Science of Earthquake Prediction: SCEC-USGS-CGS Workshop

February 20, 2004

Contents

1	Intr	roduction	3										
2 Short-Term Prediction													
	2.1	Formulation of the problem	Z										
	2.2	Regions considered	4										
		2.2.1 Central and Southern California	Z										
		2.2.2 Japan	8										
		2.2.3 Eastern Mediterranean	8										
	2.3	Methodology: Reverse Detection of Precursors (RDP)	8										
		2.3.1 Short-Term Premonitory Pattern: Chains	(
		2.3.2 Intermediate-term patterns	12										
		2.3.3 Formulation of Prediction Algorithm	14										
	2.4	Performance	1'										
		2.4.1 California	1'										
		2.4.2 Southern California	1′										
		2.4.3 Japan	1′										
		2.4.4 Eastern Mediterranean	20										
	2.5	Stability tests	20										
		2.5.1 Variation of chain parameters	20										
		2.5.2 "Historical" experiment	23										
3	Adv	dvance Predictions											
	3.1	Advance prediction of Hokkaido earthquake	20										
	3.2	Advance prediction of San Simeon earthquake	20										
	3.3	Current alarm: Southern California	20										

Α	$\mathbf{W}\mathbf{h}$	What is Earthquake Prediction?											
A.1 Formulation of the Problem													
	A.2	Data		34									
	al Scheme of Data Analysis	35											
		A.3.1	Earthquakes to be Predicted	35									
		A.3.2	Premonitory Seismicity Patterns	35									
		A.3.3	Reduction to Pattern Recognition	35									
		A.3.4	Earthquake sequence.	36									
		A.3.5	Premonitory patterns	36									
		A.3.6	Prediction algorithms	37									
в	Disa	aster n	nanagement point of view	37									
\mathbf{C}	Intermediate-Term Earthquake Prediction												
	C.1	Valida	ted Prediction Algorithms	38									
		C.1.1	Algorithm M8	38									
		C.1.2	Algorithm MSc or "The Mendocino Scenario"	41									
		C.1.3	Prediction of Largest Earthquakes, $M \ge 8. \ldots \ldots \ldots$	42									
		C.1.4	Algorithm SSE (Second Strong Earthquake)	44									
		C.1.5	Algorithm CN	48									
D	Pre	diction	Validation	51									
	D.1	Error 1	Diagram	52									
	D.2	The O	ptimal Prediction Strategy	55									
	D.3	Statist	cical Problems	56									
		D.3.1	The Performance of Prediction Algorithms	56									
		D.3.2	Estimation of (n, τ)	57									
		D.3.3	Estimation of $r(t)$	58									

1 Introduction

Recent advance predictions of Tokachi-oki earthquake, Hokkaido, Japan (September 25, 2003, M8.3) and San Simeon earthquake, CA (December 22, 2003, M6.5) made by the international group led by Prof. Vladimir Keilis-Borok present a perfect opportunity to revisit the problem of earthquake prediction and rise specific questions like *What is the progress made recently in short-term earthquake prediction? How reliable is the method used for the recent predictions? How can the geophysical community join the efforts in improving existing understanding of premonitory processes?*

The SCEC Workshop is targeted first of all at a new prediction recently issued by the same algorithm for a region in Southern California (see Sect. 3.3). Even though there is a not negligible chance that this prediction is a false alarm, the odds of having the next big one within this territory seem to be much higher than random. There is a realistic (and quite unique) possibility to study the place of a coming earthquake months prior to its occurrence, paying not too high a price in case if this prediction will turn to be a false one. It would be very important not to miss the chance to ambush a coming earthquake by joint efforts using different methods and ideas from the whole geophysical community.

Presented materials are aimed to provide necessary methodological and technical detail of the ongoing experiment in short-term prediction. They discuss *Reverse Detection of Precursors* — the methodology which is the core of the recent advances in short-term prediction, — give an overview of the advance predictions made so far, and describe tests of stability of prediction algorithm. The essential techniques and methods of prediction evaluation and the existing intermediate-term prediction algorithms are described as well.

Disclaimer: The methods and techniques that have led to the recent advance predictions are based on the experience of some 20 years of broad multidisciplinary research which merged geodynamics, non-linear dynamics, statistical physics, and pattern recognition. These materials is NOT a comprehensive review NOR a research paper, they merely a set of technical notes targeted at collecting together what have been done, and helping to overview the progress made so far. Many important issues are left aside as well as many questions are not discussed in the full detail required from a scientific publication.

2 Short-Term Prediction

Here we describe the general methods and specific techniques that lead to short-term earthquake prediction as well as provide the detailed information about the recent advance predictions.

2.1 Formulation of the problem

Commonly known are five major stages of earthquake prediction. The first, background, stage provides the maps with territorial distribution of maximal possible magnitude and recurrence time of destructive earthquakes of different magnitude. The four subsequent stages, fuzzily divided, include the time-prediction; they differ in characteristic time intervals for which an alarm is declared: *long-term* stage corresponds to alarms of tens of years, *intermediate-term* stage to years, *short-term* to months or weeks, and *immediate* to days or less (see also Sect. A.1, Table 6.) Accordingly, the problem of short-term prediction is formulated here as follows:

To outline a spatial region within which an earthquake with magnitude $M_{low} < M < M_{up}$ is expected during the next k months.

Predictions described below use k = 9 months.

2.2 Regions considered

We have considered three regions: California, Japan, and Eastern Mediterranean and used an appropriate earthquake catalog for each region. The detailed description follows.

2.2.1 Central and Southern California

Catalog: Advanced National Seismic System [2].

Region: Southern California is defined as a rectangle within 31.5°N-36°N, 120°W-114°W, and Central California defined as the polygon with vertices 41°N, 123°W; 36°N, 117°W; 33°N, 120°W; 38°N, 126°W. The union of Central and Southern California regions is shown in Fig. 1.

Magnitude: as given in ANSS (one value)

Preprocessing: Aftershocks are removed using the space-time windowing [Aft] with parameters given in Table1. Note that magnitude of aftershocks in this method is by definition smaller than the magnitude of the main shock. No other preprocessing.

Target earthquakes: are those with $M \ge 6.4$.



Figure 1: California region. Stars mark large earthquakes, targeted for retrospective prediction. Dots show background seismicity; aftershocks are excluded.



Figure 2: Japan region. Small stars mark large earthquakes, targeted for retrospective prediction. Dots show background seismicity with magnitude $M \geq 5$; aftershocks are excluded. Large star is the epicenter of the M8.1 Hokkaido (Tokachi-oki) earthquake, predicted in advance by the precursory chain shown in Fig. 11



Figure 3: Eastern Mediterranean region. Stars mark large earthquakes, targeted for retrospective prediction. Dots show background seismicity; aftershocks are excluded.

2.2.2 Japan

Catalog: Japan Meteorological Agency.

Region: Polygon with vertices 23°N, 135°E; 31°N, 124°E; 50°N, 145°E; 42°N, 156°E, see Fig. 2

Magnitude: as given in JMA catalog (one value). Note that recently magnitudes were revised for the entire JMA catalog. The new catalog has not been analyzed.

Preprocessing: Aftershocks are removed using the space-time windowing [Aft] with parameters given in Table1. No other preprocessing.

Target earthquakes: are those with $M \ge 7.0$.

2.2.3 Eastern Mediterranean

Catalog: The catalog of Geophysical Institute of Israel ([3].

Region: Rectangle with vertices 27°N-36°N, 32°E-38°E, see Fig. 3

Magnitude: Maximal magnitude from reported in GII catalog.

Preprocessing: Aftershocks are removed using the space-time windowing [Aft] with parameters given in Table1. No other preprocessing.

Target earthquakes: are those with $M \ge 6.5$.

2.3 Methodology: Reverse Detection of Precursors (RDP)

RDP methodology is based on the concept of self-organization of the lithosphere culminated by a strong earthquake [Aki, 2003; Bird, 1998; Blanter and Schnirman, 1997; Bowman et al., 1998; Gabrielov et al., 2000; Jin et al., 2003; Keilis-Borok, 1996, 2002; Keilis-Borok and Soloviev, 2003; Newman et al., 1994; Rundle et al., 2000; Scholz, 1990; Sornette, 2000; Turcotte, 1997; Zaliapin et al., 2003]. Being a part of geodynamics, this process is not localized but extends over large time and space, evolving in multiple scales. Among its various manifestations is the sequence of *premonitory seismicity patterns* the spatio-temporal patterns of seismicity that emerge as a strong earthquake approaches.

The well-known difficulty in working with short-term patterns is that they typically produce many false alarms (i.e. not all of them are followed by a strong earthquake). Contrary to intermediate-term prediction, where the rate of false alarms is non-zero yet acceptable (1 false alarm per 1 correct alarm in the worst cases), the rate of false alarms

Magnitude	$R_{\rm aft},{\rm km}$	$T_{\rm aft}$, hours
M < 2.5	20	60
$2.5 \le M < 3.0$	23	120
$3.0 \le M < 3.5$	26	180
$3.5 \le M < 4.0$	30	360
$4.0 \le M < 4.5$	35	720
$4.5 \leq M < 5.0$	40	1440
$5.0 \le M < 5.5$	47	2880
$5.5 \le M < 6.0$	54	5760
$6.0 \leq M < 6.5$	61	11520
$6.5 \le M$	70	23040

Table 1: Windows for identification of aftershocks

for short-term premonitory patterns (precursors) is way higher reaching tens of false alarms per one correct one, and rendering the whole prediction scheme useless. The RDP methodology is targeted at overcoming this obstacle by distinguishing rare correct shortterm alarms among a large number of false ones. The main idea of RDP is that correct short-term patterns should be preceded by intermediate-term ones observed within the same spatial region, while false short-term patterns may happen just randomly. The idea of this methodology is thus to combine numerous short-term patterns with less frequent intermediate-term ones.

Technically, the analysis proceeds as follows (Fig. 4.) First, we detect the "candidates" for the short-term precursors; in our case these are the *chains* of earthquakes capturing the rise of an earthquake correlation range (Sect. 2.3.1). Then we check, for each chain, whether it was preceded by intermediate-term precursors in its close vicinity (Sect. 2.3.2). If the answer is "yes", we regard this chain as a precursory one; in prediction it would start a short-term alarm. If not — the chain is disregarded and alarm is not started. Thus we consider the short-term patterns (chains) first, although they emerge later.

A notable advantage of reverse analysis is that the chain indicates the narrow area where an intermediate-term precursor should be looked for. This makes possible to detect the finely scaled precursors undetectable by the direct analysis.

2.3.1 Short-Term Premonitory Pattern: Chains

The short-term premonitory phenomenon used in our analysis reflects increase in earthquake correlation range. The pattern we use here is a generalization of patterns "ROC" and "Accord" that capture the same phenomenon; these patterns have been found in modeled seismicity [Gabrielov et al., 2000] and then in observations [Shebalin et al., 2000;



Figure 4: Reverse Detection of Precursors (RDP).

Table 2: Parameters of chain definition

Region	M_{\min}	τ_0 , days	r_c , km	С	k_0	l_0, km	γ_0
C California	2.9	30	50	0.35	10	250	0.6
S Califonria	2.9	20	50	0.35	6	175	0.5
Japan	3.6	20	50	0.35	10	350	0.4
E Mediterranian	3.0	40	50	0.35	6	200	0.5

Keilis-Borok et al., 2002]. Qualitatively, this pattern is a dense sequence of medium magnitude earthquake that quickly extends over large distance. We call it pattern "Chain".

Definition. We call two earthquakes *neighbors* if they occurred close in time and space to each other; formally if (i) the distance between their epicenters does not exceed a threshold r_0 ; and (ii) the time interval between them does not exceed a threshold τ_0 . A *chain* is the sequence of earthquakes connected by the following rule: each earthquake has at least one neighbor in that sequence. Clearly, an earthquake can not have neighbors outside the chain it belongs to; thus each pair (r_0, t_0) corresponds to a division of all the earthquakes into non-overlapping clusters; we call them chains. To connect in a chain only the unusually close earthquakes we normalize the threshold r_0 as follows:

$$r_0 = r_c 10^{c(m-2.5)},$$

where m is the magnitude of the smaller earthquake in a pair considered. Let k be the number of earthquakes in a chain thus defined, and l the maximal pairwise distance between their epicenters. We consider only the chains which are numerous and long enough:

$$k \ge k_0, \quad l \ge l_0.$$

We have noticed that correct premonitory chains have a larger number of earthquakes with higher magnitudes. Hence the additional criterion was introduced to define a chain:

$$\gamma := \frac{\# \{\text{earthquakes with } M \ge 3.5 \text{ from the chain}\}}{\# \{\text{earthquakes with } M_{\min} \le M < 3.5 \text{ from the chain}\}} > \gamma_0,$$

where γ_0 is a numerical parameter.

The chains are detected among the earthquakes with $M \ge M_{\min}$.

The need for RDP. Data analysis demonstrated that most of the target earthquakes were preceded within a few months by a chain thusly defined; we found, however, that 80-90% of all chains would give false alarms if used as short-term precursors: they are not followed by a strong earthquake. To reduce this rate we consider the chains in conjunction with intermediate-term precursors using the RDP approach.

2.3.2 Intermediate-term patterns

Here we describe intermediate-term patterns used by the RDP methodology.

R-vicinity of a Chain is an important object in our analysis. It is defines as a spatial *R*-environment of the nearest-neighbor cluster $\Omega = w_p$ that connects the epicenters of the earthquakes that form the chain:

$$\bigcup_p S_R(w_p)$$

Here $S_R(x, y)$ is a circle of radius R centered at (x, y); p defines a parametrization of points from the cluster Ω . The intermediate-term precursors that might precede a chain are looked for within its R-vicinity.

We consider eight intermediate-term patterns; four of them capture premonitory rise of seismic activity, two — rise of clustering, one — rise of correlation range (more specifically — spreading of seismicity over the fault network), and one — transformation of magnitude - frequency (Gutenberg - Richter's) relation.

To detect a pattern P we compute a function $F_P(t)$ defined on the earthquake sequence. Emergence of a pattern at the moment t is captured by condition

$$F_P(t) \ge C_P$$

where C_P is an adjustable threshold. Functions considered here are defined below. Functions are computed in *R*-vicinity of a chain within *T* years preceding it. They are defined in the "event window" [Prozorov and Schreider, 1990; Keilis-Borok et al., 2002] for the sequence of *N* consecutive earthquakes; with sequential indexes (j - N + 1, j - N + 2, ..., j)in the catalog; value of a function $F_P(t)$ is attributed to the moment of the last earthquake in a sequence, $t = t_j$. In some versions of the algorithm and in tests of prediction stability the same functions were defined in a sliding time window of length *s*: [t - s, t]. In this case N = N(t) is the number of events that occurred within this time window.

The catalog cut-off magnitude is determined by fixing the numerical parameter n^* and finding M_{\min} that satisfies the following equation

$$n^* = \#\{\text{earthquakes with } M \ge M_{\min}\}$$

This does not add the adjustable parameters; instead of M_{\min} we now adjust n^* . In all our experiments we took $n^* = 20$ or $n^* = 10$, as suggested by the previous experience in the intermediate-term prediction.

In the equations (1)-(8) below we use the following notations: $m_{1/2}$ is the median magnitude among N consecutive earthquakes; A_r^{\cup} is the area of the union of the circles of radius r centered at N epicenters in the chain, their sequence numbers are (j - N +1, j - N + 2, ..., j); A_r^{\cap} is the area of intersections of the circles of radius r centered at N epicenters in the chain; their sequence numbers are (j - N + 1, j - N + 2, ..., j); m_{kl} is the magnitude of the l-th aftershock of the k-th main shock, as usually [e.g. Kossobokov and Shebalin, 2003], we consider immediate aftershocks within given time and magnitude intervals.

Functions measuring the rise of earthquake activity.

"Activity" measures time necessary to collect N consecutive earthquakes:

$$U(t_j) = \frac{1}{t_j - t_{j-N+1}},\tag{1}$$

where t_i is the time of *i*-th earthquake in the catalog.

"Sigma" estimates the total area of faultbreaks:

$$\Sigma(t_j) = \sum_{k=j-N+1}^{j} 10^{m_k}.$$
(2)

"Acceleration of magnitudes" measures the possible increase in earthquake magnitudes:

$$A_m(t_j) = \frac{2}{N} \left\{ \sum_{k=j-N/2+1}^j m_k - \sum_{k=j-N+1}^{j-N/2} m_k \right\}.$$
 (3)

"Acceleration of number of earthquakes" measures the possible increase in frequency of earthquakes:

$$A_n(t_j) = \frac{2}{N} \left\{ \sum_{k=j-N/2+1}^j \frac{1}{t_k - t_{k-1}} - \sum_{k=j-N+1}^{j-N/2} \frac{1}{t_k - t_{k-1}} \right\}.$$
 (4)

Functions measuring the rise of earthquake clustering.

"Swarm" estimates the clustering of mainshocks:

$$W(t_j) = \frac{A_r^{\cap}}{\pi r^2}.$$
(5)

"b-micro" estimates the clustering of aftershocks:

$$b_{\mu}(t_j) = \sum_{k=j-N+1}^{j} \sum_{l} 10^{m_{kl}}.$$
 (6)

Function "Accord" estimates the rise of earthquake correlation range:

$$A(t_j) = \frac{A_r^{\cup}}{\pi r^2}.$$
(7)

Function "Gamma" estimates the transformation of frequency-size (Gutenberg-Richter) distribution:

$$\gamma(t_j) = \frac{2}{N} \sum_{m_k \ge m_{1/2}} m_k.$$
(8)

2.3.3 Formulation of Prediction Algorithm

In this section we formulate the prediction algorithm based on the RDP methodology described above. We started tests of this algorithm in California, Japan, and Eastern Mediterranean. Note, that the algorithm described below is not final and can be improved as new information is obtained in the course of the ongoing experiment.

The algorithm works in two steps: at the first — short-term — step we determine chains of earthquake following the procedure described in Sect.2.3.1. When the chains are detected, we proceed to the second — intermediate-term — step. At this step we consider the functions described in Sect. 2.3.2 for each chain within its *R*-vicinity and within time *T* prior to the chain. We consider eight variants for each of eight functions, each variant corresponds to a distinct set of numerical parameters: *r* for "Swarm" and "Accord" was kept fixed at 50 km; the pair (R, n^*) takes one of the values (100km, 20), or (50km, 10); *N* is 20 or 50, and *T* is 6 or 24 months. Parameters (R, n^*) , *N*, and *T* are varied independently. Altogether, for each chain we consider $8 \times 8 = 64$ functions.

If the function reaches or exceeds the critical threshold C_P within the time interval T, we say that it "votes" for making the chain precursory. If it is lower than the threshold, we say that it "votes" against making the chain precursory. Finally, we have 64 individual voices V_i , each of which can be "pro" ($V_i = 1$) or "con" ($V_i = 0$). Considering the result Δ of the collective voting:

$$\Delta = \sum_{i} V_i,$$

we finally decide whether we consider the chain precursory or not; this decision is made by comparison the voting outcome Δ with a threshold C.

The important question is how to define the individual critical thresholds C_P and the final voting threshold C. The thresholds C_P are automatically determined for each functional based on its performance within the learning set. The value for C_P for a given function F_P should minimize the sum of rates of prediction errors:

$$\frac{n_{\text{fail}}}{n_{\text{strong}}} + \frac{k_{\text{false}}}{k_{\text{chain}}},$$



Figure 5: Detection of optimal threshold C_p for the function N(t) calculated in sliding time window for Southern California. The values of the function are shown in Fig. 6

where n_{fail} is the number of failures to predict, that is the number target earthquakes not preceded by a precursory chain; n_{strong} is the number of target earthquakes; k_{false} is the number of false alarms, that is the number of precursory chains that do not precede a strong earthquake; k_{chain} is the total number of chains. The detection of the optimal threshold C_P is illustrated in Fig. 5. Here we show the sum of errors for the function N(t)measured in a sliding time window for Southern California. The values of this function within vicinities of 14 chains (7 false and 7 correct) are shown in Fig. 6.

The choice of the threshold C should be made considering the goals of the end-user of the prediction. The ideal prediction, from the common — and very reasonable — view point, should have as small number of false alarms as possible, predicting as many target earthquakes as possible. Making C very high (which makes it hard to consider a chain precursory), we will indeed have a small number of alarms at a price of missing many strong earthquakes, while making C very low (which makes it easy to consider a chain precursory), we will predict most of strong earthquakes, at a price of having a larger number of false alarms. This tradeoff can not be resolved without clear understanding of the goals of the ongoing prediction and without having a list of possible measures that can be undertaken in response to such a prediction. The methods of choosing the right prediction strategy is a well developed part of statistical analysis. Its brief outline is given in Appendix D.

The important moment is the possibility of a chain to grow after it was already recognized as precursory. Let t_e be the moment when the chain reached both required limits: l_0 for the length and k_0 the number of earthquakes (Sect. 2.3.1). If it recognized as precursory, the alarm is triggered for the time interval $(t_e, t_e + \tau)$. The chain might keep growing beyond t_e , accumulating subsequent earthquakes. And it might remain or become precursory until some moment t_e^1 (for example until it ended). In that case the



N(t) ={events within sliding time window}

Figure 6: Function N(t) in a sliding time window during 5 years prior 14 chains in Southern California. Left column — correct chains; right column — false chains. The last 6 months prior to each chain are highlighted. The optimal critical threshold $C_P = 45$ is shown by horizontal lines in each panel. The procedure of detecting this threshold is illustrated in Fig. 5.

alarm is extended to the time $(t_e^1 + \tau)$. We refer to all the moments when chain was precursory but still keep growing as "states" of the chain.

If a target earthquake occurs near a chain, and its epicenter should formally be added to the chain, then we disregard all the states of that chain starting from the moment of the strong earthquake. However, the already declared alarm is not called off in that case.

2.4 Performance

Here we describe the performance of the above algorithm for the three regions considered. The algorithm was applied by the following scheme to each of the regions. The time interval considered is divided into two parts. The beginning part is used for learning (adjustment of the thresholds), then the algorithm is applied to the second part with all the parameters fixed.

2.4.1 California

Performance of the algorithm for California is illustrated in Fig. 7. Parameters of the algorithm are given in the Table 2. The voting threshold is C = 41. The learning period for this region is 1965-1995. During this period 25 chains were detected, and 12 were eliminated by the intermediate-term patterns, according to the RDP methodology. Out of the remaining 13 chains 5 are false alarms, and 8 are correct. Since 1995 4 chains were detected, 2 of them were eliminated by RDP approach, and the remaining two are correct alarms that predict the Hector Mine earthquake, and the San Simeon earthquake. The total spatio-temporal volume of alarms is 9.6%.

2.4.2 Southern California

Performance of the algorithm for Southern California is illustrated in Fig. 8. Parameters of the algorithm are given in the Table 2. The voting threshold is C = 36. The learning period for this region is 1965-1995. During this period 40 chains were detected, and 32 were eliminated by the intermediate-term patterns, according to the RDP methodology. Out of the remaining 8 chains one is false alarm, and 7 are correct, they predict 6 target earthquakes. Since 1995 14 chains were detected, 12 of them were eliminated by RDP approach, and out of the 2 remaining one is correct alarm that predict Hector Mine earthquake, and the last one gives the current alarm. See Sect. 3.3 for more discussion. The total spatio-temporal volume of alarms is 8.2%.

2.4.3 Japan

Performance of the algorithm for Japan is illustrated in Fig. 9. Parameters of the algorithm are given in the Table 2. The voting threshold is C = 36. The learning period for this region is 1975-1995. During this period 60 chains were detected, and 46 were eliminated by the intermediate-term patterns, according to the RDP methodology. Out



Figure 7: Performance for California. Gray background depicts learning period, yellow background — testing period. Correct chains and their vicinities are shown in red, false chains and their vicinities — in blue. The chain corresponding to advance prediction of San Simeon earthquake is zoomed up in the lower panel.



Figure 8: Performance for Southern California. Gray background depicts learning period, yellow background — testing period. Correct chains and their vicinities are shown in red, false chains and their vicinities — in blue. The chain corresponding to the current alarm is zoomed up in the lower panel.

of the remaining 14 chains 5 are false alarms, and 9 are correct. Since 1995 29 chains were detected, 26 of them were eliminated by RDP approach, and out of the 3 remaining 2 are false alarms, and the last one is correct: it predicted the recent Tokachi-oki earthquake. See Sect. 3.1 for more information. Note that there is one more strong earthquake (M = 7.0) that occurred within this last alarm. This earthquake confirms the possibility of having more than one strong earthquake within a given alarm. The total spatio-temporal volume of alarms is 12.4%.

2.4.4 Eastern Mediterranean

Performance of the algorithm for Eastern Mediterranean is illustrated in Fig. 10. Parameters of the algorithm are given in the Table 2. The voting threshold is C = 45. The learning period for this region is 1983-1997. During this period 7 chains were detected, and 5 were eliminated by the intermediate-term patterns, according to the RDP methodology. The remaining 2 chains produced correct alarms. Since 1995 3 chains were detected, all 3 were eliminated by the intermediate-term patterns, and no strong earthquake occurred. The total spatio-temporal volume of alarms is 2.4%.

2.5 Stability tests

Here we describe different stability tests that were performed for Southern California region. The same (or similar) tests have been done for all the considered regions. In general, such tests are the most important (and time- and effort-consuming) part of development of a prediction algorithm, which in lieu of adequate theory hardly relies on stability of the performance in respect to variations of its adjustable elements.

2.5.1 Variation of chain parameters

Here we test the stability of the short-term part of our prediction algorithm, namely detection of chains. Recall that chain definition involves 7 parameters: the catalog cut-off magnitude m_{\min} , time and space thresholds τ_0 and r_c that define "neighbors", normalization constant c, and thresholds k_0 , l_0 , and γ_0 . We fixed γ_0 as shown in Table 2 and varied the other 6 parameters independently within the following limits: $m_{\min} = 2.7(0.1)4.0$; $\tau_0 = 5(5)60$ days; $r_c = 25(5)100$ km; c = 0.15(0.05)0.6; $k_0 = 4(1)9, 10(2)18, 20(5)30$; $l_0 = 100(25)1000$ km. Altogether this gives about 10^7 different sets of chains. For each set we determined the following statistics, evaluating the performance of the chains: number $n_{\text{predicted}}$ of strong earthquake targeted for prediction that lie within 75 km from one of the chains (i.e. within 75 km from one of its elements) and within 9 months after it; number k_{chain} of chains; number k_{correct} of chains that precede one of the strong earthquakes targeted for prediction.

We consider as "reliable" those sets of chains which correspond to

$$n_{\text{strong}} - n_{\text{predicted}} \le 1;$$



Figure 9: Performance for Japan. Gray background depicts learning period, yellow background — testing period. Correct chains and their vicinities are shown in red, false chains and their vicinities — in blue. The chain corresponding to advance prediction of Tokachi-Oki (Hokkaido) earthquake is zoomed up in the lower panel.



Figure 10: Performance for Eastern Mediterranean. Gray background depicts learning period, yellow background — testing period. Correct chains are shown in red.

$$k_{\text{chain}} \le 100;$$

 $k_{\text{correct}}/k_{\text{chain}} > 1/20$

Note that the fraction $1 - k_{\text{correct}}/k_{\text{chain}}$ of false chains is allowed to be very high (up to 19/20), since we hope to eliminate the bulk of these false chains during the next, intermediate-term part of the algorithm.

We found that the "reliable" sets of chains corresponds to the fairly wide domain of parameters. Besides, within this domain, the sets of chains are very similar to each other, that is the chain detection is proven to be stable under the variation of parameters. We chose a specific combinations of parameters to be used in the forward prediction from the centers of these stable domains. These parameters are listed in Table 2.

2.5.2 "Historical" experiment

Here we test the intermediate-term part of our prediction algorithm. Specifically, we fix the parameters of chains and the parameters of intermediate-term functionals, except the alarm thresholds C_P . We vary these thresholds and the voting parameter C, perform prediction and see how stable and good are our results.

We apply our prediction algorithm during the time interval $T = [t_{\text{start}}, t_{\text{end}}]$ and calculate the following statistics describing the prediction performance within this interval: number n_{target} of target earthquakes; number n_{fail} of failures to predict; rate $n^o = n_{\text{fail}}/n_{\text{target}}$ of failures to predict; total number k_{chains} of chains; total number k_{prec} of chains recognized as precursory (each chain is counted once); number k_{false} of chains recognized as precursory, but not preceding a target earthquake; rate $f = k_{\text{false}}/k_{\text{prec}}$ of false alarms; and relative volume $\tau^o(R, t, m)$ of alarms, its definition follows.

At any moment t some part of the considered territory can be in the state of alarm initiated by one or, in general case, several precursory chains in their R-vicinities for duration τ . Let $\eta(t)$ be the fraction of all the main shocks with $m \ge m_0$ in the entire catalog for the considered region, that have epicenters within the spatial area of the current alarm (i.e. within the R-vicinity of the corresponding chain(s)), independently of their occurrence time. Then

$$\tau_0 = \frac{1}{t_{\text{start}} - t_{\text{end}}} \int_{t_{\text{start}}}^{t_{\text{end}}} \eta(t) dt.$$
(9)

In all calculations of τ^{o} we take R = 75 km, (R = 100 km for Japan), $\tau = 9$ months, and $m_0 = 4.0$.

Several tests were performed.

Test 0. Here we check that the algorithm, being applied to the total time interval considered gives acceptable results. The results are summed up in Table 3 (the current chain is not included in the analysis).

Table 3: Results of Test 0

		Va	riant	1, C=	=36	Va	riant	2, C=	=39
$k_{\rm chains}$	n_{target}	$k_{\rm prec}$	$k_{\rm false}$	$n_{\rm fail}$	$\tau^o,\%$	$k_{\rm prec}$	$k_{\rm false}$	$n_{\rm fail}$	$\tau^o, \%$
54	7	10	2	0	9.3	7	0	1	7.0

With the optimal values of the thresholds C_P determined during this test, the current chain would have 37 intermediate-term patterns voting for making this chain precursory. Accordingly, it would be recognized as precursory in variant 1, and as non-precursory in variant 2.

The idea of the following tests is to use for learning (e.g. establishing values of the thresholds C_P and C) only part of the chains, and then use the fixed values of C_P and C for the remaining chains. In all the tests we keep the same number (64) of functions for the intermediate-term step of the analysis, and the same values of their parameters. All 64 thresholds C_P are found automatically on the basis of the learning set using strategy of individual minimization of the error $(n^o + k_{\text{false}}/k_{\text{prec}})$ for each function individually.

Test 1. We found that all precursory chains preceding targets earthquakes have had at least one third of events with $M \ge 3.5$ (recall that the threshold to construct chains was $M \ge 2.9$). The question is: if to use for learning only chains matching this rule, will the remaining chains be recognized correctly as non-precursory? Results are summarized in the Table 4.

		Variant 1, $C=37$				Variant 2, C=41				
	$k_{\rm chains}$	n_{target}	$k_{\rm prec}$	$k_{\rm false}$	$n_{\rm fail}$	$ au^o,\%$	$k_{\rm prec}$	$k_{\rm false}$	$n_{\rm fail}$	$ au^o,\%$
Learning set	36	7	12	4	0	12.5	7	0	1	7.0
Test set	18	0	2	2	0		0	0	0	
Learning+test	54	7	14	6	0	13.5	7	0	1	7.0

Table 4: Results of **Test 1**

With the optimal values of the thresholds C_P determined during this test, the current chain would have 38 intermediate-term patterns voting for making this chain precursory. Accordingly, it would be recognized as precursory in variant 1, and as non-precursory in variant 2. If we put C = 38, then 11 chains, including the current one, would be recognized as precursory, with one failure to predict, three false alarms and one current alarm.

Test 2. Recall that the total number of the detected chains is 54 (excluding the current one). Let us split this amount, by some time t_{start} , into two parts: $T_{\text{learning}} = [t_0, t_{\text{start}}]$

and $T = [t_{\text{start}}, t_{\text{end}}]$. The first part is used for learning, and the second part for test. Successively increasing time t_{start} , so that each time one chain is reassigned from test to learning, we shall see the performance of the algorithm and its stability. For this experiment we have to set up the rule to choose the value of the threshold C.

We considered the following variants:

- a) C minimizes $n^o + k_{\text{false}}/k_{\text{chain}}$;
- b) C minimizes $n^o + (k_{\text{chain}}\tau)/(t_{\text{start}} t_0);$
- c) Maximal value of C giving one false alarm;
- d) Maximal value of C giving one failure to predict;
- e) Fixed values: C = 36, C = 37, C = 38.

Results of these tests show that

- The learning set which includes 5 target earthquakes (Landers being the last one) and consists of 36 first chains is enough for very good "advance" prediction, with no failure to predict and at maximum one false alarm. This is 2/3 of the total data.
- The performance is worse, but obviously not random, with low fixed threshold C (strategy e), even with minimal learning set we could imagine (did not try smaller)
 20 chains and 2 targets (1/3 of the data). It is remarkable that portion of false alarms above the threshold C is approximately the same in the test set and in the learning set.
- Stability of the results to the consecutive update of the learning set is quite impressive.
- The performance is very stable to the choice of the threshold C; the range of its acceptable values is quite broad.

Test 3. This is similar to the Test 2, but each consecutively growing learning set, from 20 to 53 chains, is used only to classify one next chain. Statistics of false alarms and failures to predict for same strategies is in the Table 5

strategy		a	b	с	d	e36	e37	e38	e39	e40
Test starts in 1979	$n_{\rm fail}$	2	2	3	4	0	0	0	1	2
(34 chains tested)	$k_{\rm false}$	2	2	1	0	12	9	9	7	5
Test starts in 1993	$n_{\rm fail}$	0	0	1	2	0	0	0	1	1
(16 chains tested)	$k_{\rm false}$	2	2	2	0	2	1	1	1	1

Table 5: Results of **Test 3**

3 Advance Predictions

Here we describe three predictions made up to the moment in an ongoing experiment on short-term earthquake prediction that covers Japan, California, and East Mediterranean. Recall that these predictions are based on a short-term premonitory pattern "Chain" which reflects, qualitatively, an increase of the correlation range among small earthquakes (See Sect. 2.3,2.3.1).

3.1 Advance prediction of Hokkaido earthquake

The M8.1 Hokkaido (Tokachi-oki) earthquake of September 25, 2003 was predicted 6 months in advance. The precursor was detected in Japan in March, 2003. The results of the experiment, including the reported precursor, were presented at the XXIII Assembly of IUGG (Hagiwara Symposium on earthquake prediction) in July, 2003, six months prior to the Hokkaido earthquake.

The region considered for prediction is shown in Fig. 2. The premonitory chain that produced a correct alarm is shown in Fig. 11. Parameters used to detect the premonitory chain are given in Table 2. Parameters of intermediate-term functions are described in Sect. 2.3.2. The voting threshold is C = 36.

3.2 Advance prediction of San Simeon earthquake

The M6.5 San Simeon earthquake of December 22, 2003 was predicted 7 months in advance. The precursor was detected in May, 2003. On June 21, 2003 the information on the detected precursor was sent to a group of leading experts and officials.

The region considered for prediction is shown in Fig. 1. The premonitory chain that produced a successful alarm is shown in Fig. 12. Parameters used to detect the premonitory chain are given in Table 2. Parameters of intermediate-term functions are described in Sect. 2.3.2. The voting threshold is C = 41.

3.3 Current alarm: Southern California

A precursory chain was detected in Southern California on October 29, 2003. This chain is shown in Fig. 13 The chain was extended four times since then, the last time on December 5-th, 2003. Accordingly, the alarm was declared within the R-vicinity of the chain for the period October 29, 2003 — September 5, 2004.

The region considered for prediction is shown in Fig. 1. The premonitory chain that produced a successful alarm is shown in Fig. 13. Parameters used to detect the premonitory chain are given in Table 2. Parameters of intermediate-term functions are described in Sect. 2.3.2. The γ condition was not applied. The voting threshold is C = 41.



Figure 11: Precursory Chain for Hokkaido Earthquake. The precursory chain was reported on July 2, 2003 at the XXIII Assembly of IUGG. Red circles show earthquakes that form the precursory chain; their size is proportional to magnitude. The gray shadowed area shows R-vicinity of the chain; a strong earthquake is expected here during 9 months after the chain emergence. Blue stars show epicenters of two large earthquakes that happened within the chain vicinity during the 9 months after its emergence. One of them is the M8.1 Hokkaido (Tokachi-oki) earthquake.



Figure 12: Precursory Chain for San Simeon Earthquake. Earthquakes that form the precursory chain are shown by red circles; an earthquake with magnitude 6.4 or above is expected within gray area by Feb. 28, 2004; the epicenter of San Simeon earthquake is shown by blue star.



Figure 13: Current Precursory Chain. An earthquake with magnitude 6.4 or above is expected within the gray area within the period Oct. 29, 2003 - Sep. 05, 2004. Red circles show the earthquakes that formed the precursory chain.

References

- Aki, K., Introduction to seismology for earthquake prediction, preprint, Proceedings of the Seventh Workshop on Non-Linear Dynamics and Earthquake Prediction, International Center for Theoretical Physics, Trieste, 2003.
- [2] ANSS/CNSS Worldwide Earthquake Catalog, produced by Advanced National Seismic System (ANSS) and hosted by the Northern California Data Center (NCEDC), 1965 - 2003. http://quake.geo.berkeley.edu/anss
- [3] Israeli Earthquake catalog is available at http://www.gii.co.il
- [4] Bird, P., Testing hypotheses on plate-driving mechanisms with global lithosphere models including topography, thermal structure, and faults, J. Geophys. Res., 103, B5, 10115-10129, 1998.
- [5] Blanter, E.M., and M.G. Shnirman, Simple hierarchical systems: stability, selforganized criticality and catastrophic behavior, *Phys. Rev. E*, 55(6), 6397-6403, 1997.
- [6] Bowman, D.D., G. Ouillon, C.G. Sammis, A. Sornette, and D. Sornette, An observational test of the critical earthquake concept, J. Geophys. Res., 103, 2435924372, 1998.
- [7] Caputo M., R. Console, A.M. Gabrielov, V.I. Keilis-Borok, and T.V. Sidorenko, Long-term premonitory seismicity patterns in Italy, Geophys. J. R. Astron. Soc., 75, 71-75, 1983.
- [8] Gabrielov, A., V. Keilis-Borok, I. Zaliapin, and W. I. Newman, Critical transitions in colliding cascades, Phys. Rev. E, 62(1), 237-249, 2000.
- [9] Gabrielov, A.M., I.V. Zaliapin, W.I. Newman, and V.I. Keilis-Borok, Colliding cascades model for earthquake prediction, Geophys. J. Int., 143, 427437, 2000.
- [10] Jin, A., K. Aki, Z. Liu, and V.I. Keilis-Borok, Brittle-ductile interaction hypothesis for earthquake loading, Abstract, Hagiwara Symposium, XXIII General Assembly of IUGG, 2003, Sapporo, Japan, 2003.
- [11] Keilis-Borok V., Fundamentals of earthquake prediction: Four paradigms, in Nonlinear Dynamics of the Lithosphere and Earthquake Prediction, edited by V.I. Keilis-Borok and A.A. Soloviev, pp. 1-36, Springer-Verlag, Heidelberg, 2003.
- [12] Keilis-Borok, V., Earthquake prediction: state-of-the-art and emerging possibilities, invited to Annu. Rev. Earth Planet. Sci., 30, 2002.
- [13] Keilis-Borok, V.I., and A.A. Soloviev, editors, Nonlinear Dynamics of the Lithosphere and Earthquake Prediction, 335 pp., Springer-Verlag, Heidelberg, 2003.

- [14] Keilis-Borok, V.I., and I.M. Rotwain, Diagnosis of time of increased probability of strong earthquakes in different regions of the world: algorithm CN, Phys. Earth Planet. Inter., 61 (1-2), 57-72, 1990.
- [15] Keilis-Borok, V.I., and P.N. Shebalin, editors, Dynamics of the lithosphere and earthquake prediction, Phys. Earth Planet. Inter., special issue, 111, 179-330, 1999.
- [16] Keilis-Borok, V.I., Intermediate-term earthquake prediction, Proc. Nat. Acad. Sci., 93, 3748-3755, 1996.
- [17] Keilis-Borok, V.I., P.N. Shebalin, and I.V. Zaliapin, Premonitory patterns of seismicity months before a large earthquake: Five case histories in Southern California, Proc. Natl. Ac. Sci. 99 (26), 16562-16567, 2002.
- [18] Knopoff, L., T. Levshina, V.I. KeilisBorok, and C. Mattoni, Increased longrange intermediatemagnitude earthquake activity prior to strong earthquakes in California, J. Geophys. Res., 101, 57795796, 1996.
- [19] Kossobokov, V., and P. Shebalin, Earthquake prediction, in Nonlinear Dynamics of the Lithosphere and Earthquake Prediction, edited by V.I. Keilis-Borok and A.A. Soloviev, pp. 141-207, Springer-Verlag, Heidelberg, 2003.
- [20] Kossobokov, V.G., and J.M. Carlson, Active zone size vs. activity: A study of different seismicity patterns in the context of the prediction algorithm M8, J. Geophys. Res., 100, 6431-6441, 1995.
- [21] Ma, Z., Z. Fu, Y. Zhang, C. Wang, G. Zhang, and D. Liu, Earthquake Prediction: Nine Major Earthquakes in China, Springer, New York, 1990.
- [22] Mogi, K., Earthquake Prediction, Academic Press, Tokyo, 1985.
- [23] Molchan, G.M., Earthquake prediction strategies: A theoretical analysis, in Nonlinear Dynamics of the Lithosphere and Earthquake Prediction, edited by V.I. Keilis-Borok and A.A. Soloviev, pp. 209-237, Springer-Verlag, Heidelberg, 2003.
- [24] Newman, W.I., A. Gabrielov, and D.L. Turcotte, editors, Nonlinear dynamics and predictability of geophysical phenomena, Geophys. Monographs Ser. 83. Am. Geophys. Union, Washington, DC, 83, 1994.
- [25] Press, F., editor, Earthquake Prediction: A Proposal for a Ten Year Program of Research, 134 pp., Ad Hoc Panel on Earthquake Prediction, White House Office of Science and Technology, Washington, DC, 1965.
- [26] Prozorov, A.G., S.Yu. Schreider, Real time test of the long-range aftershock algorithm as a tool for mid-term earthquake prediction in Southern California, Pure Appl. Geophys. 133, 329-347, 1990.

- [27] Rundle, J., D. Turcotte, and W. Klein, editors, Geocomplexity and the Physics of Earthquakes, American Geophysical Union, Washington, DC, 2000.
- [28] Rundquist, D.V., and A.A. Soloviev, Numerical modeling of block structure dynamics: an arc subduction zone, Phys. Earth and Planet. Inter., 111 (3-4), 241-252. 1999.
- [29] Scholz, C.H., The Mechanics of Earthquakes and Faulting, Cambridge University Press, Cambridge, 1990.
- [30] Shaw, B.E., J.M. Carlson, and J.S. Langer, Patterns of seismic activity preceding large earthquakes, J. Geophys. Res., 97, 479, 1992.
- [31] Shebalin, P., I. Zaliapin, and V.I. Keilis-Borok, Premonitory rise of the earthquakes' correlation range: Lesser Antilles, Phys. Earth Planet. Int., 122 (34), 241249, 2000.
- [32] Shebalin, P., V.I. Keilis-Borok, I. Zaliapin, S. Uyeda, T. Nagao, and N. Tsybin, Short-term Premonitory Rise of the Earthquake Correlation Range, Abstract, Hagiwara Symposium, XXIII General Assembly of IUGG, 2003, Sapporo, Japan, 2003.
- [33] Sornette D., Critical Phenomena in Natural Sciences: Chaos, Fractals, Selforganization and Disorder: Concepts & Tools, Springer Ser. Synerg., 432 pp., Springer-Verlag, Heidelberg, 2000.
- [34] Turcotte, D.L., Fractals and Chaos in Geology and Geophysics, 2nd ed., 412 pp., Cambridge University Press, Cambridge, 1997.
- [35] Uyeda, S., and S. Park, editors, special issue, Proceedings of the International Symposium on The Recent Aspects of Electromagnetic Variations Related with Earthquakes, 20 and 21 December 1999, J. of Geodynamics, 33, 4-5, 2002.
- [36] Vorobieva, I.A., Prediction of a subsequent strong earthquake, Phys. Earth Planet. Inter., 111, 197-206, 1999.
- [37] Zaliapin, I., V.I. Keilis-Borok, and G. Axen, Premonitory spreading of seismicity over the faults' network in southern California: Precursor Accord, J. Geophys. Res., 107, B10, 2221, 2002.
- [38] Zaliapin, I., V.I. Keilis-Borok, and M. Ghil, A Boolean delay equation model of colliding cascades, Part II: Prediction of critical transitions, J. Stat. Phys., 111, 3-4, 839-861, 2003.

Appendices

A What is Earthquake Prediction?

Vulnerability of our world to the earthquakes is rapidly growing, due to the well known global trends: proliferation of high-risk constructions, such as nuclear power plants, high dams, radioactive waste disposals, lifelines, etc.; deterioration of ground and destabilization of engineering infrastructures in megacities; destabilization of environment; population growth; and other factors, up to escalating socio-economic volatility of the global village. Today a single earthquake with subsequent ripple effects may take up to a million of lives; cause material damage up to 10¹2; destroy a megacity; trigger a global economic depression (e.g. if it occurs in Tokyo); trigger ecological catastrophe, rendering a large territory inhabitable; destabilize military balance in a region (e.g. Middle East). Highly vulnerable became the regions of low seismicity; among them are European and Indian platforms, Central and Eastern US, harboring megacities like New York, Moscow, Rome.

As a result the earthquakes joined the ranks of the major disasters that, in the words of J. Wisner, became "a threat to civilization survival, as great as was ever posed by Hitler, Stalin or the atom bomb". Earthquake prediction on either stage would open a possibility to reduce the damage by undertaking a set of disaster preparedness measures.

A.1 Formulation of the Problem

The United States National Research Council, Panel on Earthquake Prediction of the Committee on Seismology suggested the following definition [AEH+76, p7]: "An earthquake prediction must specify the expected magnitude range, the geographical area within which it will occur, and the time interval within which it will happen with sufficient precision so that the ultimate success or failure of the prediction can readily be judged. Only by careful recording and analysis of failures as well as successes can the eventual success of the total effort be evaluated and future directions charted. Moreover, scientists should also assign a confidence level to each prediction." Accordingly, one can identify an earthquake prediction of certain magnitude range by duration of time interval and/or by territorial specificity. Commonly, temporal classification loosely distinguishes long-term (tens of years), intermediate-term (years), short-term (tens of days), and immediate (days and less) prediction.

Rethinking Earthquake Prediction, Lynn Sykes et al. [SSS99] wrote: "The public perception in many countries and, in fact, that of many earth scientists is that earthquake prediction means short-term prediction, a warning of hours to days. They typically equate a successful prediction with one that is 100% reliable. This is in the classical tradition of the oracle. Expectations and preparations to make a short-term prediction of a great earthquake in the Tokai region of Japan have this flavor. We ask instead are there any time, spatial and physical characteristics inherent in the earthquake process that might

Temporal, in y	vears	Spatial, in Source Zone Size			
Long-term	10	Long-range	Up to 100 L		
Intermediate-term	1	Middle-range	510~L		
Short-term	0.01 – 0.1	Narrow	2–3 L		
Immediate	0.001	Exact	1 L		

Table 6: Classification of Predictions

lead to other modes of prediction and what steps might be taken in response to such predictions to reduce losses?"

Following common perception many investigators usually overlook spatial modes of predictions and concentrate their efforts on predicting the "exact" fault segment to rupture (e.g. Parkfield earthquake prediction experiment [BL85]), which is by far more difficult and might be an unsolvable problem. Being related to the rupture size L of the incipient earthquake, such modes can be summarized in a classification that distinguishes wider prediction ranges in addition to the "exact" location of a source zone (Table 6). The accuracy of prediction might improve when independent observations are brought into analysis.

From a viewpoint of such classification, the earthquake prediction problem might be approached by a hierarchical, step-by-step prediction technique, which accounts for multiscale escalation of seismic activity to the main rupture [KB90A, KMU99]. It starts with recognition of earthquake-prone zones for earthquakes from a number of magnitude ranges, then follows with determination of long- and intermediate-term areas and times of increased probability, and, finally, may come out with an exact short-term or immediate alert.

A.2 Data

Catalogs of earthquakes remain to be the most objective record of seismic activity on the Earth. It is common knowledge that catalogs have some errors [HC94]. It is desirable to have independent records of seismicity for their identification and elimination. In some cases where there is a number of catalogs with common overlap, the pattern recognition technique may help in detecting possible errors and duplicate entries [She87]. Such a technique permits to develop automated procedures, which could reduce the percentage of errors (see e.g. [She92]). The analysis of the frequency-magnitude graph of the catalog for consistency, as well as special searches for duplicates and possible errors are the essential preliminary part in every application of the methodology described below.

In general, the errors in the data can be neutralized in two ways: First, by postponing

the analysis until the data are refined, and second, by a robust analysis of existing data within the limits of its applicability. We follow the second way and use routine catalogs of earthquakes to describe the dynamics of seismic regions, derive *precursory seismicity patterns* at the approach of large earthquakes and make predictions based upon these determinations.

A.3 General Scheme of Data Analysis

A.3.1 Earthquakes to be Predicted

We first define a strong earthquake as the one we aim to predict, by the condition that its magnitude $M \ge M_0$. Naturally, the magnitude scale we use should reflect the size of earthquake sources. Accordingly, M_s is usually taken for larger magnitudes while m_b is used for smaller ones, for which M_s determinations are not available. For many catalogs, one can use the maximum reported magnitude (e.g., we do so when using the National Earthquake Information Center/U.S. Geological Survey Global Hypocenters' Data Base [GHDB94]).

In most cases the choice of M_0 is predetermined by the condition that the average recurrence time of strong earthquakes must be sufficiently long in the territory considered. In order to establish a value of M_0 for a seismic territory, we consider values of M_0 with an increment of 0.5, unless actual distribution of earthquake size suggests a natural cutoff magnitude that determines earthquakes sometimes called characteristic [SC84]. The analysis may distinguish a number of intervals $M_0 \leq M < M_0 + 0.5$ and deliver a hierarchy of predictions.

A.3.2 Premonitory Seismicity Patterns

Studies of observed and modeled seismicity have demonstrated that a strong earthquake of magnitude M is preceded by certain spatio-temporal patterns of seismicity in a lower magnitude range. The latter form within a territory and magnitude range normalized by M.

We consider patterns of the following four types: (i) rise of seismic activity; (ii) rise of earthquakes clustering; (iii) rise of earthquake correlation range; and (iv) certain changes in the size distribution (Gutenberg-Richter relation).

Patterns of the first two types have been found in observations first and then in models; patterns of other two types first in models and then in observations. In observed seismicity statistical significance has been established so far only for intermediate-term patterns of the first two types.

A.3.3 Reduction to Pattern Recognition

Inevitably, in the absence of an adequate theory, the patterns above have been found by a pattern recognition analysis of observed or modeled seismicity. This approach, schematically illustrated in Fig. 1, can be summarized as follows:

- Strong earthquakes are identified by the condition $M_0 + c \ge M \ge M_0$, with c about 0.5.
- Prediction is aimed at determination of alarms: time intervals within which a strong earthquake has to be expected.
- A seismic region under investigation is overlaid by areas whose size depends on M_0 . In each area the sequence of earthquakes is robustly described by several functionals $F_k(t, s_k), \ k = 1, 2, \ldots$, defined on sliding time windows $(t - s_k, t)$. We look for the functionals with different distribution of values inside and outside the alarms. Single or joint distributions are explored.
- We have to look for premonitory phenomena common for a wide variety of regions and magnitude ranges, as well as within sufficiently long time periods. Otherwise the test of a prediction algorithm would be practically impossible. Accordingly, as often in dealing with high complexity, we have to find the robust and normalized definitions of premonitory patterns, applicable in different conditions. This applicability is achieved at a price: reduction of the accuracy of predictions.
- Pattern recognition of infrequent events happened to be very efficient for that purpose. This methodology has been developed by the school of I. Gelfand [GGK+76] for the study of rare phenomena of highly complex origin, a situation where classical statistical methods are inapplicable. It is akin to exploratory data analysis, developed by J. Tukey [Tu].

A.3.4 Earthquake sequence.

The simplest routine catalogs of observed earthquakes provide the sequence:

$$(t_k, m_k, h_k), \ k = 1, 2, \dots; \ t_k \le t_{k+1}.$$
 (10)

Here, t_k is the starting time of the rupture, m_k is the magnitude, and h_k is the vector that comprises the coordinates of the hypocenter. We exclude aftershocks from the analyzed earthquake sequence. Premonitory patterns are defined on the sequence of the main shocks. However the number of aftershocks is retained for each main shock.

A.3.5 Premonitory patterns

We determine for this sequence different functionals F(t) depicting the above characteristics of seismicity [KB90A]. A premonitory pattern is defined by the following condition

$$F(t) \ge C_F \tag{11}$$

The threshold C_F is usually defined as a certain percentile of the functional F.

A.3.6 Prediction algorithms

Prediction algorithm based on a single pattern is formulated as follows. An alarm is declared for the time period τ_F after each moment when the condition (11) holds. The alarm is terminated after a major earthquake occurs or the time t expires, whichever comes first. The first case is a confirmed prediction ("success"), the second - a false alarm. A failure to predict is the case when a major earthquake occurs outside an alarm. Note that one or several informative patterns can be used for prediction. Moreover, a combination of patterns can be useful for prediction if some or even all of them show unsatisfactory performance when applied separately. Many algorithms are based on combination of the patterns.

Obviously this scheme is open for including other patterns and other data, not necessarily seismological ones.

B Disaster management point of view

"Of course, things are complicated. But in the end every situation can be reduced to a simple question: Do we act or not? If yes, in what way?" /E. Burdick/

"Far better is an approximate answer to the right question, which is often vague, than an exact answer to the wrong question which can always be made precise." /J. Tukey/

How to reduce the damage from the earthquakes on the basis of predictions, given their current accuracy, not necessarily high? The key to this is to escalate or de-escalate disaster preparedness measures depending on the content of a current alarm: (what and where is predicted; probability of the false alarm; and cost/benefit ratio of different measures. Such is the standard practice in preparedness to all disasters, war included. A costly mistake — that only a precise short-term prediction is practically useful (besides estimation of seismic hazard) is sometimes emerging in seismological literature. Actually, as in the case of defense, prediction may be useful if its accuracy is known, even if it is not high. Diversity of damage. Earthquakes may hurt population, economy, and environment in very different ways: destruction of buildings, lifelines, etc.; triggering of fires; release of toxic, radioactive and genetically active materials; triggering of floods, avalanches, landslides, tsunamis, etc. Equally dangerous are the socio-economic and political consequences of the earthquakes: disruption of vital services - supply, medical, financial, law enforcement etc.; epidemics; disruptive anxiety of population, profiteering and crime; drop of production, slowdown of economy, and unemployment; destabilization of military balance; disruptive anxiety of population, profiteering and crime. The socio-economic consequences may be inflicted also by the undue release of predictions. Different kinds of damage are developing in different time — and space scales, ranging from immediate damage to chain reaction, lasting tens of years and spreading regionally

if not worldwide. A hierarchy of disaster preparedness measures is required by such diversity of possible damage. Permanent measures include: restriction of land use; building codes; insurance; preparedness of civil defense type; R&D. Temporary measures, activated in response to a time prediction include: enhancement of permanent measures safety control, simulation alarms etc.; partial neutralization of high — risk objects; mobilization of post - disaster services; emergency legislation, up to martial law; evacuation of population etc. These measures are required in different forms on local, provincial, national and international levels. Different measures require different lead time, from seconds to years, to be activated; having different cost they can be realistically maintained for different time — periods, from hours to decades; and they have to be spread over different territories — from selected points to large regions. No single type can replace another one for damage reduction and no single measure is sufficient alone. On the other hand many important measures are in expensive and do not require high accuracy of prediction. Response to predictions. Traditionally a prediction is made with a single combination of adjustable elements, chosen as the best in some sense. Disaster preparedness measures would be more flexible and efficient, if prediction is made in parallel with several such combinations, that is several versions of an algorithm.

C Intermediate-Term Earthquake Prediction

C.1 Validated Prediction Algorithms

C.1.1 Algorithm M8

This intermediate-term earthquake prediction method was designed by retrospective analysis of dynamics in seismicity preceding the greatest, magnitude 8.0 or more, earthquakes worldwide, hence its name. Its prototype [KBK84] and the original version [KBK87] were tested retrospectively in the vicinities of 143 points, of which 132 are recorded epicenters of earthquakes having magnitude 8.0 or greater from 1857 to 1983. In 1986 algorithm M8 was tested in retrospective application to predict earthquakes of smaller magnitudes, down to 6.5 [Kos86], by using independent regional seismic databases (note that the USGS/NOAA global database available to us at that time covered the period through 1983 only). By 1990 the list of the territories, where the original and other versions of algorithm M8 have been applied, extended to 19 regions listed in Table reftab4.2 [M8].

Algorithm M8is based on a simple physical scheme of prediction briefly described below. The values of constant parameters entering the algorithm are listed in the end of the description.

Prediction is aimed at earthquakes of magnitude M_0 and higher. Overlapping circles of diameters $D(M_0)$ scan the territory of the seismic region under study. The sequence of earthquakes with aftershocks removed is considered within each circle. Denote this sequence by

 $\{t_i, m_i, h_i, b_i(e)\}, i = 1, 2, \ldots,$

Region	M_0	Time period	$N_{\rm T}/N$	$V_{\rm TIPs}$	$N_{\rm suc}/N_{\rm all}$
		Learning	,		
1. World	8.0	1967 - 1982	5/7	5	7/16
Tes	ting	of the Origi	nal Version		
2. Central America	8.0	1977 - 1986	1/1	16	1/2
3. Kuril Islands					
and Kamchatka	7.5	1975 - 1987	2/2	17	2/3
4. Japan and Taiwan	7.5	1975 - 1987	5/6	20	6/8
5. South America	7.5	1975 - 1987	3/3	18	3/8
6. Western U.S.A.	7.5	1975 - 1987	-/-	5	0/1
7. Southern California	7.5	1947 - 1987	1/1	12	1/1
8. Western U.S.A.	7.0	1975 - 1987	2/2	24	2/2
9. Baikal and					
Stanovoy Range	6.7	1975 - 1986	-/-	0	-/-
10. Caucasus	6.5	1975 - 1987	2/3	12	2/2
11. East Central Asia	6.5	1975 - 1987	4/5	24	5/6
12. Eastern Tien Shan	6.5	1963 - 1987	4/4	27	5/5
13. Western Turkmenia	6.5	1979 - 1986	-/-	0	-/-
14. Apennines	6.5	1970 - 1986	1/1	10	1/1
15. Koyna reservoir	4.9	1975 - 1986	1/1	42	1/1
Te	estin	g of Modified	d Versions		
16. Greece	7.0	1973 - 1987	3/3	18	4/5
17. Himalayas	7.0	1970 - 1987	2/2	8	3/4
18. Vrancea	6.5	1975 - 1986	2/2	58	2/2
19. Vancouver Island	6.0	1957 - 1985	4/4	20	5/7
Regions $1-19$	toge	e ther	39/44 (89%)	18	49/72
Regions $2-15$	toge	25/28 (89%)	16	28/38	

Table 7: Summary of TIPs Diagnosed by Algorithm M8 (after [M8])

N and $N_{\rm T}$ are the number of all earthquakes and their number within TIPs. $V_{\rm TIPs}$ is the space-time fraction of TIPs

 $N_{\rm all}$ and $N_{\rm suc}$ are the number of all and successful TIPs, respectively

where t_i is the origin time, $t_i \leq t_{i+1}$; m_i is the magnitude; h_i is focal depth; and $b_i(e)$ is the number of aftershocks with magnitude M_{aft} or greater during first e days. The sequence is normalized by the lower magnitude cutoff $\underline{M} = M_{\min}(\widetilde{N})$, \widetilde{N} being the standard value of the average annual number of earthquakes in the sequence.

Several running averages are computed for this sequence in the trailing time window (t-s, t) and magnitude range $M_0 > m_i \ge \underline{M}$. They depict different measures of intensity in earthquake flow, its deviations from the long-term trend, and clustering of earthquakes. The averages include:

- $N(t) = N(t | \underline{M}, s)$, the number of main shocks of magnitude \underline{M} or larger in (t s, t);
- $L(t) = L(t | \underline{M}, s, t_0)$, the deviation of N(t) from longer-term trend, $L(t) = N(t) N_{\text{cum}}(t t_0)/(t t_0 s)$, where $N_{\text{cum}}(t) = N(t | \underline{M}, t t_0)$ is the cumulative number of main shocks with $M \geq \underline{M}$ from the beginning of the sequence t_0 to t;
- $Z(t) = Z(t \mid \underline{M}, \overline{M}, s, \alpha, \beta)$, linear concentration of main shocks i from the magnitude range $(\underline{M}, \overline{M} = (M_{\min}(\widetilde{N}, M_0 - g \text{ and interval } (t - s, t); \text{ the linear concentration}$ is estimated as the ratio of the average source diameter l to the average distance rbetween sources; and
- $B(t) = B(t | \underline{M}, \overline{M}, s', M_{aft}, e) = \max_{\{i\}} \{b_i|\}$, the maximum number of aftershocks (i.e. a measure of earthquake clustering). The sequence $\{i\}$ is considered in the trailing time window (t s', t) and in the magnitude range $(\underline{M}, \overline{M}) = (M_0 p, M_0 q)$.

Each of the functions N, L, and Z is calculated twice with $\underline{M} = M_{\min}(N)$, for N = 20and N = 10. As a result, the earthquake sequence is given a robust averaged description by seven functions: N, L, Z (twice each), and B.

"Very large" values are identified for each function from the condition that they exceed Q percentiles (i.e., they exceed Q percent of the encountered values).

An alarm or a TIP, Time of Increased Probability, is declared for five years when at least six out of seven functions, including B, become "very large" within a narrow time window (t - u, t). To stabilize prediction, this criterion is checked at two consecutive moments, t and t + 0.5 years. In the course of a forward application, the alarm can extend beyond or terminate in less than five years when updating causes changes of the magnitude cutoffs and/or the percentiles.

The following standard values of parameters indicated above are prefixed in the algorithm M8: $D(M_0) = (exp(M_0 - 5.6) + 1)^\circ$ in degrees of meridian (this is 384 km, 560 km, 854 km and 1333 km for $M_0 = 6.5$, 7.0, 7.5, and 8 respectively), s = 6 years, s' = 1 year, g = 0.5, p = 2, q = 0.2, u = 3 years, and Q = 75% for B and 90% for the other six functions. Usually, the average diameter l of the source, is estimated by $(n)^{-1} \sum_{\{i\}} 10^{\beta(M_i - \alpha)}$ where N is the number of main shocks in $\{i\}$, $\beta = 0.46$ to represent the linear dimension of source, and $\alpha = 0$ (which does not restrict generality), while the average distance r between sources is set proportional to $(N)^{-1/3}$. The performance of the algorithm can be improved by estimating the linear concentration of main shocks more accurately [RK96].

Running averages are defined in a robust way, so that a reasonable variation of parameters does not affect predictions. At the same time, discrete character of seismic data and strict usage of the prefixed thresholds result in a certain discreteness of the alarms.

C.1.2 Algorithm MSc or "The Mendocino Scenario"

This second approximation prediction method [KKS90] was designed by retrospective analysis of the detailed regional seismic catalog prior to the Eureka earthquake (1980, M = 7.2) near Cape Mendocino in California, hence its name abbreviated to MSc. Given a TIP diagnosed for a certain territory \vec{U} at time \vec{T} , the algorithm is designed to find within \vec{U} a smaller area \vec{V} where the predicted earthquake can be expected. To execute the algorithm, one needs a reasonably complete catalog of earthquakes with magnitudes $M \ge M_0 - 4$, which is lower than the minimum threshold usually utilized by M8. When this condition does not hold, we assume that the dynamics of earthquakes available in the database inherits the behavior from lower levels of seismic hierarchy. The detection of the MSc criteria in such a case is more difficult and might result in additional failuresto-predict.

The essence of MSc can be summarized as follows. Territory \vec{U} is coarse-grained into small squares of size $s \times s$. Let (i, j) be the coordinates of the centers of the squares. Within each square (i, j) the number of earthquakes $n_{ij}(k)$, aftershocks included, is calculated for consecutive short time windows u months long starting from time $t_0 = T - 6$ years onward, to include earthquakes that contributed to the TIP's diagnosis; k is the sequence number of a time window. In this way, the time-space considered is divided into small boxes (i, j, k) of size $(s \times s \times u)$. "Quiet" boxes are singled out for each small square (i, j); they are defined by the condition that $n_{ij}(k)$ is below the Q percentile of n_{ij} . The clusters of q or more quiet boxes connected in space or in time are identified. Area \vec{V} is the territorial projection of these clusters.

The standard values of parameters adjusted for the case of the 1980 Eureka earthquake are as follows: u = 2 months, Q = 10%, q = 4, and s = 3D/16, D being the diameter of circles used in algorithm M8.

Qualitatively, the MSc algorithm outlines such an area of the territory of alarm where the activity, from the beginning of seismic inverse cascade recognized by algorithm M8 in declaration of the alarm, is continuously high but infrequently interrupted for a short time. Such interruption must have a sufficient temporal and/or spatial span. The phenomenon, which is used in the MSc algorithm, might reflect the second (possibly, shorterterm and, definitely, narrow-range) stage of the premonitory rise of seismic activity near the incipient source of a main shock.

	Date and time	Region	Latitude	Longitude	Depth	M
	1985/09/19 13:17	Mexico	18.19°N	$102.53^{\circ}\mathrm{W}$	27	8.1
	1986/10/20 06:46	Kermadek	$28.12^{\circ}\mathrm{S}$	$176.37^{\circ}W$	29	8.3
•	$1989/05/23 \ 10:54$	Macquarie	$52.34^{\circ}\mathrm{S}$	$160.57^{\circ}\mathrm{E}$	10	8.2
	1993/08/08 08:34	Guam	$12.98^{\circ}\mathrm{N}$	$144.80^{\circ}{\rm E}$	59	8.2
٠	1994/06/09 00:33	Bolivia	$13.84^\circ\mathrm{S}$	$67.55^{\circ}W$	631	8.2
	1994/10/04 13:22	Shikotan	$43.77^{\circ}\mathrm{N}$	$147.32^{\circ}\mathrm{E}$	14	8.3
	1995/04/07 22:06	Samoa	$15.20^{\circ}\mathrm{S}$	$173.53^{\circ}\mathrm{W}$	21	8.1
	1995/12/03 18:01	Iturup	$44.66^{\circ}\mathrm{N}$	$149.30^{\circ}\mathrm{E}$	33	8.0
	$1996/02/17 \ 05:59$	New Guinea	$00.89^{\circ}S$	$136.95^{\circ}\mathrm{E}$	33	8.2
٠	1998/03/25 03:12	Balleny	$62.88^{\circ}\mathrm{S}$	$149.53^{\circ}\mathrm{E}$	10	8.3
	2000/06/04 16:28	Sumatera	$04.77^{\circ}\mathrm{S}$	$102.05^{\circ}\mathrm{E}$	33	8.0

Table 8: Earthquakes of Magnitude M = 8.0 or more, 1985–2000

Bullets mark earthquakes outside circles of investigation

C.1.3 Prediction of Largest Earthquakes, $M \ge 8$.

We applied algorithm M8 and then MSc in the areas of current alarm in 262 overlapping circles of investigation. Specifically, 170 circles were selected from larger number that scan near-uniformly the Circum-Pacific and its surroundings, whereas the other 92 circles are taken from the Alpine-Himalayan and Burma (25 in Mediterranean, 25 in Asia Minor and Iran, 28 in Pamirs-Hindukush, and 14 in Burma). Thus, we may conclude the completeness of the NEIC GHDB is sufficient for application of the original version of M8 in 80-90% of major seismic belts. A sample prediction, as on the date of writing this text, is given in Fig. 14 (a complete set of predictions in 1985-2000 could be viewed at http://mitp.ru/predictions.html). Earthquakes of magnitude 8.0 or more are expected in 16 circles, each of radius 667 km, that form seven compact areas of alarm. In the second approximation, the MSc algorithm outlines nine smaller areas inside alarms, where the great earthquakes are most likely. It is worth mentioning that, due to a very low recurrence rate of great earthquakes, most of the areas, perhaps all of them, will not be confirmed until the next update. However, the alarms last for many years (about five on the average in accordance with τ) and indeed correspond to areas and times with increased probability of occurrence already confirmed in the test. The probability gain depends on locality and varies from 2–3 in regions of extremely high activity, like Tonga-Kermadek, to 20–100 in regions where recurrence of great earthquakes is much lower than the average, like Sumatera.

The performance of both algorithms is illustrated in Fig. 15 for one segment of the Circum-Pacific, namely from Kamchatka to Marianas Trench. For the whole territory



Figure 14: Global testing of algorithms M8 and MSc, $M_0 = 8.0$: Areas of alarm as on January 1, 2001. Circular areas of alarm in the first approximation, i.e. from algorithm M8, are shaded light gray, rectangular areas determined in the second approximation by algorithm MSc are shaded dark gray



Figure 15: Global testing of algorithms M8 and MSc, $M_0 = 8.0$: Space-time distribution of alarms from Kamchatka to Marianas. Circles of investigation (*light*) and their centers (*heavy dots*) are shown on the left; the space-time distribution of alarms in 1985–2000 is on the right (*dark* for M8 and *darker* for MSc). The space coordinate is the distance along the belt. Great earthquakes are marked by *stars*

where prediction is made, the M8 alarms cover on average one third of its length at any given time, while MSc reduces this number to 10%. All eight earthquakes of magnitude 8.0 or greater, which have occurred in the area in the time span 1985–2000 (Table 8), are predicted by M8 and only one of them, the 1996 New Guinea, is missed in the second approximation given by MSc. Table 9 summarizes success-to-failure score; two time intervals are distinguished there, 1985–1997 and 1992–1997. Seismic data for any of the two intervals were not available to the authors at the time when they designed M8 and MSc algorithms. Since 1985 the database has become sufficient for the forward prediction considered, and in 1992, upon publication of [HKD92], the rigid framework of the test has been formally established and the test has begun in real-time mode.

C.1.4 Algorithm SSE (Second Strong Earthquake)

Similar to CN, the algorithm SSE whose name is an abbreviation of Subsequent Strong Earthquake, is another example of applying pattern recognition methods to an earthquake prediction problem [LV92, VL94, VP93, Vor94, Vo]. The algorithm aims at the answer to the question: Whether or not a new strong earthquake can follow the one just occurred. The answer is important for reducing the hazard caused by destabilization of buildings, lifelines, and other constructions or natural objects, like mountain slopes, glaciers, river banks, etc., after a strong earthquake has occurred. Many authors considered similar problems [Bat65, Ver69, Pro78, RJ89, Mat86, HC90].

The advance application of the algorithm SSE started in 1989 demonstrates a high

	Test	Strong earthquakes			Pe	Percentage of alarms				Significance	
	period	Pred	icted by	Total	Circu	m Pacific	Worl	dwide	leve	l, %	
		M8	MSc		M8	MSc	M8	MSc	M8	MSc	
ĺ	1985-2000	8	7	8	37.2	20.0	34.9	18.0	99.96	99.99	
	1992 - 2000	6	5	6	34.1	18.4	30.2	15.3	99.84	99.90	

Table 9: Performance of Earthquake Prediction Algorithms M8 and M8-MSc: Magnitude M = 8.0 or more

Note: The significance level estimates use the most conservative measure μ of the alarm volume accounting for the empirical distribution of epicenters in Circum Pacific.

level of statistical significance estimated currently as more than 98%. Seventeen successful predictions have been made, including the second 1991 Rachi earthquake in Georgia (Caucasus) and three Californian earthquakes at Loma- Prieta (1989), Joshua Tree (1992), and Northridge (1994). The error score is low, consisting of two false alarms and one failure-to-predict.

The Algorithm. Assume that a strong earthquake of magnitude M_1 has just occurred at the origin time t. The task is to predict whether or not a subsequent strong earthquake with magnitude $M \ge M_1 - a$ will occur within the time interval (t + s, t + T) and the circle of radius $R(M_1)$ centered at the epicenter of the occurred one.

Suppose that precursory symptoms similar to those revealed algorithms M8 and CN (see above in this Chapter) precede the occurrence of a subsequent strong event in the vicinity of the occurred shock and their absence signify that no strong earthquake will follow in a certain interval of time.

Naturally, one can presume a similarity of premonitory phenomena only after rescaling which makes aftershock sequences of main shocks with different magnitudes comparable. In the design of algorithm SSE the following scaling rules were applied.

- All magnitude thresholds are derived from the magnitude of the occurred strong earthquake, M_1 .
- The area of investigation and prediction is the circle with radius $R(M_1) = 0.03 \times 10^{0.5M_1}$ km, centered at the epicenter of the occurred strong earthquake.
- Time constants do not scale; the period of prediction is from t + s = t + 40 days to t + T = 1.5 years.

The prediction algorithm SSE was developed in the course of retrospective analysis of 21 strong earthquakes in California with $M \ge 6.4$ [VL94]. A simple pattern recognition measure known as the Hamming distance [GZK+80], is applied to reveal a characteristic

image of an earthquake followed by a subsequent one using eight functions described below. Seven of them refer to an aftershock sequence and reflect the level of aftershock activity, the expansion from the main shock, total area of ruptures, and irregularity in the sequence. One extra function characterizes seismic activation preceding the occurred strong earthquake.

Large values of the following five functions, accounted on the sequence of aftershocks with magnitude equal or exceeding $M_1 - m$ during $(t + s_1, t + s_2)$, favor the occurrence of a subsequent strong earthquake.

N, the total number of aftershocks in the sequence.

S, the total area of aftershock ruptures normalized to the rupture area of the occurred strong earthquake. Specifically, $S = \sum 10^{m_i - M_1}$, where m_i is the magnitude of the *i*th aftershock from the sequence.

 $V_{\rm m}$, the variation of magnitude value in the sequence: $V_{\rm m} = \sum |m_{i+1} - m_i|$, it adds together the absolute values of magnitude difference between subsequent aftershocks.

 V_{med} , the variation of daily average magnitude in the sequence of aftershocks: $V_{\text{med}} = \sum |\mu_{i+1} - \mu_i|$, where μ_i is the daily average magnitude of aftershocks during the *i*th day after the occurred strong earthquake.

 R_z , the deviation of the aftershock sequence from a monotonously decaying one; specifically, $R_z = (1/2) \sum (n_{i+1} - n_i + |n_{i+1} - n_i|)$, where n_i is the number of aftershocks in the interval $(t + i, t + i + s_3)$. The sum neglects negative increments of n_i .

Small values of the following three functions favor the occurrence of a subsequent strong earthquake.

 $V_{\rm n}$, the variation of the daily number of aftershocks in the sequence: $V_{\rm n} = \sum |n_{i+1} - n_i|$, where n_i is the number of aftershocks during the *i*th day.

 R_{max} , the largest distance between epicenters of the occurred strong earthquake and an aftershock from the sequence divided by $R(M_1)$.

 N_{for} , the number of earthquakes of magnitude $M \ge M_1 - m$ during $(t - s_1, t - s_2)$ within the distance of 1.5 $R(M_1)$.

Qualitatively, a characteristic image of an earthquake followed by a subsequent one is as follows: The activity preceding the occurred strong earthquake is low, the number of aftershocks is large, as well as the total area of aftershock ruptures, the aftershock sequence is highly irregular in time and magnitude and decay neither monotonously nor rapidly, and the aftershocks concentrate near the main shock.

Advance Predictions, 1989–1998. The algorithm with pre-fixed parameters has been applied to all strong earthquakes that occurred in the nine regions; Table 10 lists the results of the advance predictions.

No earthquakes with $M \ge M_0$ occurred in Baikal and Stanovoi Range, Turkmenia, and Balkans. There were additional fifteen strong earthquakes that have not been tested; nine were too close in time (within 40 days) to earthquakes listed in Table 10; no data were available for another six earthquakes.

Earthquake	M	Will a subsequent	Outcome	Note
		shock occur?	of prediction	
		California		
Loma-Prieta,	7.1	NO	No shocks with	Success
10/18/1989			$M \ge 6.1$	
Mendocino,	6.9	NO	No shocks with	Success
7/13/1991			$M \ge 5.9$	
Mendocino,	7.1	NO	No shocks with	Success,
8/17/1991			$M \ge 6.1$	first step
Joshua Tree,	6.3	YES	Landers is	Success
4/23/1992			predicted, $M = 7.6$	
Landers,	7.6	YES	Northridge $M = 68$	False alarm
6/28/1992			occurred 19 days	
			after end of alarm	
Northridge,	6.8	NO	No shocks with	Success
1/17.1994			$M \ge 5.8$	
Mendocino,	7.1	NO	No shocks with	Success
4/25/1992			$M \ge 6.1$	
Mendocino,	7.1	NO	Earthquake with	Failure,
9/1/1994			M = 6.8 occurred	first step
Mendocino,	6.8	NO	No shocks with	Success,
2/19/1995			$M \ge 5.8$	first step
California-Nevada	6.3	YES	Earthquake with	Success
border, $9/12/1994$			M = 5.5 occurred	
		Pamir and Tien	e Shan	
Kasakhstan,	7.5	NO	No shocks with	Success
8/19/1992			$M \ge 6.5$	
China,	7.1	NO	No shocks with	Success
11/19/1996			$M \ge 6.1$	
Iran,	7.5	NO	No shocks with	Success
5/10/1997			$M \ge 6.5$	
		Caucasus		
Iran,	7.7	NO	No shocks with	Success
6/20/1990			M > 6.7	
Rachi,	7.1	YES	Earthquake with	Success
4/29/1991			M = 6.6 occurred	
Rachi,	6.6	NO	No shocks with	Success
6/15/1991			$M \ge 5.6$	
Erzincan,	6.8	YES	No shocks with	False
3/13/1992			$M \ge 5.8$	alarm

Table 10: Advance Prediction by Algorithm SSE: The Results of Monitoring in 1989–1998

Earthquake	М	Will a subsequent Outcome shock occur? of prediction		Note				
		Iberia and Maai	hrih					
10c1 iu ana Magnillo								
Morocco,	6.0	NO	No shocks with	Success				
5/26/1994			$M \ge 5.0$					
Dead Sea rift								
Gulf of Aqaba,	5.8	YES	Earthquake with	Success				
8/3/1993			M = 4.9 occurred					
Gulf of Aqaba,	7.3	NO	No shocks with	Success				
11/22/1995			$M \ge 6.3$					
Italy								
Assisi,	6.4	YES	Earthquake with	Success				
9/26/1997			M = 5.4 occurred					

Table 10: (continued)

The advance prediction results can be summarized as follows (Table 11). The rate of failures-to-predict is low, while the rate of false alarms is considerably high when compared with the retrospective testing. The statistics of the advance predictions by the algorithm SSE is not yet sufficient for reliable estimations; nevertheless some preliminary calculations are quite encouraging. Statistical significance of the method, estimated by the technique proposed in [Mol97], is more than 98% [Vo]. The relative number of failuresto-predict n equals 0.2; the rate of alarms τ equals 0.3 (six alarms were diagnosed after 20 strong earthquakes). Thus, the value of $n + \tau$ for advance predictions equals 0.5 being low enough even for a retrospective testing.

C.1.5 Algorithm CN

Algorithm CN [AKB⁺84, KBK⁺88, CN] was developed in the course of retrospective analysis of seismicity patterns preceding earthquakes with $M \ge 6.5$ in California and the adjacent part of Nevada, hence its name. The essence of this algorithm can be briefly summarized as follows.

- Areas of investigation are selected in accordance with the spatial distribution of seismicity.
- Consider earthquakes with the long-term average annual number $\widetilde{N} = 3$ (after eliminating aftershocks) within each area. Compared with $\widetilde{N} = 20$ in algorithm M8 this implies a higher magnitude cutoff M_{\min} and, therefore, relaxed requirements

Prediction: Will a subsequent strong earthquake occur?	Nu (total/	etions	
	Learning	In retrospect	In advance
NO, due to small number of aftershocks	4/0	52/1	4/1
NO, due to pattern recognition criteria	11/0	34/1	10/0
YES	6/1	12/4	6/2
Total	21/1	98/6	20/3

Table 11: Summary of Predictions by SSE

imposed on the completeness of catalogs. This apparent advantage comes at the cost of losing certain degree of robustness, due to lower chances to correctly identify the current state of system dynamics.

- The sequence of earthquakes is described by 9 functions (Table 12). Two of them, N2 and N3, are similar to N and one to B defined in Sect. C.1.1, although with different choice of numerical parameters. Other functions describe the following: the fraction of relatively higher magnitudes in the sequence considered G; the variations of the sequence in time, K and Q; the value of "source energy", SIGMA; and the maximum values of "source area and diameter", S_{max} and Z_{max} .
- Following a pattern recognition routine described in [GGK+76], values of the functions are coarse-grained to distinguish "large", "medium" and "small" separated by 66-and 33-percentiles or just "large" and "small" separated by 50-percentile, i.e., the median.
- The voting of certain pairs or triplets of the discrete values of the nine functions declares (or does not declare) an alarm in the area. The combinations were found originally by applying the pattern recognition algorithm called "Subclasses" [GGK+76] to vectors (patterns) determined from the earthquake catalog of southern California, 1938–1984 and the cutoff magnitude $M_0 = 6.4$. An alarm is declared for a certain period, T = 1 year, when the votes filed by traits D outscore by $\Delta = 5$ those issued by traits N.

Qualitatively, a TIP is diagnosed when earthquake clustering is high, seismic activity is also high, irregular and growing, and some quiescence preceded the increase of seismic activity. Fig, 16 depicts examples of prediction by algorithm CN.

Algorithm CN was first tested retrospectively with pre- fixed parameters for the following 22 areas [CN]: Northern and southern California (6.4), the Gulf of California (6.6), Cocos plate margins (6.5) and adjacent to the belt Lesser Antillean arc (5.5), the



Figure 16: Some examples of the CN algorithm predictions. Times of increased probability of strong earthquakes are determined for southern and northern California, northern Appalachian, and the Cocos Ridge (after [RN99]). Periods of advance prediction are shaded on the time axis

Table 12: Functions Describing Earthquake Sequence in Algorithm CN and Their Thresholds for Descrete Evaluation in southern California

Functions	N2(t)	K	G	SIGMA	S_{\max}	$Z_{\rm max}$	N3	Q	$B_{\rm max}$
First threshold	0	-1	0.5	36	7.9	4.1	3	0	12
Second threshold	-	1	0.67	71	14.2	4.6	5	12	24

Vrancea area of intermediate-depth earthquakes, East Carpathians (6.4), Pamirs (6.5), Tien Shan (6.5), Baikal Lake (6.4), Central Italy (5.6), Caucasus (6.4), Kangra, Nepal and Assam regions in the Himalayas (6.4), Krasnovodsk, Elbruz and Kopet Dag regions in Turkmenia (6.4), area of Dead Sea rift (5.0), Northern and Southern Appalachians (5.0), and Brabant-Ardennes (4.5).

The earthquake catalogs available allowed to retrospectively consider time intervals from 12 to 22 years in each area, amounting to 32 years in Italy and 45 years in California. 60 strong earthquakes occurred in all areas in the test period. 50 (83%) of these events occurred within alarm periods and 10 earthquakes were missed. On average, TIPs in the area considered occupy about 27% of time, from 2 to 4 years per earthquake (except for 6 to 8 years in southern part of Dead Sea rift, Kopet Dag and Vrancea).

Advance prediction in area was carried out for several periods, from a year in the southern Dead Sea rift zone to about 16 years in southern California [RN99]. Altogether, 24 strong earthquakes have occurred in all regions within test periods. Of these, 11 (46%) were predicted and 13 were missed. Total alarm time occupies 26% of the periods considered. In the case of 13 areas, where at least one strong earthquake has occurred within test periods, the time of alarms occupied 29.4% of the total, and it is 18.2% in the remaining 9 regions where no strong earthquakes did occur.

The level of significance $\alpha = 95\%$ follows from a rough statistical estimates [RN99]. However it ignores the fact that the overall statistics are collected from different areas with different rate of seismic activity. A more rigorous and cautious estimation gives $\alpha = 91\%$, which is not very stable and may change to 96 or 81% with the next strong earthquake predicted or not [RN99]. Such level of sensitivity is essentially due to a small sample accumulated so far and can be overcome by continuing the test. The presented cautious estimate ignores the results of monitoring in 9 areas, where strong earthquakes did not occur in the test period. Therefore the significance of predictions by algorithm CN might be higher [MD].

D Prediction Validation

Recent discussions of earthquake prediction make increasing appeal to the language and general properties of dissipative dynamical systems. When dealing with a dynamical system, one is usually interested in the amount of deviation of the system's position from the true one at a given time. The value of a deviation is measured in a suitable metric on the system phase space and is a measure of prediction performance. For example, the classical Kolmogoroff-Wiener problem is concerned with the prediction of a random time series x(t) based on observations available with some time delay τ by the time $t - \tau$. Prediction performance is given by a single value, namely, the relative rms error

$$E|x(t) - \hat{x}_{\tau}(t)|^2 / E|x(t)|^2$$

where \hat{x}_{τ} is the forecast of x, and E denotes mathematical expectation or averaging over the paths of x.

Speaking in the language of dynamical systems, one can treat large seismic events as anomalous states (disasters) in lithosphere dynamics; we are interested first of all in the positions and times τ_n of the disasters. To take an instance, when a time series x is discussed, the disasters may be random times when x goes beyond a critical level (the limiting acceptable load in the physical system). Dealing with a general dynamical system, the τ_n are the times at which the path of the system occurs in a selected region of the phase space. In application to the seismic process, this may be a spatial zone G and a magnitude range $M > M_0$. The Lyapunoff exponents of a dynamical system, which can be used to judge about the horizons of dynamical prediction, tell us practically nothing as to whether disasters can be predicted. In addition, a forecast of disaster time that is very accurate, but delayed $\hat{\tau}_n > \tau_n$, can be unacceptable in practical terms, because a forecast of τ_n has to be made in advance. For this reason the metrical proximity of $\hat{\tau}_n$ and τ_n is useless in disaster prediction.

D.1 Error Diagram

The performance of an earthquake prediction technique actually requires at least two quantities rather than a single one to characterize it: the rate of failures-to-predict and the relative alert time, τ . These can be given precise meaning, when more definitions have been formulated. Here and below, we will only discuss the time behavior of large earthquakes, in other words, forecasts will concern events of magnitude $M > M_0$ in a given area G. The sequence of large events is assumed to make a random point process dN(t) (N(t) is the number of events in the interval (0,t)) of finite rate $\lambda > 0$, i.e., $EdN(t) = \lambda dt$. For simplicity the point process dN(t) will be considered on the lattice $Z_{\delta} = \{\delta k, k = 0, \pm 1, \pm 2, ...\}$ as well, assuming that no more than one event can occur in the interval $(t, t + \delta)$:

$$\operatorname{Prob}\{\delta N(t) = N(t+\delta) - N(t) > 1\} = 0, \quad t \in Z_{\delta}.$$

Let J(t) be the information available at time t for prediction of events in the point process dN(t). In practice, J(t) may include earthquake catalogs for the region containing G, data on physical fields, and observations of precursors. Any type of information is relevant to a constant moving interval of the form $(t - t_i, t - \tau_i)$, where τ_i is the delay of the *i*th data type. The simplest case is where the observer uses information J(t) and makes the decision $\pi(t)$: calling $(\pi = 1)$ or not calling $(\pi = 0)$ an alert in the time interval $(t, t + \delta)$, where δ is some time unit. This may be equal to the time increment at which the information is updated. An event is considered to have been predicted, when it occurred during an alert period, and is a failure-to-predict otherwise. The set of decisions $\{\pi(t)\} = \pi$ is called the prediction *strategy*. In practice the strategy is defined by the method or by the prediction algorithm. Discrete-time strategies were defined above just in order to simplify the discussion.

We now define quantities to characterize the predictive properties of a strategy π in the interval (0, T). These are the relative number of failures-to-predict

$$\hat{n}_{\pi} = \sum_{0 < t < T} (1 - \pi(t)) \cdot \delta N(t) / N(T)$$
(12)

and the relative alert time

$$\hat{\tau}_{\pi} = \sum_{0 < t < T} \pi(t) \delta / T.$$
(13)

Without loss of generality we can assume the information flow J(t) on which the prediction strategy is based to be a multivariate random process. Suppose the process $(\delta N(t), J(t)), t \in Z_{\delta}$ is stationary and ergodic. The application of the individual ergodic theorem [Bi] to (12), (13) will give the result that \hat{n}_{π} and $\hat{\tau}_{\pi}$ converge to the respective constants n_{π} and τ_{π} with probability one. These limits define two long-term prediction errors for the strategy π , the rate of failures-to-predict n_{π} and the relative alert time τ_{π} .

Representation of a strategy π by a pair of numbers n and τ gives a subset $\mathcal{E}(J)$ of the square $[0,1] \times [0,1]$ which depends on information flow J (Fig. 17). It turns out that this set admits of an effective description. The key observation is that any two strategies π_1 and π_2 of the type considered can be combined into a new strategy that independently uses π_1 or π_2 with probabilities q and 1 - q in each time interval δ . This leads to a mixture of parameters $(n, \tau)_i$ of the original strategies with the same weights q and 1 - q. Hence the error set $\mathcal{E} = \{(n, \tau)_{\pi}\}$ corresponding to various strategies is convex, if these strategies are based on the same information J(t). Now note that the error set \mathcal{E} contains points (1, 0) and (0, 1) and, by convexity of \mathcal{E} , the diagonal $n + \tau = 1$. The first point stands for the widespread *optimistic strategy* in which an alert is never declared. The second point corresponds to the total *pessimistic strategy* in which the continuous alert is kept. Points on the diagonal $n + \tau = 1$ correspond to the strategy of a *random guess* in which an alert is declared with probability p independent of J(t).

The set \mathcal{E} has the center of symmetry (1/2, 1/2), because every prediction corresponds to the antipodal prediction π^- where an alert and non-alert swap places and errors (n, τ)



Figure 17: Error set $\mathcal{E}(J)$ for prediction strategies based on a fixed type of information J. Point A corresponds to optimistic strategy, point B to pessimistic strategy, the interval AB corresponds to strategies of random guess. C is the center of symmetry of $\mathcal{E}(J)$. π and π^- are a strategy and its antipodal strategy. Γ is the error diagram of optimal strategies. Arrows indicate a better forecast relative to the strategy π_0 . Dashed lines are contours of the loss function $\gamma = \max(n, \tau)$. Q^* are errors of the minimax strategy, $n = \tau$. Dash-dotted lines are contours of the loss function $\gamma = \tau/(1-n)$

are replaced by $(1 - n, 1 - \tau)$. Therefore all points of \mathcal{E} above the diagonal $n + \tau = 1$ correspond to strategies constructed by rejecting nontrivial strategies with $n + \tau < 1$.

Let us show that only strategies at the lower boundary Γ of the set \mathcal{E} are important. The boundary Γ connects the points (1, 0) and (0, 1). It is monotone and concave due to the properties of \mathcal{E} . So far there is no strategy with errors (0, 0), i.e., an *ideal* strategy that guarantees a 100% prediction of large events in G with no alerts at all. Consequently, Γ does not contain (0, 0). The points of Γ are incomparable, that is, if $\tau_1 < \tau_2$, then $n_1 \ge n_2$. For any point $(n, \tau) \in \mathcal{E}$ there exists another point $(n_1, \tau_1) \in \Gamma$ where $n_1 < n, \tau_1 < \tau$, which corresponds to a better prediction. Therefore, the total error set \mathcal{E} contains a minimum set of best and incomparable strategies. The number of these strategies is infinite, and they are described by an error diagram Γ (Fig. 17).

In order to be able to compare strategies, we choose some one-dimensional characteristic $\gamma = \gamma(n, \tau)$ that is a function of (n, τ) . We will call it a loss function if γ is increasing in each argument. Typical examples of γ that have been employed at the research phase of prediction are functions of the form $\gamma_1 = n + \tau$, $\gamma_2 = \max(n, \tau)$, $\gamma_3 = \tau/(1-n)$, $\gamma_4 = n/\theta(\tau_0 - \tau)$, or $\gamma_5 = \tau/\theta(n_0 - n)$, where $\theta(x) = 1$ for x > 0 and $\theta(x) = 0$ otherwise. The strategy π^* will be called γ -optimal, when it minimizes $\gamma(n_{\pi}, \tau_{\pi})$. For example, the optimal strategy minimizes the mean prediction error $(n + \tau)/2$ in the γ_1 case and optimizes the number of successes at a given level of alert time in the γ_4 case. Knowing the Γ diagram, one can easily find the errors of the γ -optimal strategy graphically. We assume that the sets of levels of γ , $A_u = \{(n, \tau) : \gamma(n, \tau) < u\}$ are convex for any level u. The sets A_u are increasing with increasing u. Obviously, there is a critical level u_* where A_u and $\mathcal{E}(J)$ touch each other. Since A_u and $\mathcal{E}(J)$ are convex, this will be a single point (the regular case) or a line segment (not a typical case). The point of contact $Q^* = (n^*, \tau^*)$ will determine the γ -optimal errors. By construction it belongs to the Γ error diagram. It is easily seen that any point of the Γ curve can be made γ -optimal by a suitable choice of the loss function. This can be demonstrated as follows. The Γ curve is always on one side of its tangent. Let $a(n - n^*) + b(\tau - \tau^*) = 0$ be the equation of the tangent to Γ at the point (n^*, τ^*) . Then a and b have the same sign, since Γ is decreasing along the n-axis, while $f = |a|n + |b|\tau$ is the desired loss function for the point (n^*, τ^*) .

To sum up, if there is no ideal forecast with zero errors, the Γ curve consists of a continuum of points corresponding to γ -optimal strategies. The absence of a universal prediction strategy for a given information flow was far from being quickly grasped in prediction practice.

When the information flow is updated, $J \subset J'$, the set $\mathcal{E}(J)$ expands; for this reason the diagram $\Gamma(J')$ will be below (to be more accurate, not higher) $\Gamma(J)$. The $\Gamma(J)$ curve can be regarded as characterizing the limiting capability of information J in the prediction of large events in region G. The paradox here consists in the fact that the Γ curve always includes the end points (0, 1) and (1, 0) corresponding to the trivial strategies of an optimist and a pessimist. These ignore all information and become optimal with a special choice of the loss function. For example, looking from the economic point of view, there is no sense in predicting seismic events where there is no threat to economy and population.

D.2 The Optimal Prediction Strategy

We will try to find out what is the structure of the optimal prediction strategies based on the information flow J(t). To do this, we define the *hazard function* r(t), which is the conditional (with respect to the information J(t)) rate of predicted events:

$$r(t) = \text{Prob}\{\text{an event occurs in } (t, t + \delta) | J(t) \} / \delta$$

The symbol λ above stands for the unconditional rate of large events, i.e., $E\delta N(t)/\delta = \lambda$.

The statement that follows provides a description of the optimal strategy with errors $(n^*, \tau^*) \in \Gamma$. The description is not unique.

Statement 1. If the flow (n(t), J(t)) is stationary and ergodic, then there exists a threshold r^* depending on the loss function γ such that the optimal prediction strategy declares an alert every time when $r(t) > r^*$. In rare cases in which the relation $r(t) = r^*$ has a nonzero probability, an alert is selected with some probability p^* . If Q^* is the point where the isoline $\gamma = \gamma^*$ touches the error curve Γ , then the threshold r^* is expressed in terms of derivative common for Γ and lines $\gamma = \gamma^*$ at Q^* .

$$r^* = -\lambda \frac{dn}{d\tau}(Q^*). \tag{14}$$

If one of the curves is not differentiable at Q^* , then the derivative is the slope of any straight line tangent to Γ at Q^* .

The statement is remarkable in that it holds for a very broad class of the processes (n(t), J(t)). Nevertheless, the proof [Mo] is quite elementary and reduces to the classical Neyman-Pearson lemma in the statistical theory of hypothesis testing. The generality results from the choice of the class of loss functions which involve two prediction characteristics only, n and τ . It is just statistics such as these which are considered at the research phase of prediction. This is not at all sufficient from the practical point of view. For example, the goal function of the form $\gamma(n, \tau)$ ignores the rate of points where the strategy changes state. Frequent changes from alert to non-alert make for lower trust in the prediction involved. This circumstance is well known from practical forecasts of aftershocks, the population beginning to ignore seismologists' warnings when they frequently call short-lived alerts.

Example. Consider $\gamma = \alpha \lambda n + \beta \tau$. This loss function can be given an economic meaning, even though a naive one. Let α be the mean loss prevented by successful prediction. The use of a strategy with errors (n, τ) will fail to predict λn events per unit time. Consequently, $\alpha \lambda n$ will give the loss per unit time resulting from failures-to-predict. The quantity $\beta \tau$ gives the loss due to alerts per unit time, when β denotes the cost of maintaining the state of alert per unit time. It follows that γ gives the loss per unit time. The example of a loss that is linear in n and τ is important in that one can find the optimal strategy for it without knowing the relevant error diagram Γ . This can be demonstrated as follows. The isolines $\gamma = c$ form a set of parallel straight lines with the slope $dn/d\tau = -\beta(\alpha\lambda)^{-1}$. The latter determines the slope of the tangent line to Γ at the point Q^* . Therefore, the use of (14) gives the optimal γ -strategy as

$$\pi(t) = \begin{cases} 1 & r(t) > \beta/\alpha \\ 0 & r(t) < \beta/\alpha. \end{cases}$$
(15)

We have assumed that $P(r(t) = \beta/\alpha) = 0$, which generally holds.

Relation (15) is highly important for the analysis of the earthquake prediction problem as a whole. In the present case it separates into two independent problems. The one (seismological) reduces to estimation of the hazard function r(t), while the other (economic) problem is to estimate the economic parameter β/α . Nevertheless, it is important during the research phase of prediction to know at least the order of β/α , since one can hardly hope to get stable estimates of r in the entire range of values.

D.3 Statistical Problems

D.3.1 The Performance of Prediction Algorithms

Intermediate-term prediction techniques recently developed actually address the theoretical problem of whether earthquakes are predictable. Therefore, the techniques mostly reduce to the simplest two-phase alert, as characterized by the errors (n, τ) . The positive answer to that question will be found in the proof that the error diagram Γ is significantly different from the straight line $n + \tau = 1$. The diagram can be estimated by the lower



Figure 18: Comparison of algorithms by error diagrams Γ_A . The *solid line* and *dash-dotted* line are error diagrams for two algorithms. The line (n^*, τ^*) is the common tangent line for these diagrams

bound to the convex hull of points $(n, \tau)_A$ relevant to different prediction algorithms $\{A\}$ based on the same data set, the same prediction domain, and the magnitude range of large events.

The above idea can be used for comparisons among algorithms. Most algorithms involve internal parameters that are subsequently held fixed in an arbitrary manner. Varying the essential parameters θ of an algorithm A, one gets an error set $(n, \tau)_{\theta}$. Considering again the lower bound of its convex hull, one arrives at the error curve Γ_A representing the prediction power of the algorithm based on the data set chosen. Suppose the curves Γ_A for two algorithms (Fig. 18) intersect at an intermediate point (the end points are always the same). Let Γ_{A_1} and Γ_{A_2} have a common tangent of slope -p(Fig. 18). When the goal is a linear loss of the $\gamma = an + b\tau$ type, then it follows from Fig. 18 that A_1 is to be preferred when a/b > p and A_2 otherwise (a/b < p).

D.3.2 Estimation of (n, τ)

Statistical estimators of (n, τ) are generally unstable owing to the short history of a forward forecast. When the errors (n, τ) refer to a time-space forecast, the value of τ measures the relative space-time occupied by alerts. Unjustified extension of the space by adding aseismic areas can make τ as small as one likes. One way out of this difficulty is to collect estimates of (n, τ) for prediction algorithms having a common prediction space and a common magnitude range of large events. Anomalies in the (n, τ) estimates become evident in the (n, τ) -diagram.

D.3.3 Estimation of r(t)

The amount of information provided by J(t) as to the appearance of a large event in the interval δt is given by the Shannon quantity $I = \ln PG(J(t))$, where

$$PG(J(t)) = \operatorname{Prob}(\delta N(t) = 1|J(t))/\operatorname{Prob}(\delta N(t) = 1) = r(t)/\lambda$$

Aki [Ak] has called PG(J) the probability gain and put forward the following prediction program based on a combination of precursors:

- choose simple, weakly correlated and sufficiently informative precursors $A = \{A_1, ..., A_k, ...\},\$

- estimate the quantity $PG(A_k)$ for each,

- find PG(A(t)) for the whole set of precursors $A(t) = (A_{j_1}, ..., A_{j_k})$ observed up to the time t from the relation

$$PG(A(t)) = q_t PG(A_{j_1}) \cdots PG(A_{j_k}) \tag{16}$$

The factor q_t is related to precursors that have not been observed by time t. The above program was implemented for the Caucasus region [SCZ]. Unfortunately, this method of estimating r(t) seems oversimplified. The equality (16) means that the precursors $A_1, ..., A_n$ are conditionally independent with respect to the event $\{\delta N(t) = 1\}$. That is impossible physically, when the $\{A_k\}$ are really precursors, even though independent ones. Consider a formal example: suppose ξ_1 and ξ_2 are independent random variables and a large event occurs, when $|\xi_1 + \xi_2 - 1| < \varepsilon$. Hence $\xi_1 + \xi_2 \simeq 1$ in the conditional situation.

Vere-Jones [VJ], Utsu [Ut] and Aki [Ak] considered the estimation of r(t) as the foremost task of a prediction specialist. This only in part is true (see Statement 3). The estimation of r(t) in the entire range of values is unstable. This has been demonstrated by using the simple problem of the characteristic earthquake prediction. Stability is potentially possible in the case of an "academic" forecast involving two alert states, because exact knowledge of r(t) is needed in the neighborhood of a fixed level. The number of such levels in a realistic situation have to be greater than two or even infinitely many (Statements 2 and 3). For this reason special importance will be attached to the maximum possible reduction of dimensionality for the prediction functionals. The best solution to that problem today is provided by the M8 algorithm.

An important example of estimating $r(\cdot)$ is statistical modeling of earthquake catalogs dating back to Hawkes [Ha] and Kagan [Ka1]. We are speaking of the so-called self-exciting model in which events x = (time t, magnitude M, location g) divide into "main" and "offspring". The main ones make a Poisson process with the rate $\lambda_0(x)$; once occurring, any event x_i will generate offspring events which form a Poisson process with the rate $\lambda(x|x_i)$ (note that $\int \lambda(x|x_i) dx \leq \rho < 1$). Reproduction of any event (both a main one or offspring) occurs once and independently of one another. Consequently, the probability of finding an event on the interval dx is r(x) dx with

$$r(x) = \lambda_0(x) + \sum_{t(x_i) < t(x)} \lambda(x|x_i)$$
(17)

where t(x) is the time coordinate of the point x.

The offspring events with a common main event as progenitor x_0 play the part of aftershocks of x_0 , while offspring events for x can be considered as direct aftershocks of x. The properties of aftershocks are well known: the Gutenberg-Richter law for the magnitudes and the Omori law for the time decay. The properties of direct aftershocks are not known, and this causes difficulties and ambiguities in the parameterization of $\lambda(x|x_i)$ (compare the solutions of this problem in [KK] and [KJ]).

References

- [AEH+76] C. R. Allen (Chaiman), W. Edwards, W. J. Hall, L. Knopoff, C.B. Raleigh, C. H. Savit, M. N. Toksoz, and R. H. Turner. Predicting earthquakes: A scientific and technical evaluation — with implications for society. Panel on Earthquake Prediction of the Committee on Seismology, Assembly of Mathematical and Physical Sciences, National Research Council, U.S. National Academy of Sciences, Washington, D. C., 1976.
- [SSS99] L. R. Sykes, B. E. Shaw, and C. H. Scholz. Rethinking earthquake prediction. Pure appl. Geophys., 155: 207–232, 1997.
- [BL85] W. H. Bakun and A. G. Lindh. The Parkfield, California, earthquake prediction experiment. Science, 229: 619–624, 1985.
- [KB90A] V.I. Keilis-Borok. The lithosphere of the Earth as a nonlinear system with implications for earthquake prediction. *Rev. Geophys.* 28: 19–34, 1990.
- [KMU99] V. G. Kossobokov, K. Maeda, and S. Uyeda. Precursory activation of seismicity in advance of Kobe, 1995, M =7.2 earthquake. *Pure and Appl. Geophys.*, 155: 409– 423, 1999.
- [HC94] R. E. Habermann and F. Creamer. Catalog errors and the M8 earthquake prediction algorithm. Bull. Seism. Soc. Am., 84: 1551–1559, 1994.
- [She87] P. N. Shebalin. Compilation of earthquake catalogs as the task of clustering with learning. Transactions (Doklady) Acad. Sci. SSSR, 292: 1083–1086, 1987.
- [She92] P. N. Shebalin. Automatic duplicate identification in set of earthquake catalogues merged together. U.S. Geol. Surv. Open-File Report 92-401, Appendix II, 1992.

- [SC84] D. P. Schwartz and K. J. Coppersmith. Fault behavior and characteristic earthquakes: examples from the Wasatch and San Andreas fault zones. J. Geophys. Res., 89: 5681–5698, 1984.
- [SCZ] G. A. Sobolev, T. L. Chelidze, A. D. Zavyalov, L. B. Slavina, and V. E. Nicoladze. Maps of expected earthquakes based on cobmination of parameters. *Tectonophysics*, 193: 255–265, 1991.
- [VJ] D. Vere-Jones. Earthquake Prediction A Statistician's View. J. Phys. Earth., 26: 129–146, 1978.
- [Ut] T. Utsu. Probabilities in earthquake prediction. Zizin, J. Seismol. Soc. Jpn., 30: 179–185, 1977.
- [Ak] K. Aki. A probabilistic synthesis of precursory phenomena. In D. V. Simpson and P. G. Richards, editors, *Earthquake Prediction. An International Review*, Manrice Ewing Ser., Vol.4, pp.566-574, AGU, Washington, D. C., 1981.
- [Ha] A. G. Hawkes. Spectra of Some Self-exciting and Mutually Exciting Point Processes Biometrica, 58: 83–90, 1971.
- [Ka1] Y. Kagan. On a Probabilistic Description of the Seismic Regime. *Fizika Zemli*, 4: 110–123, 1973.
- [KK] A. A. Khokhlov and V. G. Kossobokov. Seismic Flux and Major Earthquakes in the Northwestern Pacific. (in Russian), *Transactions (Doklady) of Russian Acad. Sci.*, 325(1): 60–63, 1992.
- [KJ] Y. Y. Kagan and D. D. Jackson. Seismic Gap Hypothesis: Ten Years After. J. Geophys. Res., 96: 21419–21431, 1991.
- [Mo] Molchan G., Earthquake Prediction Strategies: A Theoretical Analysis. In V.I. Keilis-Borok and A.A. Soloviev (Eds.) Nonlinear Dynamics of the Lithosphere and Earthquake Prediction. Springer-Verlag, Berlin Heidelberg, 2003, pp. 337.
- [Bi] P. Billingsley. Ergodic theory and information. Wiley, New York, 1965.
- [GGK+76] I. M. Gelfand, Sh. A. Guberman, V. I. Keilis-Borok, L. Knopoff, F. Press, E. Ya. Ranzman, I. M. Rotwain, and A. M. Sadovsky. Pattern recognition applied to earthquake epicenters in California *Phys. Earth and Planet. Inter.*, 11: 227–283, 1976.
- [Tu] Tuckey, J.W., 1977. Exploratory data analysis. Addison-Wesley Series in Behavioral Science: Quantitative Methods, Reading, Mass. Addison-Wesley.

- [KBK84] V. I. Keilis-Borok and V. G. Kossobokov. A complex of long-term precursors for the strongest earthquakes of the world. In *Proceedings 27th Geological congress*, 61, pp. 56–66, Nauka, Moscow, 1984.
- [KBK87] V. I. Keilis-Borok, and V. G. Kossobokov. Periods of high probability of occurrence of the world's strongest earthquakes. In *Computational Seismology*, 19, pp. 45–53, Allerton Press, 1987.
- [Kos86] V. G. Kossobokov. The test of algorithm M8. In M. A. Sadovsky (Ed.). Algorithms of long-term earthquake prediction., pp.42–52, CERESIS, Lima, Peru, 1986.
- [M8] V. I. Keilis-Borok and V. G. Kossobokov. Premonitory activation of earthquake flow: algorithm M8. Phys. Earth Planet. Int., 61: 73–83, 1990.
- [RK96] L. L. Romashkova, and V. G.Kossobokov. Sources concentration parameters in an intermediate-term earthquake prediction algorithm. In V.I.Keilis-Borok and G.M.Molchan, editors, *Modern Problems of Seismology and Earth Dynamics, Comput. Seismol.* Vol. 28, pp. 56–66, Nauka, Moscow, 1996 (in Russian).
- [KKS90] V.G. Kossobokov, V.I. Keilis-Borok, and S.W. Smith. Localization of intermediate-term earthquake prediction. J. Geophys. Res., 95(B12): 19763–19772, 1990.
- [HKD92] J. H. Healy, V. G. Kossobokov, and J. W. Dewey. A test to evaluate the earthquake prediction algorithm, M8 U.S. Geol. Surv. Open-File Report 92-401, 23 pp. with 6 Appendices, 1992.
- [LV92] T. Levshina and I. Vorobieva. Application of Algorithm for Prediction of a Strong Repeated Earthquake to the Joshua Tree and Landers. In *Fall Meeting AGU, Ab*stracts, p. 382, 1992.
- [Vo] I. A. Vorobieva. Prediction of second large earthquake. Phys. Earth. Planet. Inter., 111: 197-206, 1990.
- [Vor94] I.A.Vorobieva. Prediction of a reoccurrence of large earthquakes based on the aftershock sequence of the first large earthquake. In *Seismicity and Related Processes* in the Environment. pp. 33–37, Russ. Acad. Sci., Moscow, 1994.
- [VL94] I. A.Vorobieva and T. A. Levshina. Prediction of the second Large Earthquake based on aftershock sequence. In *Computational Seismology and Geodynamics*, Vol 2, pp. 27–36, AGU, Washington, D. C., 1994.
- [Bat65] M. Bath. Lateral inhomogeneties of the upper mantle. *Tectonophysics*, 2: 483– 514, 1965.
- [Ver69] D. Vere-Jones. A note on the statistical interpretation of Bath's law. Bull. Seismol. Soc. Amer., 59: 1535–1541, 1969.

- [Pro78] A.G. Prozorov. A statistical analysis of P-wave residuals and the prediction of the origin times of strong earthquakes. In V.I. Keilis-Borok, editor, *Earthquake Prediction and the Structure of the Earth, Comput. Seismol.* Vol. 11, pp. 4–18, Allerton Press, N.Y., 1987.
- [RJ89] P. A. Reasenberg and L. M. Jones. Earthquake hazard after a main shock in California. *Science*, 243: 1173–1176, 1989.
- [Mat86] R. S. Matsu'ura. Precursory quiescence and recovery of after- shock activities before some large aftershocks. Bull. Earth. Res. Inst. University of Tokyo, 61: 1–65, 1986.
- [HC90] R. E. Haberman and F. H. Creamer. Prediction of large aftershocks on the basis of quiescence. The 7th US - Japan Seminar on Earthquake Prediction, Vol.1, pp. 93–96, 1990.
- [GZK+80] A. D. Gvishiani, A. V. Zelevinsky, V. I. Keilis-Borok, and V. G. Kosobokov. Computational Seismology 13: 30–43, Allerton Press Inc, N.Y., 1980.
- [Mol97] G. M. Molchan. Earthquake Prediction as a Decision-making Problem. *Pure and Appl. Geophys.*, 149: 233-247, 1997.
- [AKB⁺84] C. Allen, V. I. Keilis-Borok, I. V. Kuznetsov, and I. M. Rotwain. Long-term earthquake prediction and self-similarity of seismic precursors. In Achievements and Problems of Modern Geophysics, pp.152–165, Nauka, Moscow, 1984 (in Russian).
- [KBK⁺88] V. I. Keilis-Borok, L. Knopoff, I. M. Rotwain, and C. R. Allen. Intermediateterm prediction of occurrence times of strong earthquakes. *Nature*, 335(6192): 690– 694, 1988.
- [CN] V. I. Keilis-Borok and I. M. Rotwain. Diagnosis of time of increased probability of strong earthquakes in different regions of the world: algorithm CN. *Phys. Earth Planet. Int.*, 61: 57-72, 1990.
- [RN99] I. M. Rotwain and O. V. Novikova. Performance of the earthquake prediction algorithm CN in 22 regions of the world. *Phys. Earth. Planet. Inter.*, 111: 207–213, 1999.
- [MD] G. M. Molchan, O. E. Dmitrieva, I. M. Rotwain, and J. Dewey. Statistical analysis of the results of earthquake prediction, based on burst of aftershocks. *Phys. Earth. Planet. Inter.* 61: 128–139, 1990.
- [VP93] I. A. Vorobieva and G. F. Panza. Prediction of the occurrence of related strong earthquakes in Italy. *Pure Appl. Geophys.*, 141(1): 25–41, 1993.

[Aft] Gardner, J., and Knopoff, L. Is the sequence of earthquakes in southern Califonria, with aftershocks removed, Poissonian? Bull. Seismol. Soc. Amer., 64, No.5, 1363-1367, 1974.