United Institute of Physics of the Earth, Russian Academy of Sciences, and the Chair of Physics of the Earth, Physics Department, Moscow University (see Table 1).

Method for estimating completeness. Because the statistical procedures employed here was comprehensively described in [4], [5], we will consider a flow chart of the algorithm used to estimate the magnitude of complete reporting (Fig. 1).

All estimates are derived from frequency distributions of earthquake magnitudes (or energy classes) in the range M_{min} to M_{max}, the length of class intervals ΔM being fixed by the user. The problem is to find the lowest value of M_{min} for which the magnitude-frequency relation is linear in the range $M \geq M_{min}$. The problem is solved by trying increasing M_{min} values. The step δM_{min} is determined by magnitude uncertainty in the catalog under study; we used $\delta M_{min}=0.1$. The starting value $M_{min}=M_{min0}$ may be arbitrary, but not above the magnitude M_0 at the maximum of the histogram. The starting value M_{min0} was chosen from the requirement

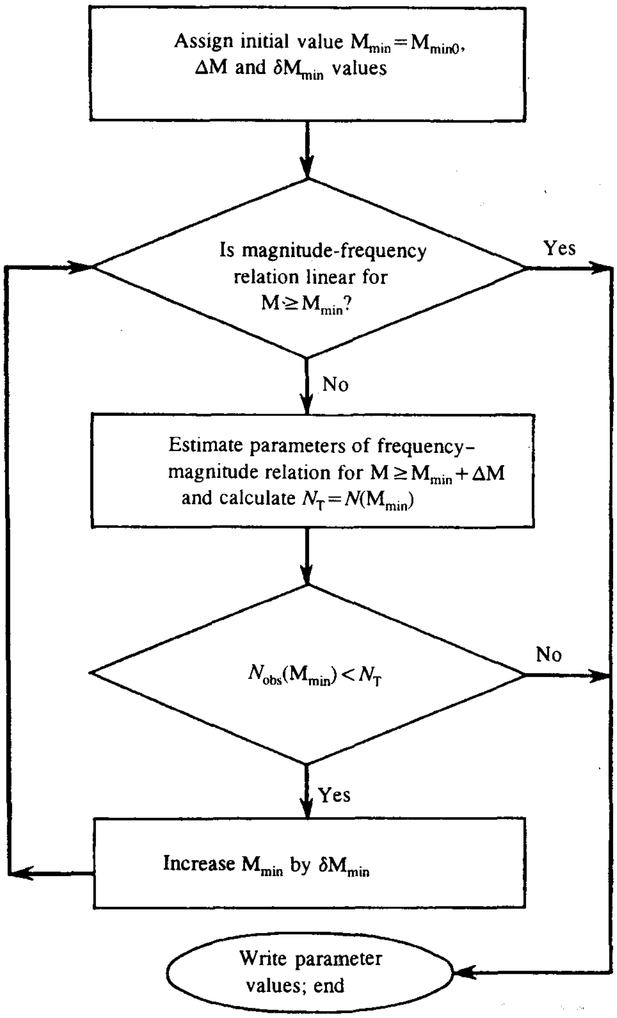

Figure 1 Flow chart of the algorithm for estimating the lowest completely reported magnitude based on earthquake catalog data. For explanations see the text.

$$M_{min0} = \max M : \left\{ \begin{array}{l} M < M_0, \\ N(M_0) - N(M) \leq \Delta N(M_0) + \Delta N(M), \end{array} \right.$$

where $N(M)$ is the number of events in a class interval; $\Delta N(M) = (N(M))^{1/2}$ is its uncertainty.

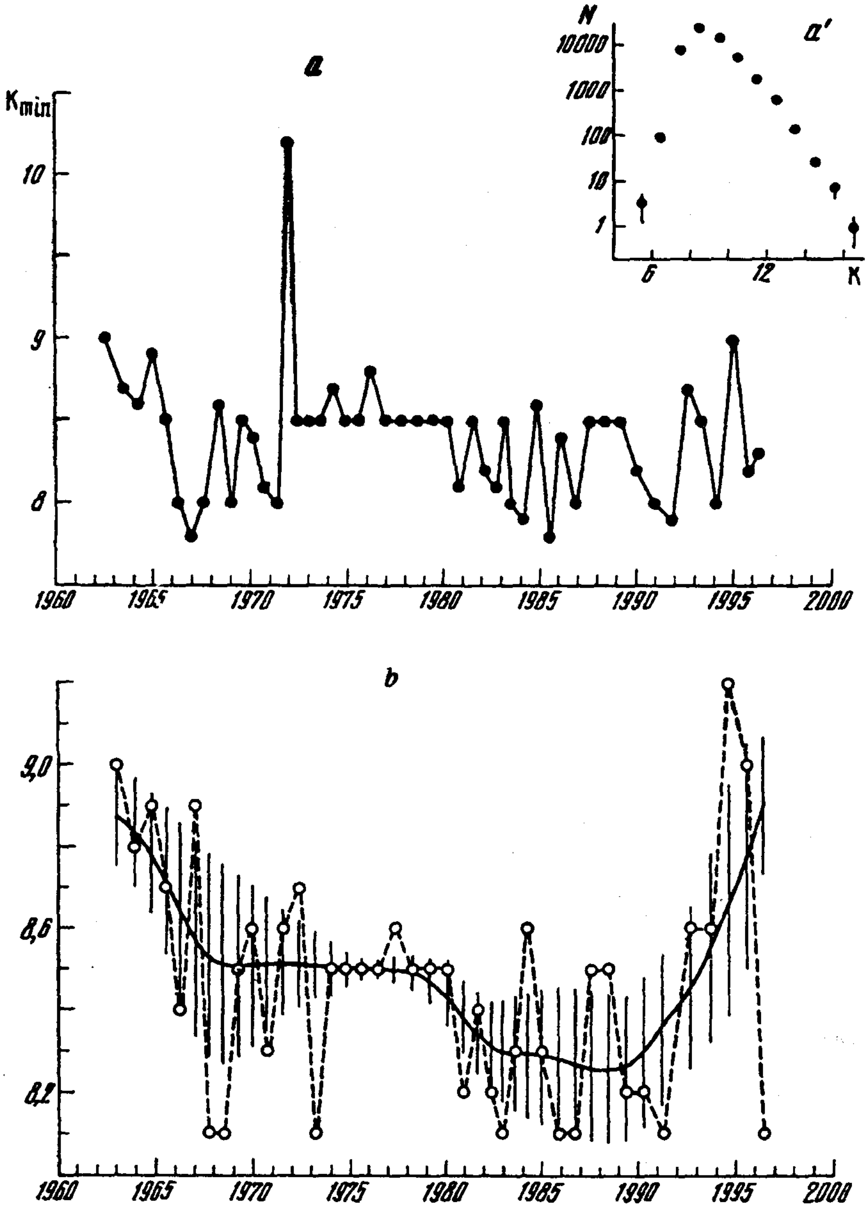

Figure 2 Variation of the lowest completely reported energy class with time for Kamchatka catalog: *a* – full catalog (*a'* – class-frequency plot); *b* – same catalog with aftershocks eliminated.

The first step is to test whether the observed magnitude–frequency relation is a straight line. The test is to see whether the observed histogram is in agreement with a theoretical frequency distribution corresponding with the Gutenberg-Richter law (a polynomial distribution in our case). We used information theory techniques to test the hypothesis, because they are more stable than the conventional Pearson test. If the hypothesis of agreement is accepted at a given significance level, then the respective M_{min} is the desired estimate of the magnitude of complete reporting. The procedure is terminated and the relevant parameter values are written in the file.

The departure of a magnitude–frequency relation from a straight line may either result from a bend at low magnitudes (incomplete reporting) or may be due to natural causes. The question of why the relation is not linear is dealt with at the next step of the procedure. First, the method of maximum likelihood is used to estimate the parameters of the relation (the b-value and the constant term) in the range $M \geq M_{min} + \Delta M$. These parameters are used to find the number of events N_T in the range $[M_{min}, M_{min} + \Delta M]$ (that is, the straight magnitude–frequency relation is extrapolated to M_{min}), the resulting theoretical number of events being then compared to the observed number $N_{obs}(M_{min})$. The condition $N_{obs}(M_{min})$ is tested statistically at a level of significance chosen. When it is found that $N_{obs}(M_{min}) < N_T$, it is concluded that the departure from linearity has been caused by the incompleteness of reporting for magnitude M_{min} events. The value of M_{min} is then increased by the amount δM_{min}, and the entire procedure is repeated. If the condition $N_{obs}(M_{min}) < N_T$ is not true in the statistical sense of the word, it is concluded that the curvilinearity stems from natural causes. The procedure is then terminated, and an appropriate signal is entered in the output file.

This algorithm is implemented as a set of programs which estimate completeness of reporting in earthquake catalogs using moving time windows in fixed spatial areas. The level of significance used in the testing is assigned depending on the amount of data and is entered in the output file.

Results. Figure 2*a* presents a plot showing the variation of the lowest energy class of complete reporting for the regional Kamchatka catalog. The estimates were based on nonoverlapping time windows containing 1000 events each. The 1971 spike was caused by the aftershocks of the large Ust-Kamchatsk earthquake ($K = 15.4$). Similar aftershock-caused spikes were observed for some other catalogs. Aftershock sequences contain an abnormally great number of comparatively large earthquakes compared to the background seismicity. These events are superposed upon the background level and produce a knee in the magnitude–frequency relation. This knee is interpreted by the algorithm as the lowest size reported completely. Some small events may be missed on account of too many large earthquakes being recorded on seismograph channels. For this reason aftershocks should be eliminated from an earthquake catalog before estimating its completeness. Whenever necessary, the completeness within an aftershock sequence can be estimated separately.

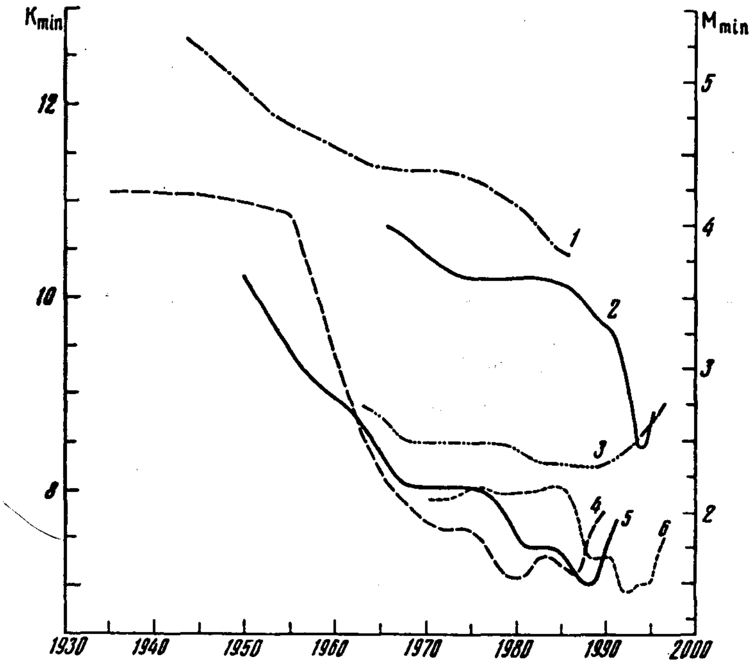

Figure 3 Average curves showing the variation of the lowest completely reported magnitude for the following earthquake catalogs: *1* - New Zealand, *2* - Greece, *3* - Kamchatka, *4* - Caucasus, *5* - Kirgizia, *6* - North China.

Aftershocks were identified using an algorithm reported in [2], [3]. The principle underlying the separation of aftershocks from the other events relies on comparison between their distributions in space and time. These distributions are assumed to be uniform for background seismicity. A bell-shaped (Gaussian) distribution on a plane is assumed for aftershocks (epicenters alone are used) along with the Omori law for time distribution. The parameters involved in these distributions are estimated in an iterative manner as more aftershocks are being identified. The classification criterion, that is, the rule classifying an earthquake as a background event or an aftershock, is derived from the requirement that the probabilities of the two classification errors should be equal, namely, classifying an aftershock as a background event and the inclusion of a background event in the aftershock set. In this way one minimizes the total number of classification errors and makes the mathematical expectation of the number of aftershocks equal to its true value. The performance of this algorithm has been tested by its authors [2], [3] and by the writer of this paper by comparing the aftershocks identified in a regional catalog to the aftershocks recorded by the relevant local seismic network. The discrepancy in the number of aftershocks is about 5%.

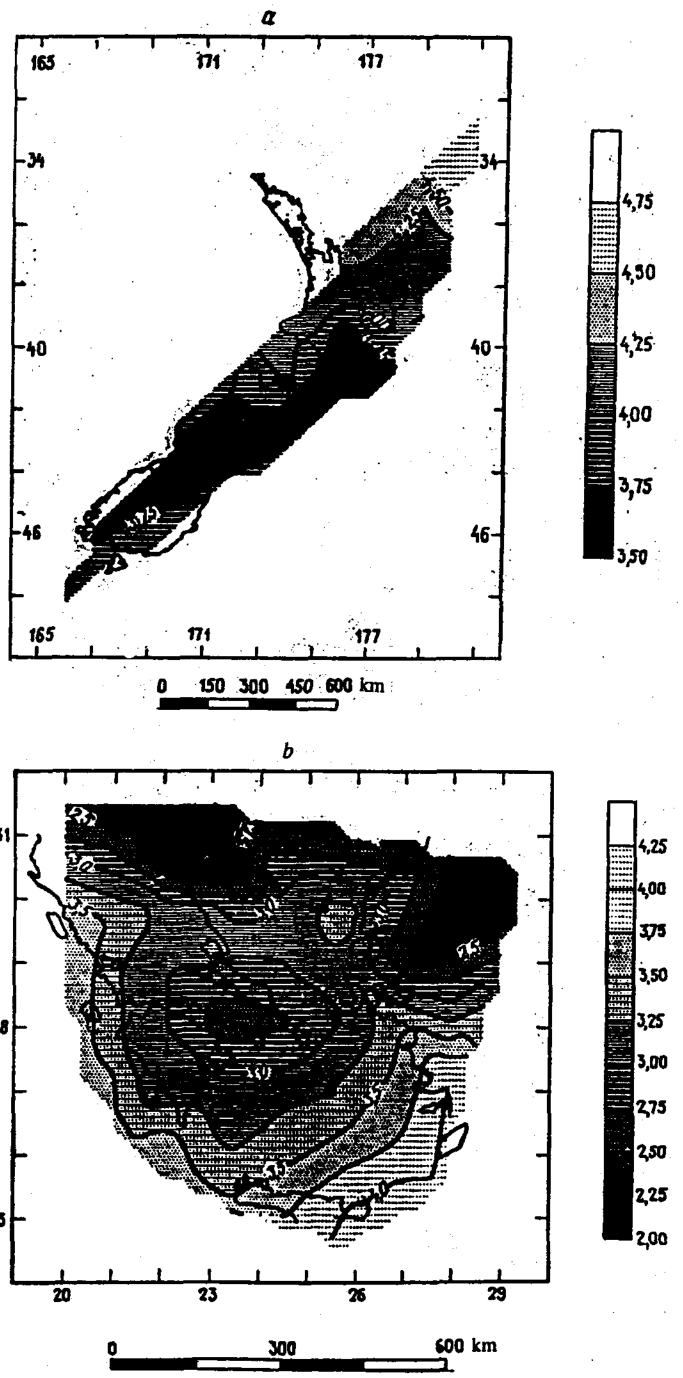
a
165
171
177
34
40
46
4,75
4,50
4,25
4,00
3,75
3,50
0 150 300 450 600 km
b
41
38
35
20
23
26
29
4,25
4,00
3,75
3,50
3,25
3,00
2,75
2,50
2,25
2,00
0 300 600 km

Figure 4 *a, b*

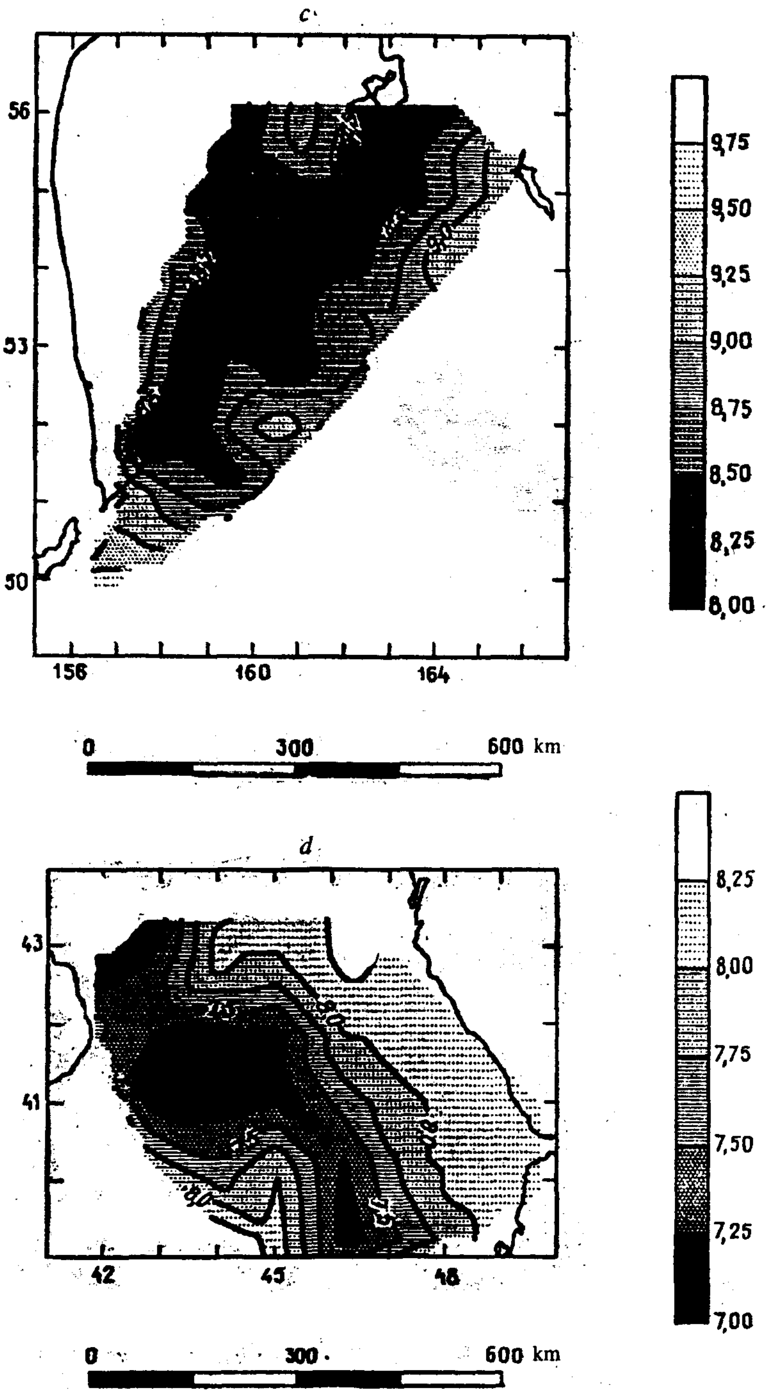
c
56
53
50
156
160
164
9,75
9,50
9,25
9,00
8,75
8,50
8,25
8,00
0
300
600 km
d
43
41
42
45
48
8,25
8,00
7,75
7,50
7,25
7,00
0
300
600 km

Figure 4 *c, d*

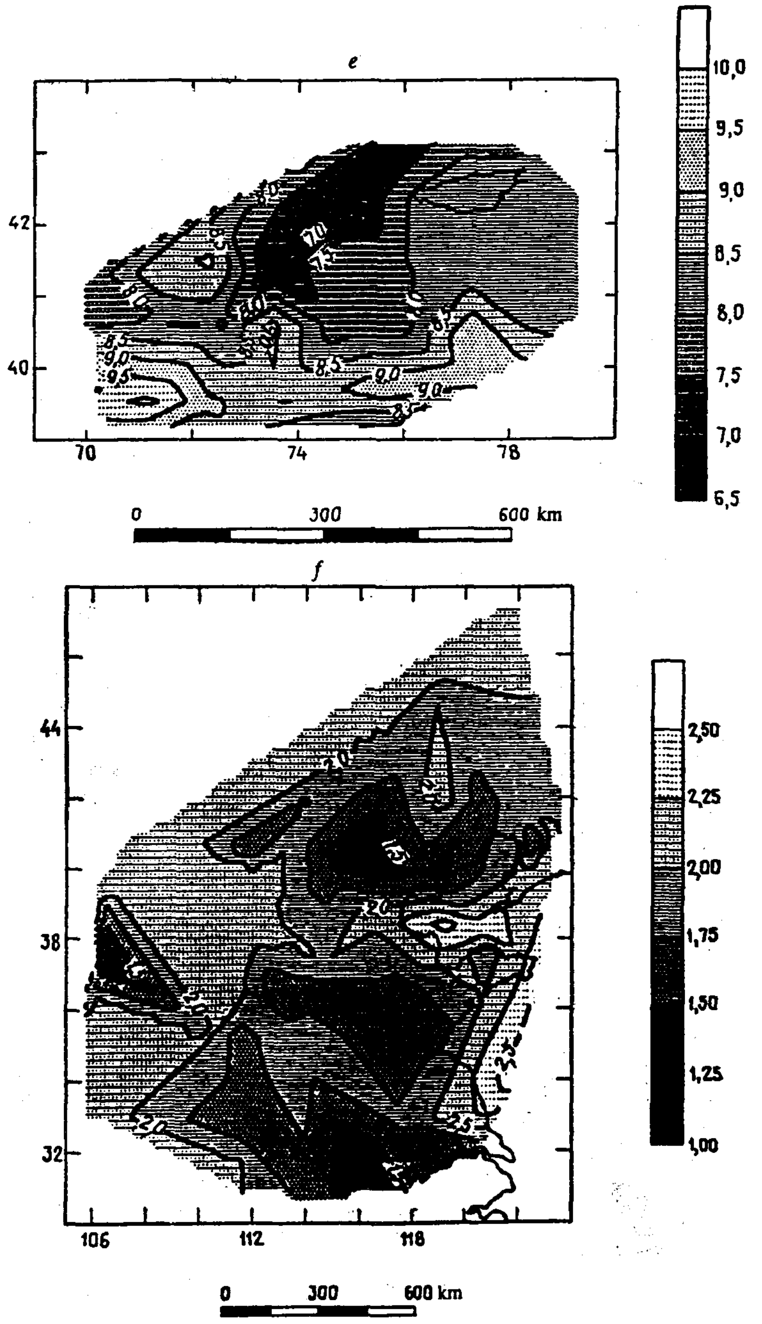

Figure 4 Spatial distribution of the lowest completely reported magnitude (energy class) for the following earthquake catalogs: *a* - New Zealand, *b* - Greece, *c* - Kamchatka, *d* - Caucasus, *e* - Kirgizia, *f* - North China.

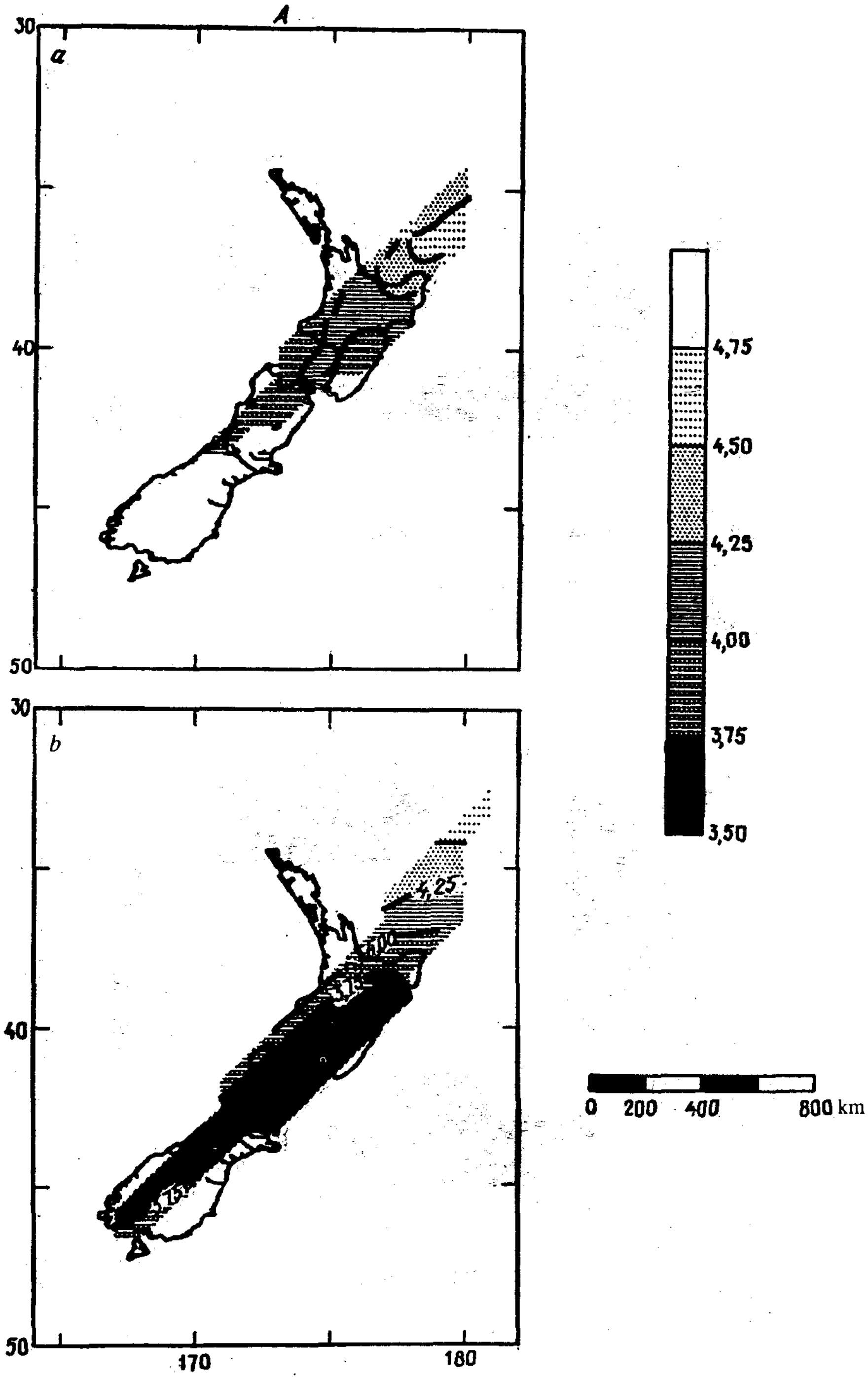
A
a
b
30
40
50
30
40
50
170
180
4,25
4,00
3,75
3,75
4,75
4,50
4,25
4,00
3,75
3,50
0 200 400 800 km

Figure 5 *A*

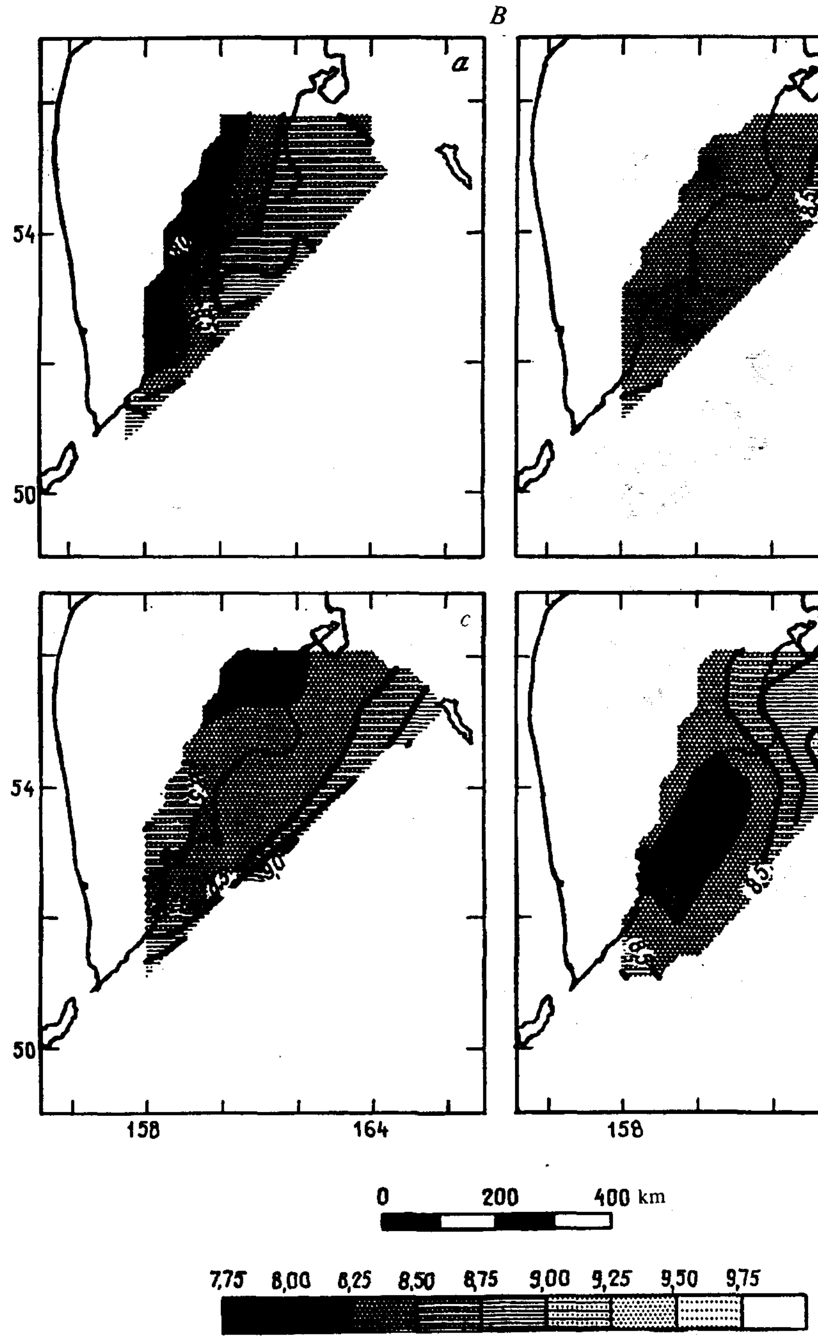
B
a
b
c
d
54
50
54
50
158
164
158
164
0
200
400 km
7,75
8,00
8,25
8,50
8,75
9,00
9,25
9,50
9,75

Figure 5 *B*

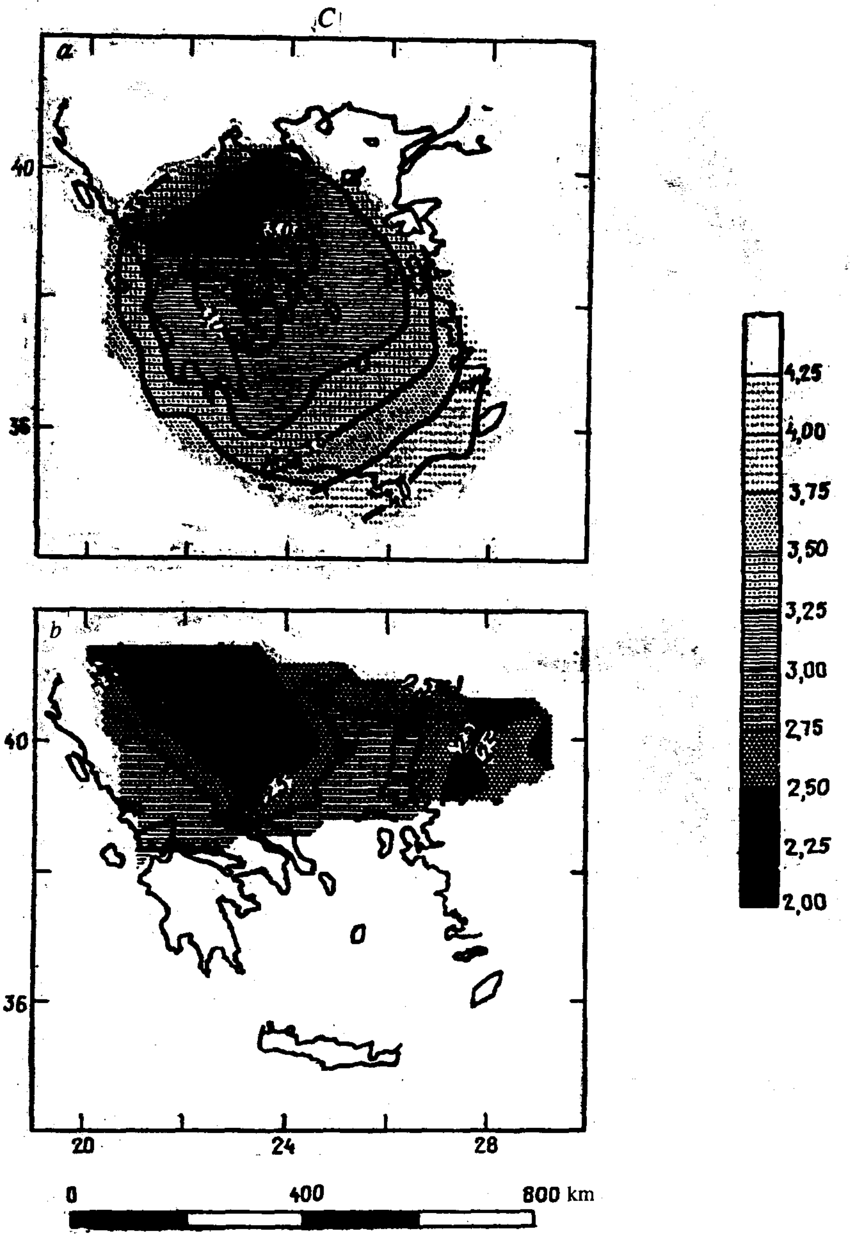

Figure 5 Spatial distribution of the lowest completely reported magnitude (energy class) during various time periods for the following earthquake catalogs: *A* - New Zealand (*a* - 1950 to 1972, *b* - 1973 to 1986); *B* - Kamchatka (*a* - 1962 to 1970, *b* - 1970 to 1979, *c* - 1979 to 1987, *d* - 1987 to 1996); *C* - Greece (*a* - 1964 to 1992, *b* - 1993 to 1995).

Figure 2*b* shows the variation of the cutoff energy class for the Kamchatka catalog with aftershocks eliminated (dashed line). A smooth curve is drawn through the data points to show the overall trend in K_{min}. The scatter of the data points about the smooth curve, indicated by vertical bars, is 0.2–0.4. This is in agreement with the error ± 0.5 caused by the class interval length $\Delta K = 1$ used in this study.

One can see a trend of decreasing cutoff energy class as the seismograph network was expanded prior to 1991. The increase after that date seems to have been due to failing financial support of research and a general deterioration in the economic situation of the region. Some improvement is noticeable in 1996, inspiring the hope that the situation is improving. It should be added that the 1995–1996 data were derived from the fast-processing catalog and are liable to be refined in the future.

Figure 3 presents the smoothed time variation plots of the lowest magnitude or energy class of complete reporting for the catalogs considered in this study. The plots correlate with certain stages in the development of seismic networks concerned. One can notice invariable increases in the cutoff magnitude in the catalogs of the former Soviet republics after the breakup of the USSR.

The plots in Figs 2 and 3 characterize each of the catalogs as a whole. That does not mean however that the magnitudes given in these figures are representative of the entire area sampled. Figure 4 shows spatial distributions of magnitude and energy class calculated for the entire period of instrumental observation in about 100 × 100 km squares. The lowest completely reported magnitude is seen to increase generally from the middle of an area toward the periphery. This tendency is in agreement with theoretical predictions [1]. The more complicated pattern in northern China seems to be due to the fact that data from several networks deployed there were lumped in the respective catalog [6]. The same circumstance is responsible for differences between northern and southern Greece. As to the Caucasus, the lowest completely reported energy class is seen in the Dzhavakhet Upland, where good seismic network coverage is available. The Frunze Earthquake Prediction Site area is prominent in Kirgizia.

The spatial distribution seems to vary in time. Large amounts of data are needed to estimate the distribution of spatial magnitude for successive epochs. Practice shows that samples of at least 500–1000 events are necessary to get reliable estimates of the lowest completely reported magnitude. This restriction determines the detail of the spatio-temporal distributions of the lowest completely reported size in earthquake catalogs attainable with the method discussed. Examples of the spatial distributions of the lowest completely reported magnitude for various time periods are presented in Fig. 5.

Conclusion. The above results provide evidence of considerable time and space variations in the lowest completely reported magnitude for earthquake catalogs. This circumstance should be taken into account when estimating seismicity parameters. When dealing with certain problems, e.g., with seismic zonation, it is sufficient to restrict the data from below by applying a cutoff magnitude, uniform for the entire area of study and time period. In other problems, e.g., in the study of the nature of seismicity and its structure, this approach seems to be unduly wasteful, because some observations that are completely reported in certain space–time volumes would be excluded from consideration. When this is the case, one should carefully examine the spatio-temporal distribution of completely reported earthquake size and make appropriate selections from the catalogs. The procedure discussed here is quite suitable for the purpose.

This work was supported by the Russian Foundation for Basic Research, grant 96-05-64970.

REFERENCES

1. V. Yu. Burmin, *Optimizatsiya seismicheskikh setei i opredelenie koordinat zemletryaseniy* (Optimization of seismic networks and earthquake location)(Moscow: OIFZ RAN, 1995).
2. G. M. Molchan and O. E. Dmitrieva, *Vychislitelnaya Seismologiya* (Computational Seismology) iss. 24: 19-50 (1991).
3. G. M. Molchan and O. E. Dmitrieva, in: *Seismichnost i seismicheskoe raionirovanie Severnoi Evrazii*, iss. 1 (Seismicity and seismic zonation of Northern Eurasia)(Moscow: IFZ RAN, 1993): 62-69.
4. V. F. Pisarenko, in: *Diskretnye svoistva geofizicheskoi sredy* (Discrete properties of the geophysical medium)(Moscow: Nauka, 1989): 47-60.
5. M. A. Sadovskiy and V. F. Pisarenko, *Seismicheskiy protsess v blokovoi srede* (The seismic process in a block medium)(Moscow: Nauka, 1991).
6. Zhang Guomin and Zhang Zhaocheng, *J. Earthq. Pred. Res.* **1**, N1: 71-86 (1992).