Izvestiya, Physics of the Solid Earth, Vol. 41, No. 10, 2005, pp. 798–812. Translated from Fizika Zemli, No. 10, 2005, pp. 30–45.

Multiscale Order Grouping in Sequences of Earth's Earthquakes

A. A. Gusev

Institute of Geology and Geochemistry, Far East Division, Russian Academy of Sciences, Petropavlovsk-Kamchatskii, Russia

Received May 25, 2004

Abstract—Time sequences of earthquakes with removed aftershocks often possess the property of grouping, forming groups (clusters) of events on the time axis. The special phenomenon of order grouping, in contrast to the similar "ordinary" grouping, is examined in the paper. This phenomenon manifests itself in any time-ordered sequence of events of various intensities and consists in the fact that the largest events in such a sequence exhibit the tendency toward grouping. The length of time intervals between events is insignificant in this case. Ordinary and order groupings are different phenomena that can be independent of each other. Tendencies toward grouping of both types can be of multiscale nature. This study revealed multiscale order grouping of earthquakes from world catalogs of 1900–1989 (Abe) and 1977–2000 (Harvard CMT) and estimated its characteristics.

INTRODUCTION

The statistical structure of time sequences of events has been extensively studied. The consideration of events as belonging to the same type is often justified, and in this case there are known such models as the "purely random" Poisson process and its generalizations, the recovery process, and others. Models more adequate in a number of problems are processes with unlimited aftereffect (in particular, multiscale self-similar processes such as fractal and multifractal noises). An approach considering earthquake sequences as self-similar objects has been successfully developed [Prozorov, 1982; Rykunov *et al.*, 1987; Smalley *et al.*, 1987; Kagan and Jackson, 1991; Ogata and Abe, 1991; Hirabayashi *et al.*, 1992; Stakhovsky, 2000]. At a qualitative level, the self-similar behavior is characterized by the tendency toward the formation of groups (clusters) of events closely spaced in time that have various scales and do not exhibit explicit periodicities. Aftershock sequences and earthquake swarms are the most noticeable phenomena of this type. Such short-term grouping is well known and is not the object of this study. We are interested in the tendency toward long-term grouping [Ogata and Abe, 1991; Kagan and Jackson, 1991] observable on larger time scales (usually longer than half a year). Therefore, we analyze sequences (catalogs) that do not contain aftershocks. The tendency toward long-term grouping of occurrence times of events (below, referred to as ordinary grouping) has already been discovered for several data sets and one more study of this tendency is not of great interest. The approach applied in this paper to the study of grouping consists in combined analysis of times and seismic moments ("weights") of catalog events.

As has been well known since the times of Gutenberg and Richter, the distribution of earthquakes over energy or seismic moment is close to a power (hyperbolic) law with the exponent $b = 0.6$–0.7. Under these conditions, the tectonic contribution of the several largest events is predominant within any specific catalog. However, the properties of grouping as studied up to now characterized primarily the most numerous events, which have the smallest magnitudes and thereby contribute relatively little to tectonic processes (and seismic hazard). Our approach takes into account the weight of events explicitly. Below, we define an efficiency function characterizing the generation rate of the seismic moment or energy (or, in other applications, mass, volume, etc.)

Since the ordinary grouping is a widespread phenomenon, two questions arise in relation to a concrete catalog: (1) whether the efficiency function describing a given process possesses properties of self-similar behavior and, if so, (2) whether this self-similar behavior is a consequence of the ordinary grouping of individual events (points) or has a more general nature. The second alternative is realized as a possible occurrence of such a specific phenomenon as the order grouping. This phenomenon can either supplement the well-known ordinary grouping or arise independently. The order grouping can be present in any time-ordered sequence of events of various intensities and consists in the tendency of the largest events of the sequence to be near neighbors. The ordinary and order groupings are qualitatively different phenomena, and methods of their parameterization are developed below independently. Stakhovsky [2000] successfully applied the multifractal approach to the analysis of the efficiency function but dealt with not a catalog but a discrete time series of data summed over narrow time windows without discriminating between the contributions of ordinary and order grouping.

The order grouping of earthquakes was discovered for the first time by Ogata and Abe [1991] in time-ordered sequences of magnitudes from world and Japanese catalogs of earthquakes. These sequences were processed with a constant step as time series (the number of an event in the sequence was meant as its occurrence time). Ogata and Abe discovered that the series they obtained possess a self-similar structure. However, they considered this result as secondary; it has not drawn any attention and has not been further developed in seismology. Order grouping in a sequence of volcanic eruptions was discovered in [Gusev *et al.*, 2003].

In this work, the phenomenon of order grouping is studied on the basis of the Abe and Harvard world catalogs, using a physical value, the seismic moment M_0, as the weight of an event. In the case of the Abe catalog, covering the period 1900–1989, values of M_0 were mostly estimated indirectly. The Harvard CMT catalog (1977–2001) includes values of M_0.

To perform this task, one should have methods for recognizing self-similar behavior of efficiency functions and measuring their characteristics, which is the subject of the introductory part of the paper. The correlation dimension D_c is used for studying this behavior from a limited set of data. The dimension $D_c \equiv D_2$ is the generalized fractal dimension D_q at $q = 2$ for a hypothetical multifractal measure defining the time structure of the efficiency function.

APPROACH TO THE ANALYSIS OF SELF-SIMILAR GROUPING IN A SEQUENCE OF EVENTS OF VARIOUS INTENSITIES

Initial Data Model

Initially, we assume that an idealized catalog of earthquakes can be considered as an object generated by the following chain of operations.

(1) Realizations of two random multifractal measures $X(t)$ and $Y(t)$ are specified in an interval $(0, T)$ of the time axis t. In general, these measures are statistically interdependent.

(2) A realization of the generalized Poisson process of a density $X(t)$ is generated in $(0, T)$. The realization consists of N points $\{t_i; I = 1, 2, \ldots, N\}$ defining the time moments (dates) of catalog events.

(3) The interval $(0, T)$ is divided into N small intervals, each containing one point-event. For definiteness, let a half-interval $(t_{i-1}, t_i]$ contain the ith event.

(4) The measure $Y(t)$ is integrated over each of the smaller intervals, and the result is assigned to the corresponding point-event as its weight.

In this scheme, here referred to as scheme A, $X(t)$ specifies the instantaneous density of the flow of events and $Y(t)$ gives the efficiency; in the case of earthquakes, the latter is the generation rate of the seismic moment. If $X(t)$ = const, the event times form a Poisson process. The measure $Y(t)$ is integrated over small intervals, and their distribution has a finite variance, whereas the distribution of seismic moments is hyperbolic with an exponent of about –0.65. We could not construct a simple model for $Y(t)$ with a hyperbolic distribution of weights in both the general case and the case of the null hypothesis of mutually independent weights of neighboring points. Therefore, although this scheme is physically clear, its direct application to data analysis encounters difficulties. However, it can be checked indirectly, by examining its consequences.

In this connection, we considered an alternative scheme of the data structure (scheme B). As before, $X(t)$ is a random multifractal density generating a set of time moments $\{t_i; i = 1, 2, \ldots N\}$, while the weights of each of the N events are defined as follows.

(1) A set Γ of N independent random values is specified in accordance with a power (Gutenberg–Richter) law.

(2) The set is enumerated so as to generate a positive correlation between the weights of neighboring points (with numbers i and $i + k$); this correlation is described by a function slowly decreasing with increasing k, e.g., a power function. (Technically, this rearrangement can easily be made, for example, by creating a realization Φ of a correlated discrete Gaussian random process of N readings in length, after which each of the N readings in Φ is replaced by the "corresponding" reading from Γ. Here, the correspondence means that the replaced (in Φ) and replacing (from Γ) readings have identical numbers in their variational series, i.e., in permutations of the sets Φ and Γ ordered by increasing weight. If a hyperbolic function of correlation is chosen for Φ, the generated process will be a nearly self-similar multiscale process. However, there is no reason to expect multifractal behavior proper from such a process.)

Scheme B is largely artificial and therefore less appealing but enables a direct analysis of data, as was done in [Ogata and Abe, 1991]. Note that both schemes A and B aim at estimating the exponent in the correlation function (or spectrum). If the initial object is treated as a multifractal, this exponent is the correlation dimension.

Direct Estimation of the Correlation Dimension

To reveal self-similar grouping in sequences of events of various intensities, we used concurrently two approaches, each determining the correlation dimension in the ideal case of a multifractal object. In the case of such an object consisting of points (of unit weight), the correlation dimension can be determined on the basis of the first correlation integral [Bozhokin and Parshin, 2001]:

$$C(d) = (1/N_p)N(d_{ij} < d), \qquad (1)$$

where N_p is the number of pairs of points on the time axis indexed as $1, 2, \ldots i, \ldots j, \ldots n$ ($N_p = n(n-1)/2$), $d_{ij} = t_j - t_i$ is the time interval between the components of a pair with the numbers i and j, and $N(d_{ij} < d)$ is the number of pairs with values $d_{ij} < d$. Formula (1) is standard but, if data from a limited time interval (of the length T_0) are treated, gives a biased estimate, which is usual in estimating the correlation function. The pertinent correction is easily obtained in this case, and the unbiased estimate has the form

$$C(d) = \frac{N(d_{ij} < d)}{N_p(1 - d/2T_0)}. \quad (2)$$

Estimates (1) and (2) can be derived directly from the "observed efficiency function"; in the given case, the latter has the form

$$p(t) = \sum_{i=1}^{n} V_i \delta(t - t_i), \quad (3)$$

where V_i is the mass of the ith point. We set temporarily $V_i = 1$ and denote this variant of the function $p(t)$ by $p_1(t)$. Note that the smoothing with a window of the width T transforms $p(t)$ into an empirical estimate of efficiency with the resolution T, whereas an instantaneous value of the efficiency cannot be determined. Similarly, smoothing of $p_1(t)$ yields an empirical estimate of the flow density of events. The ordinary estimate of the second multivariate correlation moment of the function $p_1(t)$ has the form

$$B_1(d) = \frac{1}{\mathrm{T}}\int p_i(t) p_1(t+d)dt, \quad (4)$$

and the integral $\int_0^d B_1(x)\,dx$ is identical with $N(d_{ij} < d)$. This representation elucidates the meaning of generalization (2) for the case of points with different masses V_i. The accumulated number of pairs $N(d < d_{ij})$ is replaced by the accumulated sum of weights of the pairs $W(d < d_{ij})$, so that the contribution of a pair of events (i, j) to $W(d < d_{ij})$ is equal to the product of their masses $V_iV_j = w(d_{ij})$. Then, the generalized analogue of formula (2), i.e., the estimate of the correlation integral allowing for masses, assumes its final form:

$$C_w(d) = \left(\frac{1}{1 - d/2\mathrm{T}}\right)\frac{W(d_{ij} < d)}{W(\text{все } d_{ij})}$$

$$= \left(\frac{1}{1 - d/2\mathrm{T}}\right)\frac{\sum_{d_{ij} < d} w(d_{ij})}{\sum_{\text{все } d_{ij}} w(d_{ij})}. \quad (5)$$

If the behavior of the efficiency function is self-similar (scale-invariant), $C(d)$ and $C_w(d)$ must be power functions. It is shown in [Bozhokin and Parshin, 2001] that, in the 1-D case,

$$C(d) \sim d^{D_c}, \quad (6)$$

where D_c is the correlation dimension. Let D_{cw} be the value of D_c characterizing $C_w(d)$. To estimate D_c from observations, we can take the logarithm of both parts in (6) and apply the linear regression procedure to the resulting linear relation

$$\log C(d) = \text{const} + D_c \log f. \quad (6a)$$

In this case, it is convenient to represent $C(d)$ not by all readings but by a small number (7–20) of values corresponding to a set of d values with a constant (or nearly constant) step on a logarithmic scale. This is justified because (1) weakly correlating points are selected and (2) different scales have the same weight, although the points are unequally spaced. Below, we adhere to this technique and use a step of the order 1/2–1/3 of the octave (0.1–0.15 of a common logarithm unit).

Estimation of the Correlation Dimension with the Use of the Power Spectrum

An alternative method of correlation dimension determination uses the Fourier transform of $B(d)$, which is the estimate of the power spectrum of the efficiency function $p(t)$ [Pisarenko and Pisarenko, 1991]. If $p(t)$ has a self-similar structure, this spectrum $B(f)$ must also be a power function; in this case, we have [Bozhokin and Parshin, 2001]

$$B(f) \sim f^{D_c - 1}. \quad (7)$$

In using (7) for the estimation of D_c from observations, it is appropriate to integrate the power spectrum (by analogy with (4)); this eliminates the problem of invalid pointwise estimates of the spectrum [Pisarenko and Pisarenko, 1991]. In what follows, we proceed from the relation

$$U(f) = \int_0^f B(f')df \sim f^{D_c}. \quad (8)$$

Then, data are treated in a way quite similar to the case of $C(d)$ (formulas (6) and (6a)):

$$\log U(f) = \text{const} + D_c \log f. \quad (8a)$$

Note that, if the Poisson process is not self-similar ($D_c = 1$), $C(d) \sim d$, $B(d)$ and $|B(f)|$ are white noises, and $U(f) \sim f$. If $D_c < 1$, (6) is the spectrum of flicker noise. To simplify the notation, the estimates of D_c and D_{cw} from the spectrum are denoted below as D_s and D_{sw}.

Real data often have a more complicated structure compared to the representations described above. Below, particularly important is the case when a tendency toward periodicity of events is superimposed on the mainly self-similar grouping. Because of the definition of the functions $C(d)$ and $U(f)$ through integrals, the presence of one or several outliers in the spectrum and/or the autocorrelation curve should distort the hypothetical power structure of these functions. In the simplest case of a single characteristic period T^*, one may expect that this will not lead to overly large distortions in the estimates of D_c from $U(f)$ at frequencies lower than $1/T^*$ and from $C(d)$ at delays d no longer than $(0.5–0.8)T^*$.

Significance Test of the Self-Similar Grouping and the Estimation Accuracy of the Parameter D_c

At first glance, this problem should not be difficult because the estimation of the parameter D_c using (6) and (8) is linear regression and the grouping significance test consists in checking whether the confidence interval of a given estimate of D_c includes the value $D_c = 1$. Actually, the situation is more complex. We discuss this problem in the specific case of formula (6a). As noted above, in order to construct the regression line, a set of arguments d_k is specified and the values $C(d_k) \equiv C_k$ are selected under the condition that the values C_k are correlated and nonequally spaced. Moreover, since actual volumes of data are limited, the pairs (C_k, d_k) have distortions due to the effect of a small sample, complicating the estimation, particularly with D_c close to unity. Under these conditions, reliable estimates of significance are difficult to obtain.

Therefore, we chose the Monte Carlo method for the significance test. We analyzed the observed catalog and a series of its randomized analogues ("synthetic data") and compared the results; below, these randomized analogues, devoid of the time structure, are referred to as "pseudocatalogs." Information on the time structure includes the set of event times t and an ordered sequence of their values M_0. Observations can differ from nonstructural synthetic data in both of these aspects. First, the observed times can be non-Poissonian and have a tendency toward the formation of groups and clusters, which is called ordinary grouping. Second, the observed sequence of values M_0 can differ from a random permutation of such values, forming groups of large events, and such behavior is called order grouping. Therefore, two randomizations should be made for each pseudocatalog. The first randomization, by time (indicated below by the index "RT"), destroys the ordinary grouping: the observed set of times t_i is replaced by a model set of times corresponding to a randomization of a Poisson sequence of the same length T_0, whereas the set of mass values remains the same. The second randomization, by order (indicated below by the index "RO"), destroys the order grouping: values of masses of n events are randomly reshuffled, and the times of these n events remain the same. In constructing each pseudocatalog, both randomizations should be performed (in an arbitrary order) for testing the simplest hypothesis "$D_{cw} < 1$." Applying this scheme, we discuss estimation of D_{cw} alone (analysis of estimates for D_{cw}, as well as for D_c and D_s, is the same).

We created sufficient numbers of pseudocatalogs ($N = 500–5000$), calculated N estimates $D_{cw} = B_{cs}^{(\mathrm{synth})}$, and constructed their empirical distribution function. This function can be reduced to the case of the null hypothesis "true D_{cw} is zero" and can then be directly applied to the determination of the significance level for a concrete value $D_{cw} = D_{cw}^{(\mathrm{obs})}$ obtained from observations. If the estimate $D_{cw}^{(\mathrm{synth})} < D_{cw}^{(\mathrm{obs})}$ is obtained in N_Q cases from N_{MC} pseudocatalogs, the direct estimate of the significance level is $Q = N_Q/N_{MC}$. At small values of Q, the relative accuracy (variation coefficient) of the Q determination by the Monte Carlo method is close to $\delta Q_{MC} = N_Q^{-0.5}$. The value of N_{MC} is chosen large enough to attain an acceptable accuracy of the Monte Carlo Q determination (usually no worse than 10%). To guarantee the result, we always used the modified value of the significance level $Q_{\mathrm{mod}} = Q_{MC}(1 - N_Q^{-0.5})$, close to the upper 84% confidence level for Q. At the final step, the inferred estimate of the significance level was rounded upward.

In the case of the null hypothesis, the empirical estimate of the variance $\sigma_0^2(D_{cw}^{(\mathrm{synth})})$ from pseudocatalogs yields an acceptable (albeit somewhat elevated) estimate of the error in the values $D_{cw}^{(\mathrm{obs})}$. Finally, it is useful to have the average value $\overline{D_{cw}^{(\mathrm{synth})}}$ from the estimates $D_{cw}^{(\mathrm{synth})}$, which must be equal to unity in the ideal case. However, at $n < 1000$ and even more so at $n = 25–100$, systematic distortions arise due to a small size of the sample. Heavy weights in the tail of the distribution of M_0 also have a certain effect. In a first approximation, the influence of these distortions on the estimates derived from observations can be compensated by using the deviation of $\overline{D_{cw}^{(\mathrm{synth})}}$ from unity as a correction to the direct estimate $D_{cw}^{(\mathrm{obs})}$ from the observed catalog. Below, the following corrected values are used as observed parameters:

$$D_{cw}^{(\mathrm{c})} = D_{cw}^{(\mathrm{obs})} + (1 - \overline{D_{cw}^{(\mathrm{synth})}}). \quad (9)$$

In what follows, we do not consider results from individual pseudocatalogs and, for simplicity of notation, $D_{cw}^{(\mathrm{synth})}$ means the average over pseudocatalogs.

The Role of the Two Types of Grouping

To gain the simplest idea of the factors responsible for the observed behavior of the efficiency function, it is desirable to examine separately the contributions of both types of grouping. One of the approaches to this problem is based on scheme A described above and consists in the comparison between the estimates of the efficiency function D_{cw} obtained from observations and a partially randomized catalog. Thus, in order to verify the significance of the ordinary grouping, one can apply the randomization RT alone and then perform the same procedures as in the preceding section. Namely, the estimate $D_{cw}^{(\text{obs})}$ from observations should be compared with the estimate $D_{cw(\text{RT})}^{(\text{synth})}$ obtained as the average over estimates derived from the set of pseudocatalogs randomized in accordance with scheme RT alone. The significance level for the hypothesis "$D_{cw}^{(\text{obs})} < D_{cw(\text{RT})}^{(\text{synth})}$" is estimated from this set as the relative amount of such estimates that are smaller than $D_{cw}^{(\text{obs})}$.

In addition to the significance level, it is useful to estimate the contribution of the ordinary grouping to the overall result D_{cw} by the relation

$$D_{cw(\text{RT})} = 1-(D_{cw(\text{RT})}^{(\text{synth})} - D_{cw}^{(\text{obs})}). \quad (10)$$

As is easily seen, this estimate is qualitatively acceptable. In the absence of ordinary grouping, we have $D_{cw(\text{RT})}^{(\text{synth})} \approx D_{cw}^{(\text{obs})}$ (on average, the RT randomization, changing times but preserving the order, does not change anything) and $D_{cw(\text{RT})} \approx 1$, which means that the contribution of the ordinary grouping to the self-similar behavior of the efficiency function is absent. In the absence of the order grouping, we have $D_{cw(\text{RT})}^{(\text{synth})} \approx 1$ (the RT randomization destroys completely the time and, thereby, any structure) and $D_{cw(\text{RT})} \approx D_{cw}^{(\text{obs})}$, which means that the contribution of the ordinary grouping to the self-similar behavior of the efficiency function is predominant. Finally, it is convenient that the test of the clear hypothesis "$D_{cw(\text{RT})} < 1$" is strictly equivalent to the test of the more sophisticated "$D_{cw}^{(\text{obs})} < D_{cw(\text{RT})}^{(\text{synth})}$," discussed in the preceding paragraph. In a quite analogical way, using the RO randomization alone, the significance of the presence of order grouping is assessed and the parameter $D_{cw(\text{RO})}$, estimating its individual contribution to $D_{cw}^{(\text{obs})}$, is constructed.

Another approach to discrimination between the contributions of the ordinary and order groupings relies on the data representation using scheme B. In this case, following Ogata and Abe, time intervals between events are artificially set equal to each other in order to verify the presence of order grouping and D_{cw} is estimated on this modified scale for both a real catalog and a pseudocatalog. Such estimation is performed through the replacement of real or model dates by a sequence of dates with a constant time step, retaining the same total duration T_0. The modified time (the argument of the thus-constructed time series) is denoted by t^*, and the corresponding frequency, by f^*. In this approach, the estimates D_{cw} and D_{sw}, denoted below as D_{cw1} and D_{sw1}, assume a new, unconventional meaning and, what is more important, become independent of the actual set of exact dates of events. As a result, this yields a "pure" measure of the extent of order grouping. Estimating D_{sw1} from the spectrum, we used the discrete Fourier transformation applied to a set of n points. The contribution of ordinary grouping is determined in this approach traditionally, from the estimates D_c and D_s obtained with unit weights of points (events). We should emphasize that the estimates D_{cw} and D_c are independent and the same is true of D_{sw1} and D_s.

Thus, five estimates of the self-similar behavior can be examined for each catalog: the total estimate D_{cw}, reflecting the self-similar behavior of the efficiency function on the whole; the estimates $D_{cw(RO)}$ and D_{cw1}, reflecting the order grouping alone; and the estimates $D_{cw(\text{RT})}$ and D_c, reflecting ordinary grouping alone. Of course, estimates within each of these two pairs are not independent, but their joint analysis can be helpful. In addition, we have five analogical spectral estimates: D_{sw}, $D_{sw(\text{RO})}$, D_{sw1}, $D_{sw(\text{RT})}$, and D_s.

To illustrate the notion of order grouping, Fig. 1 presents several artificial sequences differing in grouping properties. The plots (a)–(c) in this figure show variants of the same sequence with artificially created order grouping: large events are neighbors. In order to attain a graphically clear correlation of neighboring values, we used a smooth function for the generation of weights and did not intend to construct a self-similar sequence. Thus, the parameters D_{cw1}, D_{cw1}, and D_c were invoked here only to illustrate their capabilities of recognizing various types of grouping, without relation to the property of self-similarity. Plots (a)–(c) differ in time structure: (a) a constant time step, (b) a Poisson sequence, and (c) a sequence with ordinary grouping (dense clusters are clearly seen in this sequence). In all of the three cases, the parameter D_{cw1} (its accuracy is about ±0.5) is significantly lower than unity, which indicates the presence of order grouping. The similar plots (d)–(f) in Fig. 1 involve the same set of times and the same set of event weights, but these weights are randomly permuted here. Now, the property of "neighborhood of large events" becomes invalid and D_{cw1} is close to unity. The parameter D_c (its accuracy is about ±0.02) in plots (c) and (f) is significantly smaller than unity, as is characteristic of the ordinary grouping; on the contrary, no clusters are present ($D_c \approx 1$) in plots (a), (b), (d), and (e). (The apparent temporal variations in the density of events visible in Figs. 1b and 1e are statistically insignificant, as is evident from the value $D_c \approx 1$.) The values of the parameter D_{cw} (its accuracy is about ±0.07) clearly reflect the absence or presence of any type of grouping and are smallest in the case (c), when both types are equally present.

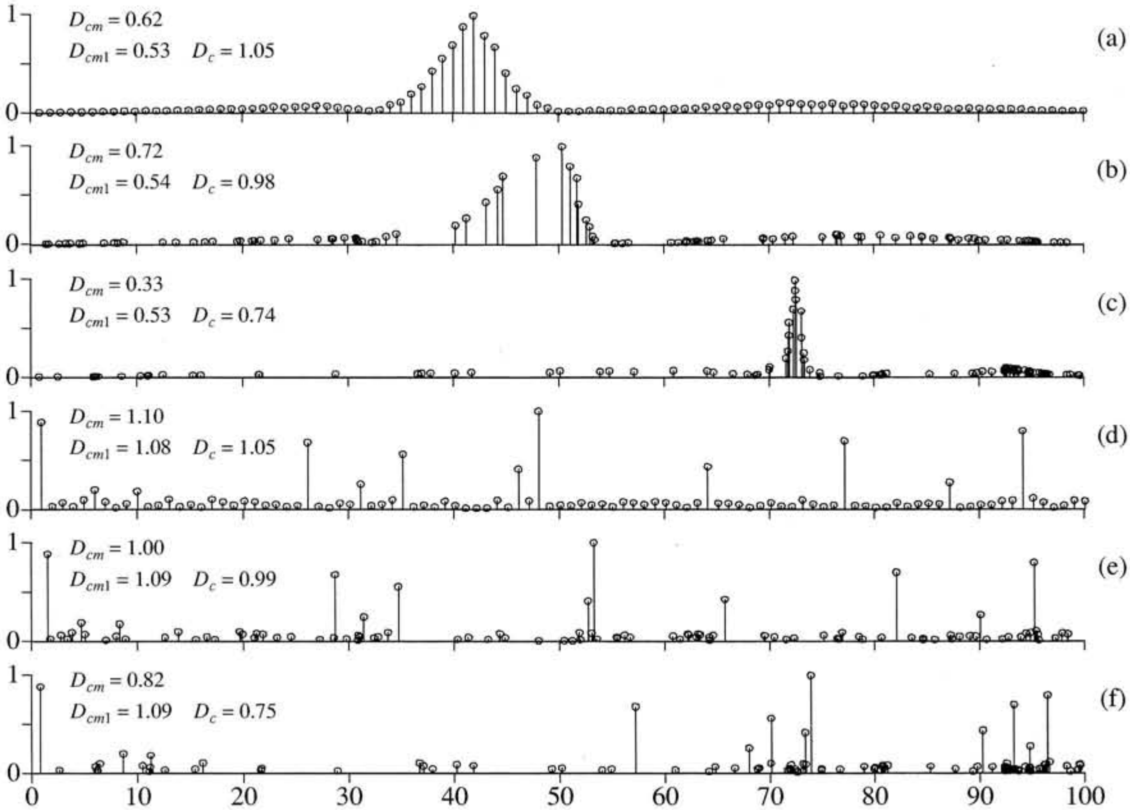

Fig. 1. Artificial examples of short ($N = 100$) sequences of points differing in weight and grouping properties: (a) order grouping for weights with a constant (unity) step in t; (b) order grouping for weights with a Poisson sequence of times t; (c) order grouping for weights with ordinary grouping in t; (d–f) the same as in (a–c) but with the order grouping being destroyed by permutations of numbers of points. Italicized are values of parameters statistically indistinguishable from unity.

INITIAL DATA

We used two world catalogs of earthquakes: the Abe catalog [Abe, 1981, 1984; Abe and Noguchi, 1983a, 1983b], containing only shallow earthquakes and covering the period 1900–1989, and the Harvard catalog, covering the period 1977–2001. The initial version of the Abe catalog was compiled after [Pacheco and Sykes, 1992]. It is important to note that Pacheco and Sykes, following [Perez and Scholz, 1984], "corrected" values of M_s from the Abe catalog, proceeding from the disputable a priori idea that the yearly number of earthquakes having magnitudes above a certain threshold must be stable. In the opinion of the authors of these two papers, variations in the yearly number of events in the Abe catalog are fictitious, reflecting systematic errors that varied in a supposedly consistent manner at the majority of the world seismic stations. Figure 2 plots the cumulative number of $M_s > 7.2$ events for the original version of the Abe catalog (a) and after the introduction of corrections from [Pacheco and Sykes, 1992] (c). It is seen that, decreasing M_s values by 0.1 in 1900–1915 and by 0.2 in 1915–1948, these corrections did smooth the behavior of the yearly number of events at the level M_s = 7.2–7.5 and decreased drops in the density of events in 1949–1966 and particularly in 1970–1989, as well as a less pronounced drop in 1900–1915. However, if the magnitude threshold is elevated to M_s = 7.9–8.0, it is seen from Figs. 2b and 2d that the same drops persist even after the introduction of corrections from [Pacheco and Sykes, 1992]. This implies that the variations in the flow density of events are by no means an artifact. For this reason, the corrections from [Pacheco and Sykes, 1992] were discarded in the construction of the working catalog. The lower threshold of magnitudes was set at $M_s = 7.2$.

The Abe catalog gives values of the magnitude M_s, whereas a catalog presenting values of the seismic moment M_0 is preferable for our purposes. For the bulk of events, the values M_0 were calculated from M_s by the correlation formula [Ekström and Dziewonski, 1988]

$$\log M_0 \text{ [dyn cm]} = 1.5 M_s + 16.14. \qquad (11)$$

In the case of events with $M_s > 8.2$, for which this approach is incorrect, we used the work [Fujita, 2001], where published values of the seismic moment M_0 are summarized for many strong events. In some cases, estimates of M_0 were taken from [Pacheco and Sykes, 1992]. The Kanamori formula $M_w = (2/3)\log M_0 - 10.7$ was utilized for the determination of the moment magnitude.

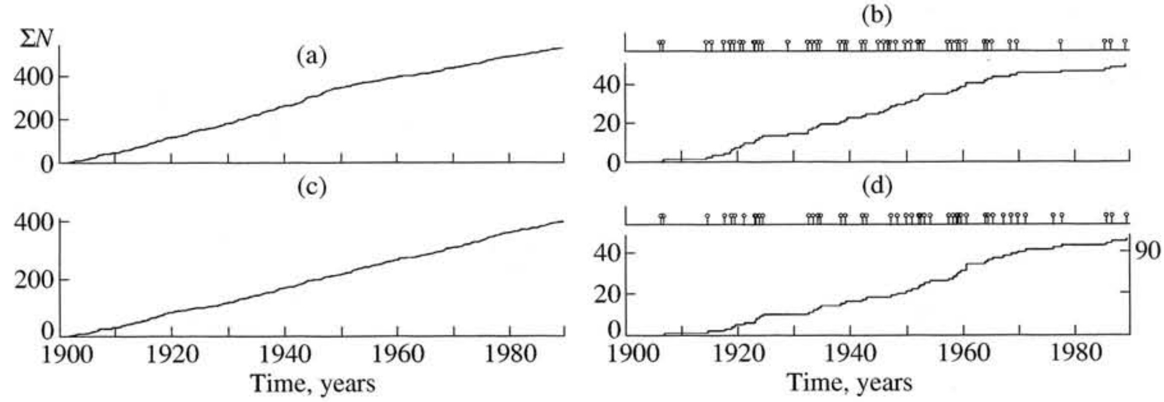

Fig. 2. Time behavior of the cumulative number of earthquakes from the Abe catalog in the initial variant (a, b) and with corrections after [Pacheco and Sykes, 1992] (c, d). The lower threshold of the magnitude M_s and the number of events n are (a) 7.2 (n = 535), (b) 8.0 (50), (c) 7.2 (402), and (d) 7.9 (47). The corrections flatten the plot at the level M_s = 7.2–7.5, but the main features of nonstationarity (higher densities in 1915–1924 and 1932–1969) remain unchanged at the level M_s = 7.9– 8.0.

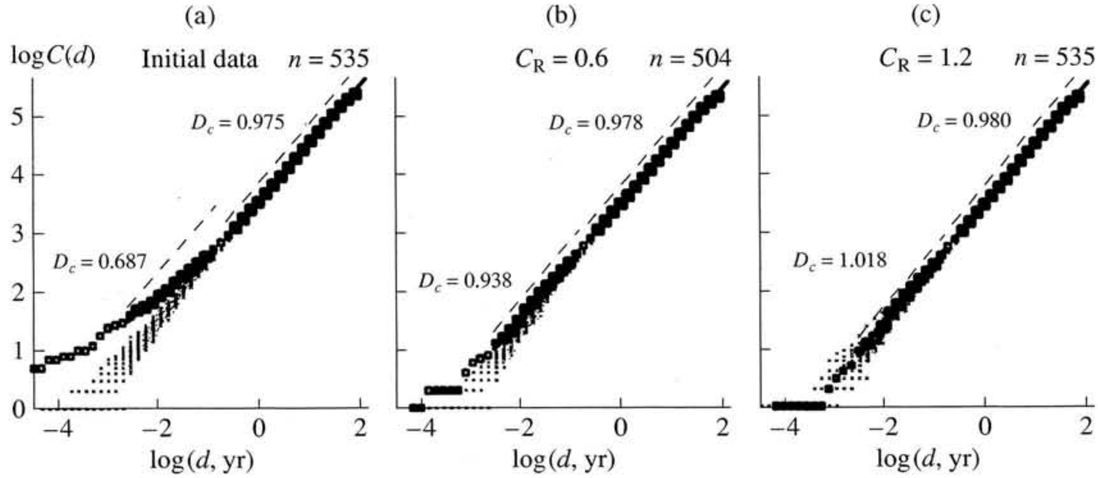

Fig. 3. The aftershock elimination effect on the correlation integral curve for the Abe catalog with (a) and without (b, c) elimination of aftershocks at C_R = 0.6 (b) and C_R = 1.2 (c). Henceforward, D_c was estimated from discrete points of the empirical dependence $\log C(\log d)$ chosen at a step of 0.15 on the $\log d$ scale. The larger squares are points included in regression. The thick line is the direct regression, and the broken line with a slope of unity is given for reference. The smaller symbols, plotting the correlation integral for the first 25 synthetic sequences, show the internal scatter in the variants of the empirical curves $\log C_w(\log d)$ and clearly illustrate deviations of the observed curve from variants of synthetic data.

The resulting catalog contained n = 536 events. Swarms and aftershocks were then eliminated from this catalog (the algorithm is described in Appendix). Figure 3 plots the dependence $\log C(\log d)$ for the original and the processed catalogs, the latter in two versions, of moderate (C_R = 0.5) and stringent (C_R = 1.2) elimination of aftershocks. In each case, dependence (6a) constructed for the range d = 0.5–90 yr yields an estimate of the slope D_c close to 0.98. It is clear that, with such a choice of the range, the influence of aftershocks is relatively small. On the other hand, estimates of D_c for the range d = 0.003–0.1 yr vary significantly.

The estimate D_c = 0.67 was obtained for the range d = 0.003–0.1 yr without elimination of aftershocks. However, this peculiar value has no clear sense because the plot $\log C(\log d)$ has a significant curvature. The branches for small and large delays join smoothly in the interval 0.1–0.5 yr. Figure 3a clearly demonstrates the presence of two qualitatively different phenomena of the "aftershock" and "long-term" grouping. We should emphasize that, although the value of D_c is close to unity in the range d = 0.5–90 yr, it is smaller than unity at a high significance level.

In the catalog version with eliminated aftershocks, the D_c estimates for the range d = 0.003–0.1 yr are closer to or even greater than (at C_R = 1.2) unity. The difference in the estimates is clearly related to the extent of elimination of aftershocks. Generally speaking, the elimination of aftershocks is by no means a rigorous procedure and the choice of the elimination extent regulates the relationship between elimination errors of two types: (1) "missing a real aftershock" and (2) "erroneous elimination of a non-aftershock" [Gusev, 1971]. For the purposes of this work, it is important that residual aftershocks result in an overestimated extent of grouping, whereas the elimination of a certain number of non-aftershocks underestimates the grouping extent and, in the worst case, will give a negative result. It is obvious that, under these conditions, a redundant elimination of aftershocks is preferable to their insufficient elimination. The formal estimate $D_c > 1$ for the range $d = 0.003$–0.1 at $C_R = 1.2$ implies that the aftershock elimination with $C_R = 1.2$ is stringent enough: no grouping is present in the catalog and aftershocks are redundantly suppressed. Below, we used solely catalogs with aftershocks eliminated at $C_R = 1.2$. To decrease distortions, the lower boundary of the working range of delays may be set at 0.25 yr because, judging from Fig. 1c, the self-similarity hypothesis (the D_c independence of scale) seems reasonable within the range 0.25–90 yr. However, further analysis showed that the lower boundary d should be still higher.

The calibration of events of the Abe catalog being somewhat reliable, we invoked data from the Harvard catalog. The small overlapping time interval of the Harvard and Abe catalogs was considered insignificant. We selected the $M_w > 6$ events of 1977–2001. The initial catalog contained 3001 events, but their number was reduced to 2572 after the elimination of aftershocks with $C_R = 1.2$. The number of residual aftershocks is much smaller than in the Abe catalog, and the slope of the curve $\log C(\log d)$ is appreciably greater than unity, which can indicate an effect of the Geiger counter dead time type. This fact suggests that some events are missing in the Harvard catalog due to "blindness" of the Harvard system of data processing within a time interval up to half a day long after a strong earthquake. Since the Harvard catalog is regarded as a "true" catalog by many seismologists, this problem is significant for gaining deeper insights into the actual reliability of its data used in various types of analysis. We tested our suggestion by analyzing data over the period from January through June 1996. The ISC bulletin of 2001 contains 52 events with M_s(ISC) ≥ 6.0. The Harvard catalog does not give M_w(HRVD) values for 17 (33%) of these events. Of course, this is an overestimated result because some of the M_s(ISC) values are unreliable. Rejecting 13 events for which a small number of stations (1–5) are involved in the M_s(ISC) estimation, we find that 4 (10%) of the remaining 3 events do not have values of M_w(HRVD). This is a rather appreciable fraction. Two of these four events are not aftershocks. Thus, one should expect that estimates from the Harvard catalog can be slightly biased toward higher values of D_c, thereby underestimating the significance of grouping effects.

ANALYSIS OF THE ABE CATALOG

The structure of the time behavior in the case of the Abe catalog is well recognizable in plots constructed for an abridged catalog (Fig. 4). Direct inspection of the plots reveals the presence of ordinary and order groupings. Figure 5a presents the plot $\log C_w(\log d)$ for the complete catalog. The plot is seen to be sharply nonlinear, is S-shaped, and has a few benches and plateaus. Taking into account the log–log scale, the most pronounced bench is observed in the interval $d = 2$–5 yr. The probable origin of this anomaly becomes clear from Fig. 4: the largest release of the seismic moment is associated with the activation of 1950–1965, and events inside this interval have a tendency toward periodicity with the periods observed. Another interesting feature of the plots is drops about 2–5 yr long in world seismicity after groups of or single earthquakes in 1906, 1923, 1952, and 1960. Both factors are responsible for the formation of "random" benches and plateaus in the empirical plot $\log C_w(\log d)$. Reliable analysis of the correlation dimension using formula (6a) is impeded by the presence of these well-expressed benches. However, for illustrative purposes, we obtained an estimate of about 0.6 for the slope of the linear segment of the plot in the interval $d = 4$–90 yr. This value cannot be used as an estimate of the correlation dimension D_{cw} for the reasons discussed above. Moreover, the plot includes other nearly linear segments with slopes of more than unity. Thus, correct estimation of D_{cw} from the plot $\log C_w(\log d)$ is impossible.

On the contrary, the spectral interval plotted in Fig. 5b for periods longer than 4 yr (in the frequency range $f = 0.011$–0.25 yr^{-1}) is basically suitable for the estimation of the dimension by formula (8a), particularly taking into account that the plot is nearly linear here. The working range cannot be broadened toward higher frequencies because a kink is present near a period of 3–4 yr in the plot of the integral of the spectrum $\log U(\log f)$ (probably, this feature is evidence of the same anomaly as in the plot $C_w(d)$). Table 1 presents results of detailed analysis in the range $f = 0.01$–0.25 yr^{-1}. First of all, we should note that the estimate of D_{sw} is 0.80, and the inequality $D_{sw} < 1$ is significant at the level $Q = 2.5\%$. This indicates that the original idea of multiscale grouping is generally valid for the seismic moment efficiency function.

It is important to clarify whether the deviation of D_{sw} from unity is related to the order grouping or it can be well accounted for by the ordinary long-term grouping. Above, the parameters $D_{cw(\mathrm{RO})}$ and D_{sw1} were proposed for this purpose. The estimate of $D_{sw(\mathrm{RO})}$ is 0.82 and its deviation from unity is significant at a level of 5%.

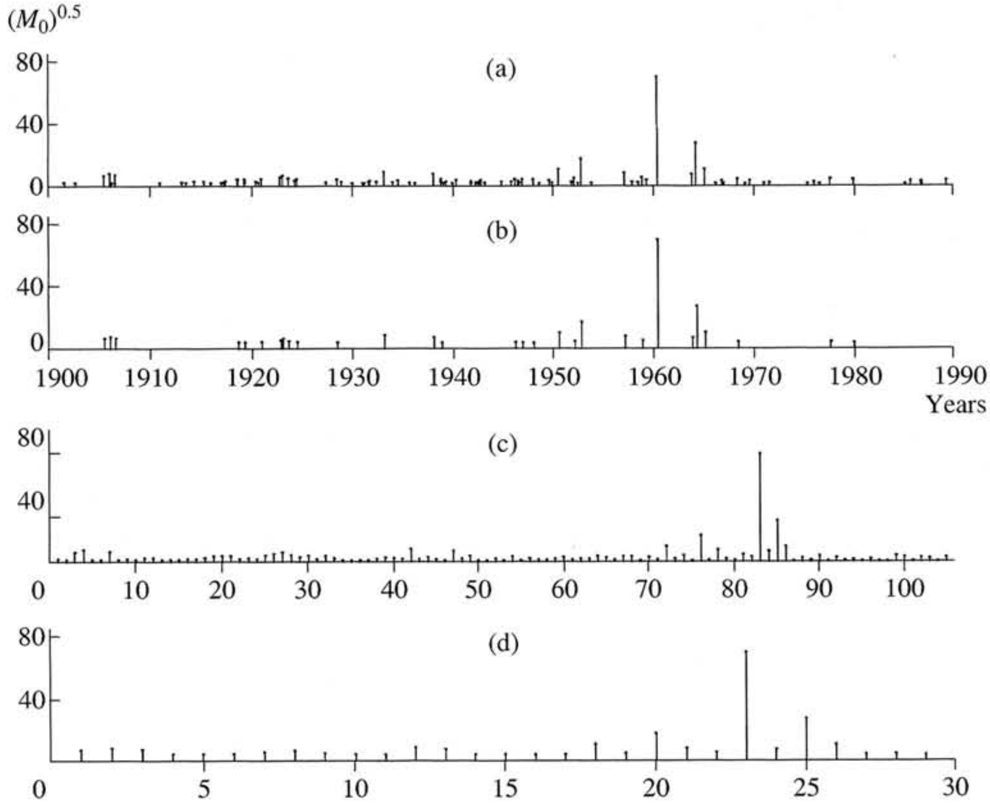

Fig. 4. Sequence of the strongest earthquakes from the Abe catalog as a function of time (a, b) and order number (c, d) for the lower thresholds $M_w = 7.8$ (a, c) and $M_w = 8.21$ (b, d). For clarity, the events are specified by the value $M_0^{1/2}$. Note the clusters of 1905–1906, 1918–1925, 1950–1952, 1957–1960, and 1963–1965, as well as the supercluster of 1950–1965. Also noticeable are the seismicity drops of 1907–1910, 1954–1956, and 1961–1962. Plots (a) and (b) clearly show the ordinary grouping (close groups separated by gaps). Plots (c) and (d) clearly show the order grouping (the strongest events gravitate toward each other and are "poorly mixed" with weaker events).

The parameter D_{sw} was estimated in the spectral window $f = 0.011$–0.7 yr^{-1}, inside which the spectrum is nearly linear. (On the d^* scale, the aforementioned abrupt bench of the plot $C_w(d)$ is preserved but shifts to the point $d^* = 1.3$ yr, so that the choice of upper bound $f^* = 0.7$ yr^{-1} is quite acceptable.) Since a rather convincing significance level was attained in this case, we briefly illustrate our calculations. Direct estimation of the slope from 12 spectral points in the interval 0.011–0.7 yr^{-1} yielded $D_{sw}^{(obs)} = 0.865$. The D_{sw1} average over 1000 pseudocatalogs is $D_{sw1}^{(synth)} = 1.010 \pm 0.056$. This

Table 1. Estimates of grouping parameters obtained by various variants of the analysis of the Abe catalog

M_w threshold (n)	$D_{sw} \pm \sigma$ Q	$D_{sw(RO)} \pm \sigma$ Q	$D_{sw1} \pm \sigma$ Q	$D_{sw(RT)} \pm \sigma$ Q	$D_s \pm \sigma$ Q
7.2	**0.81 ± 0.10**	**0.82 ± 0.11**	**0.86 ± 0.06**	**0.90 ± 0.02**	**≈0.3 ± 0.4**
(482)	**2.5%**	**5%**	**0.25%**	**<0.1%**	**1%**
7.8	**0.79 ± 0.09**	**0.82 ± 0.11**	**0.85 ± 0.05**	**0.91 ± 0.05**	**≈0.2 ± 0.5**
(105)	**0.5%**	**5%**	**0.25%**	**5%**	**1%**
8.21	**0.80 ± 0.09**	**0.80 ± 0.09**	**0.80 ± 0.08**	0.95 ± 0.05	0.77 ± 0.5
(29)	**1%**	**0.5%**	**0.1%**	15%	>20%
7.8	0.73 ± 0.3	0.89 ± 0.3	**0.66 ± 0.24**	1.06 ± 0.08	**≈0.2 ± 0.5**
(104*)	20%	>20%	**5%**	–	**1%**

Notes: Boldfaced are values with a significance level of 5% or less.
* Catalog with the eliminated strongest event of 1960.

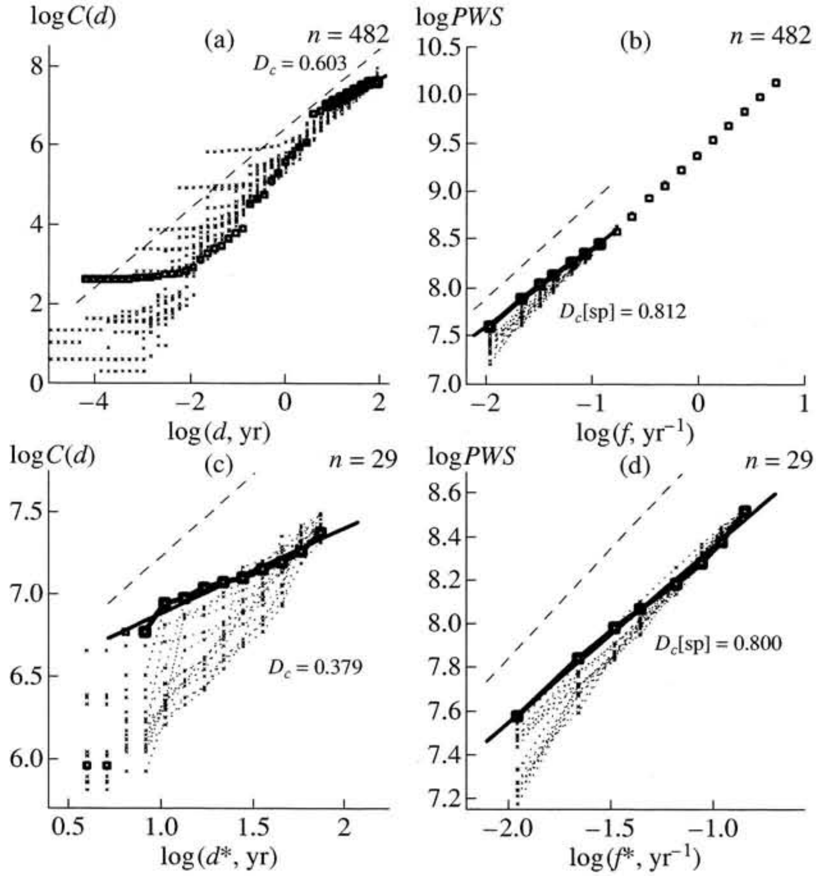

Fig. 5. The correlation integrals $C_w(d)$ (a) and $C_w(d^*)$ (c) and the integrals of the power spectra $U(f)$ (b) and $U(f^*)$ (d) from data of the Abe catalog: (a, b) at $M_w > 7.2$ ($n = 482$); (c, d) at $M_w > 8.21$ ($n = 29$). The estimates of D_{cw} and D_{sw} for plots (a) and (b) are obtained with delays (d) and periods ($1/f$) longer than 4 yr, and the estimates of D_{cw1} and D_{sw1} for plots (c) and (d) are obtained with delays (d^*) and periods ($1/f^*$) longer than 1.3 yr. The slope of the plot $\log C_w(\log d)$ in case (a) in the interval $d = 0.01$–1 yr is noticeably higher than unity, implying the absence of fractal behavior in this interval; therefore, only the estimate of D_{sw} from spectrum (b) ($n = 482$) should be regarded as correct. The estimate of D_{cw1} from spectrum (d) ($n = 29$) is also correct, whereas the reliability of the estimate of D_{cw1} from (c) is limited. The notation is the same as in Fig. 2.

yields $D_{sw1}^{(p)} = 0.855$. The rounded interval estimate is $D_{sw1} = 0.86 \pm 0.06$. Furthermore, 20 of 10 000 pseudocatalog values of $D_{sw1}^{(\text{synth})}$ are smaller than 0.865, which yields the initial estimate $Q = 20/10\,000 = 0.0020$. With regard for the statistical indeterminacy, we have $Q_{\text{mod}} = 0.0020 \times (1 + 20^{-0.5}) = 0.00247$. Rounding off upward, we obtain a final estimate of the significance level for the inequality $D_{sw1} < 1$: $Q = 0.25\%$.

Incidentally, we obtained estimates for the ordinary grouping that are given in the table. An unexpectedly small value of D_s is noteworthy. It reflects a real tendency but has low accuracy because the corresponding spectral segment is appreciably nonlinear. (This does not contradict the estimate $D_c = 0.98$ from Fig. 1: the theory does not guarantee the coincidence of dimension estimates in time and frequency domains in cases deviating from pure, frequency-unbounded self-similarity.) Overall, we may state that both types of grouping contribute to the deviation of D_{sw} from unity. Comparison of the estimates $D_{sw(\text{RO})} = 0.82$ and $D_{sw(\text{RT})} = 0.90$ with the main estimate $D_{sw} = 0.80$ suggests that the contribution of the order grouping is more significant.

Similar analysis was applied to truncated variants of the aftershock-free Abe catalog that included events with $M_w > 7.8$ (105 events) and with $M_w > 8.21$ (29). In the first variant, numerical estimates of D_{sw}, D_c, and the significance level remained virtually the same. However, much smaller estimates of these parameters were obtained for the catalog of 29 events. This is opposite to the usual behavior, when the significance decreases with decreasing amount of data, and implies that the order grouping is best expressed in sets consisting of a small number of the strongest events.

For illustration, Fig. 5 plots estimates of D_{sw} and D_{sw1} obtained for this case. The data are seen to be well consistent with power-law dependences, suggesting

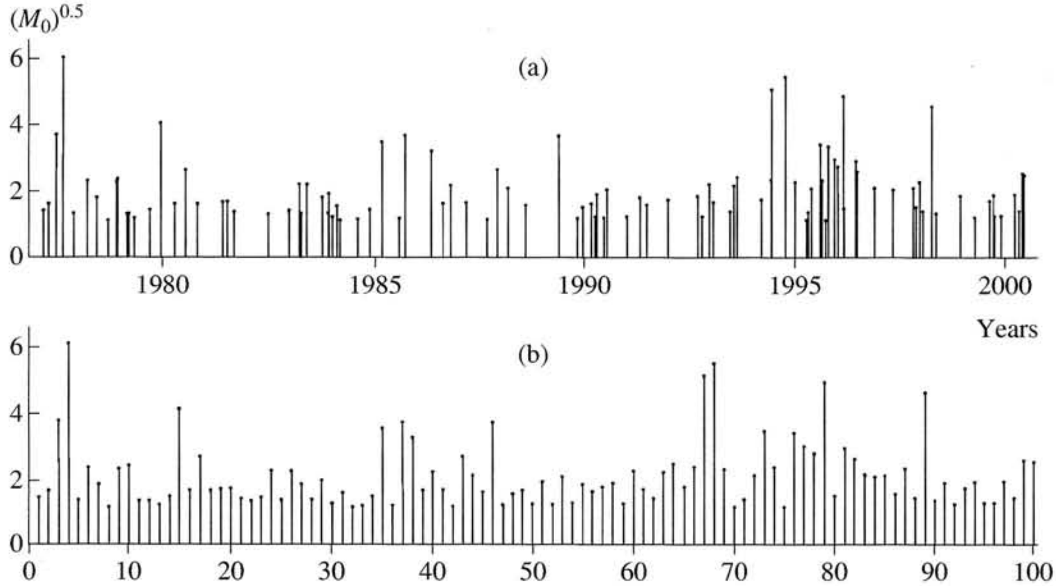

Fig. 6. Sequence of the strongest earthquakes of 1977–2000 as a function of time (a) and order number (b) with the lower threshold $M_w = 7.35$. Values of $M_0^{1/2}$ (M_0 in units of 10^{27} dyn cm) are plotted on the ordinate axes. Plot (a) displays moderately expressed ordinary grouping (concentration intervals on the time axis). Plot (b) displays order grouping (strong events gravitate toward each other).

that the idea of multiscale grouping is fairly plausible and additionally supporting the reliability of the D_{sw} estimate. The picture in the time domain is even more impressive: the estimated value of D_{cw1} is 0.37 ± 0.25 and the plot $\log C_w(\log d^*)$ is nearly linear. However, noticeable deviations from linearity at small d^* make the numerical results somewhat less reliable. (Note that the interval of anomalous behavior at $d = 0.01$–1 yr visible in Fig. 1a is virtually absent in the variant of the truncated catalog.) Thus, at a qualitative level, estimates obtained in the time domain support the inferences derived from the spectral analysis. To illustrate the general stability of the results, we removed the strongest event of 1960 from the list of 105 events with $M_w > 7.8$. Processing of the catalog of the remaining 104 events (Table 1) yielded qualitatively the same results as in the former case; in particular, the significance level for the parameter D_{sw1} amounts to 5%.

ANALYSIS OF THE HARVARD CATALOG

Figure 6 presents the time variation plot for the 100 strongest events from the Harvard catalog. Both ordinary and order groupings are noticeable at a qualitative level, although they are not so evident as in the case of the Abe catalog. Figure 7 presents the plots $\log C_w(\log d)$ and $\log U(\log f)$ for the complete catalog (2573 events with $M_w \geq 6.0$) and the plots $\log C_w(\log d^*)$ and $\log U(\log f^*)$ for 100 events with $M_w \geq 7.35$. Grouping patterns are clearly seen in the plots $\log U(\log f)$ and poorly recognizable in the plots $\log C_w(\log d)$. The dependence $\log U(\log f)$ exhibits a kink in the interval $f = 2$–3 yr^{-1} (Fig. 7b). In order to obtain correct estimates, all parameters of the slope were determined for delays d, d^* and periods $1/f$, $1/f^*$ longer than 0.5 yr. Numerical estimates of the dimension are given in Table 2; estimates of D_{cw} and D_{sw} are also presented in the table for catalogs containing 2573, 406, 100, and 29 events (with different M_w thresholds). Self-similar variations in the efficiency function are observable in both parameters D_{cw} and D_{sw} but best resolved in D_{sw}. The D_{cw} values are significantly smaller than unity only for truncated catalogs. The order grouping is unobservable for the catalog of 28 events but, for other catalogs, it is clearly observed in the parameter $D_{sw(RO)}$ and is somewhat less distinct in the parameters $D_{cw(RO)}$ and D_{sw1}. The situation with the parameters of ordinary grouping is not so clear but, even in this case, the data point to a noticeable grouping, at least for the strongest earthquakes. As noted above, values of the parameter D_c can be somewhat distorted (overestimated), so that its behavior is not unexpected. Overall, the Harvard catalog data qualitatively confirm the conclusions derived from the Abe catalog. However, both types of grouping are expressed to an appreciably smaller degree.

DISCUSSION

Our analysis showed that both self-similar behavior of the efficiency function and order grouping are clearly expressed in both studied catalogs on large time scales. Ordinary grouping is also present, but this is not sur-

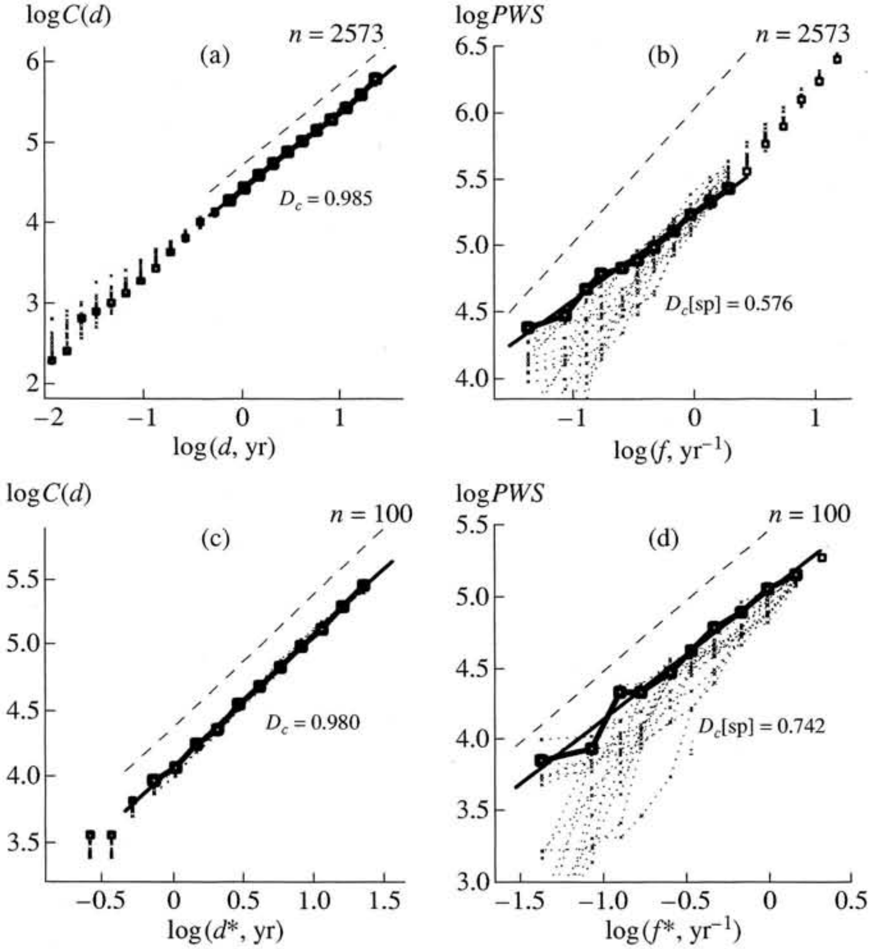

Fig. 7. The correlation integrals $C_W(d)$ (a) and $C_W(d^*)$ (c) and the integrals of the power spectra $U(f)$ (b) and $U(f^*)$ (d) from data of the Harvard catalog: (a, b) at $M_W > 6.0$ ($n = 2573$); (c, d) at $M_W > 7.35$ ($n = 100$).

prising. The behavior of the Abe catalog exhibits better resolved patterns. We may state that the period 1980–1995, which is distinguished by a seismic moment release rate that is very low for the 20th century, is also anomalous with respect to grouping tendencies.

It is noteworthy that numerical estimates of the correlation dimensions of the types D_c and D_s appreciably diverge. Thus, all estimates of the type D_c for the Harvard catalog are much closer to unity than estimates of the type D_s, whereas an inverse situation is observed for the Abe catalog (Fig. 5). Apparently, the main factor responsible for these divergences is a significant deviation of real data from an ideal self-similar pattern. This issue requires a more detailed examination in the future. As yet, it is clear that more reliable results are obtained by the spectral approach, in which all short-term effects of the aftershock or quiescence types are suppressed automatically.

G.M. Molchan (personal communication) noted that, if data are generated by scheme B, the following multiplicative relation should be valid:

$$D_{cw} = D_{cw1}D_c. \quad (12)$$

Inspection of Tables 1 and 2 shows that our results agree, in general, with this relation (with regard for wide intervals used for estimation). In some cases, it is satisfied with good accuracy; thus, according to Table 2 for $N = 100$, $D_{sw} = 0.52$ and $D_{sw1}D_s = 0.74 \times 0.76 = 0.56$. The same is true of the estimates presented in Fig. 1c, although the self-similarity of the artificial data is doubtful.

The discovery of well-expressed ordinary grouping in the Abe catalog is consistent with results reported in [Ogata and Abe, 1991; Kagan and Jackson, 1991]. We did not detect the presence of ordinary grouping in the Harvard catalog for $M_w \geq 6.0$. On the whole, this agrees with the results derived by Kagan and Jackson [1991] from the Harvard catalog for the period 1984–1988 with the suppressed effect of aftershocks: for the bulk

Table 2. Estimates of grouping parameters obtained by various variants of the analysis of the Harvard catalog

Parameter	M_w threshold (n)	$D_{cw}/D_{sw} \pm \sigma$ Q	$D_{cw}/D_{sw(\mathrm{RO})} \pm \sigma$ Q	$D_{cw1}/D_{sw1} \pm \sigma$ Q	$D_{cw}/D_{sw(\mathrm{RT})} \pm \sigma$ Q	$D_c/D_s \pm \sigma$ Q
D_{cw}	6.0	0.985 ± 0.02	0.98 ± 0.16	*0.99 ± 0.02*	*1.00 ± 0.007*	*1.00 ± 0.001*
	(2572)	20%	10%	*>20%*	–	–
D_{sw}	6.0	**0.58 ± 0.28**	**0.48 ± 0.25**	**0.49 ± 0.4**	0.98 ± 0.03	1.02 ± 0.2
	(2572)	**5%**	**5%**	**5%**	>20%	–
D_{cw}	6.8	0.97 ± 0.024	0.97 ± 0.020	0.98 ± 0.021	*0.99 ± 0.02*	*1.01 ± 0.05*
	(406)	20%	10%	20%	*>20%*	–
D_{sw}	6.8	**0.53 ± 0.27**	**0.57 ± 0.27**	0.64 ± 0.3	**0.81 ± 0.10**	0.74 ± 0.27
	(406)	**2.5%**	**2.5%**	10%	**2.5%**	20%
D_{cw}	7.35	0.95 ± 0.04	**0.94 ± 0.03**	0.98 ± 0.02	0.97 ± 0.03	*1.01 ± 0.02*
	(100)	10%	**5%**	20%	20%	–
D_{sw}	7.35	**0.52 ± 0.27**	**0.62 ± 0.25**	0.74 ± 0.3	**0.64 ± 0.20**	0.76 ± 0.26
	(100)	**2.5%**	**2.5%**	25%	**2.5%**	20%
D_{cw}	7.75	**0.82 ± 0.08**	*0.99 ± 0.07*	*1.05 ± 0.04*	**0.81 ± 0.07**	**0.83 ± 0.06**
	(28)	**2.5%**	*>20%*	–	**0.5%**	**1%**
D_{sw}	7.75	**0.60 ± 0.26**	*0.96 ± 0.10*	*1.95 ± 0.4*	**0.51 ± 0.27**	**0.56 ± 0.27**
	(28)	**5%**	*>20%*	–	**2.5%**	**5%**

Note: Boldfaced are values with a significance level of 5% or less, and italics mean the absence of grouping.

of pairs of events, they obtained $D_c = 1.00$ for epicentral distances of more than 3276 km and $D_c = 0.95$ for the interval 819–3276 km, whereas our estimate is $D_c = 1.00$ for all epicentral distances. However, with higher magnitude thresholds, our estimates from the parameter $D_{sw(\mathrm{RT})}$ definitely suggest that the ordinary grouping is actually present. The discovery of the order grouping in the Abe catalog agrees with the results of Ogata and Abe [1991]. It is worth mentioning here the following peculiar fact: for small delays (0.1–30 days) in a catalog with noneliminated aftershocks, we obtained a tentative estimate of $D_c \approx 0.67$ for both the Abe catalog (see Fig. 3a) and the Harvard catalog (not shown).

An important methodological aspect of this work is the demonstration of the fact that long-term variations in seismicity are well observed in the spectral representation. Ogata and Abe applied this approach to the flow density and the sequence of magnitudes, but the spectral properties of long-term variations in the efficiency function are studied here for the first time. We remind the reader that a process characterized by a power spectrum of the type $1/f^{\alpha}$, where the exponent α lies in the range 0.5–1.5 (or in a slightly wider range), is described as flicker noise. Since $\alpha = 1 - D_s$, the values of D_s and D_{sw} presented in Tables 1 and 2 imply that, given n = 100–400, $p_1(t)$ and $p(t)$ are pulsed flicker noises with the following spectral parameters: (1) $\alpha \approx 0.75$ for $p_1(t)$ and $\alpha \approx 0.2$ for $p(t)$ from the Abe catalog and (2) $\alpha \approx 0.25$ for $p_1(t)$ and $\alpha \approx 0.45$ for $p(t)$ from the Harvard catalog. The dependences of the seismic moment on the number of the event determined from the same catalogs have the form of flicker noises with $\alpha \approx 0.15$–0.3 and $\alpha \approx 0.15$–0.5, respectively.

We should make the following comment on the meaning of the estimates D_{cw1} and D_{sw1}. One might think that scheme 2, using the number of an event as an argument of the type "fictitious time," is a purely artificial procedure. However, following the arguments of Mandelbrot [1982] concerning the "devil's staircase," the "inner clock" of the seismotectonic process runs and measures a "modified time" only when a seismotectonic deformation occurs, i.e., during an earthquake; the rest of the time, this clock is stopped. In this context, the presence of the self-similar order grouping means that the seismotectonic deformation process possesses fractal properties on a modified time scale. On the other hand, the presence of the self-similar ordinary grouping means that the modified time itself flows "in a fractal manner," i.e., only at the time moments of the ordinary time that form a fractal point set (fractal dust).

CONCLUSIONS

(1) Methods for the recognition and analysis of the self-similar order grouping in catalogs of events of various weights, the self-similar behavior of the efficiency function, and the contributions of the ordinary and order grouping to this behavior are developed. A spectral approach to these problems is shown to be effective.

(2) Anomalies in the time structure are discovered for the Abe catalog at delays of up to 4 yr and for the

Harvard catalog at delays of up to half a year. The anomalies are of the types of periodicity and "dead time" (quiescence) and contradict the idea of self-similar behavior. Due to the presence of such anomalies, analysis of self-similar behavior is effective only in a limited range of large time scales (low frequencies).

(3) The properties of the self-similar order grouping of earthquakes, a phenomenon previously discovered by Ogata and Abe, are studied in detail and are significantly refined. This grouping is identified in two world catalogs with eliminated aftershocks in the magnitude interval $M_w = 7$–9.5 in a limited range of large time scales.

(4) Self-similar variations in the global generation rate of the seismic moment, or the efficiency function, are discovered in the same ranges of magnitudes and time scales with eliminated aftershocks. The efficiency function behaves as flicker noise with a power spectrum of the type $1/f^{0.25-0.45}$.

(5) Under the same limitations, the earthquake flow density with eliminated aftershocks behaves as pulsed flicker noise with a power spectrum of the type $1/f^{0.2-0.8}$.

(6) Under the same limitations, the seismic moment as a function of the earthquake number in a catalog behaves as pulsed flicker noise with a power spectrum of the type $1/f^{0.15-0.5}$.

ACKNOWLEDGMENTS

I am grateful to G.M. Molchan for helpful discussion. This work was supported by the Russian Foundation for Basic Research, project no. 03-05-64459.

APPENDIX

Algorithm for Joining Swarms in a Catalog

We have to solve the problem of eliminating groups of events close in space and time through the replacement of such a group by a "joint event." For this purpose, repeated attempts are made to discover close pairs ("neighbors") and to combine each pair into an (joint) event. The problem is considered as solved when close pairs are absent. The joining procedure preserves the total seismic moment. A pair is considered as close if its events are separated by a distance no greater than (or of the same order as) the typical size of a source and by a time interval d of the same order as the duration of an aftershock swarm.

The outer loop of the algorithm repeatedly processes the current state of a catalog by an inner loop procedure until no pairs remain to be joined.

(i) The inner loop consists in testing a current (in time) event of the catalog (a "basic" event with a number i and a time t_i). All potential neighbors postdating the basic event are then considered in the time window $[t_i, t_{\rm end}]$.

(ii) The further procedure consist in testing whether a current potential neighbor with a number $j > i$ is an actual neighbor, i.e., whether it meets the conditions $r_{ij} < dr$ and $t_j - t_i = d_{ij} < dt$, where r_{ij} is the distance between the epicenters of the events and $dr = dr(M_i, M_j)$ and $dt = dt(M_i, M_j)$ are thresholds for r and d. The M dependences of dr and dt are $dr = C_R \max(L(M_i), L(M_j))$ and $dt = C_T(dr/L(8))^{0.5}$, where $L(M) = 10^{0.5M-1.8}$ km, $C_R = 0.5$–1.5 (adjustable), and $C_T = 0.25$ yr. Further, if the events i and j are neighbors, the joining procedure is applied: the value $M = M_w$ of the joint event is determined through the sum of M_0 values of the components i and j; an event greater in magnitude is determined and, if, for example, M_i is greater than M_j, the record number and the time of the event i are assigned to the joint event, the "summed" magnitude is introduced into the record of the latter, and the record of the event j is deleted, after which procedure (ii) is repeated. Otherwise, the record number and the time of the event j are assigned to the joint event, the "summed" magnitude is introduced into the record of the latter, and the record of the event i is deleted, after which procedure (i) is repeated.

REFERENCES

1. K. Abe, "Magnitudes of Large Shallow Earthquakes from 1904 to 1980," Phys. Earth Planet. Inter. **27**, 72–92 (1981).
2. K. Abe, "Complements to Magnitudes, 1904 to 1980," Phys. Earth Planet. Inter. **34**, 13–23 (1984).
3. K. Abe and S. Noguchi, "Determination of Magnitude for Large Shallow Earthquakes, 1898–1917," Phys. Earth Planet. Inter. **32**, 45–59 (1983a).
4. K. Abe and S. Noguchi, "Revision of Magnitudes of Large Shallow Earthquakes, 1897–1912," Phys. Earth Planet. Inter. **33**, 1–11 (1983b).
5. S. V. Bozhokin and D. A. Parshin, *Fractals and Multifractals* (NITs Regulyarn. Khaotich. Dinamika, Izhevsk, 2001) [in Russian].
6. Bulletin of the International Seismological Centre (Thatcham, 2001, URL: http://www.isc.ac.uk).
7. G. Ekström and A. M. Dziewonski, "Evidence of Bias in the Estimation of Earthquake Size," Nature **332**, 319–323 (1988).
8. K. Fujita, *Earthquake Magnitudes* (2001, URL: http://www.msu.edu/~fujita/earthquake/bigquake.html).
9. A. A. Gusev, "A Nomogram for the Identification of Gorups of Earthquakes," Geol. Geofiz, No. 3, 36–43 (1971).
10. A. A. Gusev, V. V. Ponomareva, O. A. Braitseva, *et al.*, "Great Explosive Eruptions on Kamchatka during the Last 10000 Years: Self-Similar Irregularity of the Output of Volcanic Products," J. Geophys. Res. **108** (B2), 2126 (2003).
11. *Harvard CMT Catalog* (2001, URL: http://www.seismology.harvard.edu/CMTsearch.html).

12. T. Hirabayashi, K. Ito, and T. Yoshii, "Multifractal Analysis of Earthquakes," Pure Appl. Geophys. **138**, 591–610 (1992).

13. Y. Y. Kagan and D. Jackson, "Long-Term Earthquake Clustering," Geophys. J. Int. **104**, 117–133 (1991).

14. B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, New York, 1982; Inst. Komp'yuternykh Issled., Moscow, 2002).

15. Y. Ogata and K. Abe, "Some Statistical Features of the Long-Term Variation of the Global and Regional Seismic Activity," Int. Statist. Rev. **59** (2), 139–161 (1991).

16. J. F. Pacheco and L. R. Sykes, "Seismic Moment Catalog of Large Shallow Earthquakes, 1990 to 1989," Bull. Seismol. Soc. Am. **82**, 1306–1349 (1992).

17. O. J. Perez and C. H. Scholz, "Heterogeneities of the Instrumental Seisminity Catalog 1904–1980 for Strong Shallow Earthquakes," Bull. Seismol. Soc. Am. **74**, 669–686 (1984).

18. V. F. Pisarenko and D. V. Pisarenko, "Spectral Properties of Multifractal Measures," Phys. Lett. A **153** (4–5), 169–172 (1991).

19. A. G. Prozorov, "Clustering Characteristics of World Earthquakes," in *Earthquake Prediction and Study of the Earth's Structure (Computational Seismology, No. 15)* (Nauka, Moscow, 1982), pp. 18–26 [in Russian].

20. L. N. Rykunov, V. B. Smirnov, Yu. O. Starovoi, and O. S. Chubarova, "Seismic Radiation Self-Similarity in Time," Dokl. Akad. Nauk SSSR **297** (6), 1337–1341 (1987).

21. R. F. Smalley, J. J.-L. Chatelain, D. L. Turcotte, and R. Prevot, "A Fractal Approach to the Clustering of Aftershocks: Application to the Seismicity of the New Hebrides," Bull. Seismol. Soc. Am. **77** (4), 1368–1381 (1987).

22. I. R. Stakhovsky, "Time and Space–Time Multiscaling Analysis of Seismicity Preceding the Racha Earthquake," Fiz. Zemli, No. 4, 41–47 (2000) [Izvestiya, Phys. Solid Earth **36**, 298–304 (2000)].