Geophys. J. Int. (1997) **131**, 505–525

SPECIAL SECTION—ASSESSMENT OF SCHEMES FOR EARTHQUAKE PREDICTION

Are earthquakes predictable?

Yan Y. Kagan

Institute of Geophysics and Planetary Physics, University of California, Los Angeles, CA 90095-1567, *USA. E-mail: kagan@cyclop.ess.ucla.edu*

Accepted 1997 August 18. Received 1997 August 15; in original form 1997 January 31

SUMMARY

The answer to the above question depends on the definition of earthquake prediction. We discuss several definitions and possible classifications of earthquake prediction methods. We also consider various measures of prediction efficiency, review several recent examples of earthquake prediction, and describe the methods that can be used to verify prediction schemes. We conclude that an empirical search for earthquake precursors that forecast the size of an impending earthquake has been fruitless. Despite considerable effort in several countries, no statistically rigorous validation of proposed precursory phenomena is available; therefore, reported cases of precursors can be explained by random noise or by chance coincidence. We present evidence that earthquakes are non-linear, chaotic, scale-invariant phenomena. The most probable consequence of earthquake self-similarity is a lack of earthquake predictability as popularly defined, that is a forecast of a specific individual earthquake. Many small earthquakes occur throughout any seismic zone, demonstrating that the critical conditions for earthquake nucleation are satisfied almost everywhere. Apparently, any small shock can grow into a large event. Thus, it is likely that an earthquake has no preparatory stage. This sceptical view of current earthquake prediction efforts should not be interpreted as a statement that any further attempts to mitigate the destructive effects of earthquakes are futile. The seismic-moment conservation principle, when combined with geodetic deformation data, offers a new way to evaluate the seismic hazard, not only for tectonic plate boundaries, but also for areas of low seismicity, that is the interiors of continents. Earthquake clustering with a power-law temporal decay (Omori's law) can be used to estimate the rate of future earthquake ocurrence. Real-time seismology can facilitate relief efforts after large earthquakes and eventually provide an immediate warning of severe shaking a few seconds or tens of seconds before the shaking starts.

Key words: earthquake prediction, fractals, seismicity, statistical methods.

1 INTRODUCTION

The results of efforts to develop earthquake prediction methods over the last 30 years have been disappointing: after many monographs and conferences and thousands of papers we are no closer to a working forecast than we were in the 1960s (Geller 1991, 1996a, 1997; Geller *et al.* 1997; Kagan & Jackson 1994b). (We use the words *prediction* and *forecast* as synonyms in this paper; moreover, the term *precursor* is used for phenomena which predict the size of a future earthquake.) There are several reasons for prediction difficulties. (1) Earthquake prediction has not been properly defined. The prediction of a 'specific' earthquake in a certain time, space and magnitude window (Wallace, Davis & McNally 1984) is probably impossible because of the continuum and fractal nature of earthquake statistical distributions. (2) No acceptable criteria exist for validating an earthquake prediction (Molchan & Kagan 1992). The frequently mentioned failure-to-predict versus false-alarm criterion is unsatisfactory, since a trivial strategy of alarm declaration for the entire region leads to zero errors (see discussion below, Section 3.1). Furthermore, in using this criterion it is impossible to distinguish a method with real predictive skill from a technique that owes its success to chance coincidence—the null hypothesis is not easily defined. (3) Most earthquake-forecast attempts are not rigorously testable due to an insufficient number of predicted events, or prediction ambiguities, or a prediction time window too large for the testing to be feasible.

The earthquake process is multidimensional. Thus, its prediction requires specifying at least the space, time and magnitude aspects of an earthquake. The uncertainty in prediction parameters makes it imperative to formulate forecasts in terms of probabilities. Thus, earthquake prediction is a largely statistical problem: failure to appreciate this is at the root of many difficulties in prediction and analysis (*cf.* Vere-Jones 1995). 'Deterministic predictions' are not realistic (see below) and forecasts of future seismicity should be statistical. Furthermore, all earthquake forecasts (whether deterministic or statistical) need to be evaluated statistically to see if their success could be due to chance.

It is widely accepted that the extensive efforts of the last 30 years to find 'reliable' earthquake prediction methods—the efforts which culminated in the Parkfield prediction experiment (Roeloffs & Langbein 1994 and references therein) in the USA and the Tokai experiment (Mogi 1995) in Japan—have largely failed (Jordan 1997; Scholz 1997; Kossobokov, Healy & Dewey 1997). What are the fundamental reasons for such a failure? Why did the obvious inadequacy of available earthquake theories and the failure of previous 'predictions' not become clear to the geophysical community earlier? The answer to the first question lies in the non-linear, scale-invariant features of earthquake occurrence—properties of seismicity not fully recognized until the 1980s. However, the latter failure can be explained by the general deficiency of the models and techniques proposed to forecast earthquakes. Generally the models have not been falsifiable (Popper 1980; Engelhardt & Zimmermann 1988), and efforts to validate and test these prediction algorithms rigorously have not been encouraged. In effect, earthquake prediction efforts and, in general, the study of earthquake occurrence have been carried out as a qualitative, verbal, descriptive, 'story-telling' exercise (Morowitz 1996), rather than a quantitatively predictive science, subject to precise mathematical and statistical verification.

The dual absence of a comprehensive theory for earthquake occurrence and rigorous validation efforts is the reason that, contrary to the practice of other 'hard', quantitative sciences, earthquake prediction efforts have been judged by a committee of experts. The National Earthquake Prediction Evaluation Council (NEPEC—see for example Lomnitz 1994, p. 34) in the USA, the Committee of senior scientists in Japan (Lomnitz 1994, p. 270) and the IASPEI Sub-commission (Wyss 1991; Wyss & Dmowska 1997) evaluate proposed methods for earthquake prediction. These assessments are not based on a falsifiable test of a method's performance, but on the plausibility of physical models and other assumptions and on the review of a few case histories. Such expert reviews are insufficient for validating a prediction scheme—plausible theoretical considerations and meticulous execution of experiments do not guarantee the method's success, and vice versa. As Rhoades & Evison (1989) and Evison & Rhoades (1994) point out, the reliability of a prediction model is indicated not by the statistical significance of the data upon which it is based, but by its performance when tested against independent, preferably future, data.

Although we try to discuss as many recent efforts at earthquake prediction as reasonably possible, this article is no comprehensive review. We focus instead on the predictability of earthquakes and on the theoretical and practical reasons for the difficulties that prediction research experiences. The history of earthquake prediction efforts is reviewed by Lomnitz (1994), Aki (1995) and Geller (1996a, 1997); recent, reasonably complete, reviews are provided by Turcotte (1991) and Agnew & Ellsworth (1991). Vere-Jones (1995) discusses the statistical aspects of prediction efforts.

This paper considers only the scientific aspects of the prediction *methods*; predictions of isolated individual earthquakes (see Geller 1997) are not discussed. We need first to discover prediction algorithms with a predictive power or predictive skill, that is ones that perform better than a random guess (the null hypothesis). Thereafter, we can decide whether such schemes are practical. Molchan & Kagan (1992) and Molchan (1997) consider the application problem from a decision-theoretical point of view.

Implicit assumptions in earthquake prediction research have been that prediction is possible and that one need only find techniques and methods appropriate for solving the problem. We argue that it is not yet clear that certain types of predictions are feasible. Thus, earthquake predictability needs to be treated as an open problem, subject to careful and rigorous investigation.

In Section 2 we consider definitions of earthquake prediction and several classes of prediction. Section 3 is dedicated to describing prediction testing, and Section 4 discusses recent results on the scale-invariance of the earthquake process and its implications for predictability. In Section 5 (Discussion) we summarize the scientific challenges of earthquake prediction research.

2 PREDICTION TYPES

2.1 Prediction definition

There are several definitions of earthquake prediction. Nishenko's (1989) definition is essentially the same as that proposed by the US National Research Council (NRC) in 1976 (see Wallace *et al.* 1984): 'Earthquake prediction refers to the specification of the expected magnitude, geographic location and time of occurrence of a future event with sufficient precision that the ultimate success or failure of a prediction can be evaluated'. This definition has several defects which contribute to confusion and difficulty in prediction research.

(1) The above definition, as well as the discussion by Wallace *et al.* (1984), confirms that major effort has been directed towards predicting a *specific* individual earthquake, in other words a forecast is envisioned as the determination of parameters for the *next* strong earthquake to occur in a region. Both prediction experiments—Parkfield and Tokai (see above)—are designed to forecast an isolated individual earthquake. It is assumed that such an earthquake would be recognized, although no formal algorithm is proposed to identify it, for the obvious reason that such a rule is impossible to formulate. The next event in a 1-D process is self-evident. However, the multidimensional nature of the earthquake process makes the definition of the next event impossible (Vere-Jones 1995), unless the process is made 1-D. Such a radical transformation is attempted by the seismic-gap/characteristic-earthquake hypothesis (Nishenko 1989; Aki 1995; Kagan 1996a). We discuss this hypothesis below.

(2) This definition does not specify how the prediction is to be validated, that is what criteria are needed to evaluate its

effectiveness, how to test whether the predicted event could occur by chance, etc.

(3) The earthquake focal mechanism should be included in the list of earthquake parameters to be forecast. Modern earthquake catalogues routinely include the results of seismic moment tensor inversions (Dziewonski, Ekström & Salganik 1996; US Geological Survey 1996). Time–space–size and focal mechanism constitute the major parameters of an earthquake. Their specification allows us to calculate static deformation or low-frequency seismograms for an event.

In considering earthquake predictions, it is often implicitly assumed that 'large' earthquakes—those which are potentially damaging—can be forecast. However, because of earthquake self-similarity (see Section 4 below), it would be a major achievement, from a scientific point of view, to predict even small events.

We define an earthquake prediction as a (probabilistic) statement about future earthquakes, which reduces the uncertainty of their occurrence compared to present knowledge. The statement must be statistically testable in a reasonable amount of time. Ideally, the time-dependent prediction can be defined as a formal rule (usually a computer algorithm) that predicts the rate of occurrence of earthquakes over some multidimensional interval of time, space and the seismic moment tensor. This rule should yield results significantly better than the Poisson estimate of the rate of occurrence (or the appropriate null hypothesis, see below). Most prediction efforts considered in this paper and elsewhere (see the references above) are based on simpler ideas. To bring some order into this diverse field, we try to classify earthquake forecast techniques.

2.2 Prediction classification

Several criteria can be used to classify earthquake prediction; we list and discuss some of them below.

2.2.1 Existence of quantitative theory

The great majority of proposed methods for earthquake prediction lack quantitative theory, thus they should be classified as empirical methods. Among these are earthquake clustering, in particular foreshock–main shock–aftershock sequences, seismicity variations, changes in seismic velocities, precursory strain, anomalous animal behaviour, variations in geochemical, hydrological and electromagnetic signals, etc. (Nishenko 1989; Ma *et al.* 1990; Lomnitz 1994). However, the only quantitative forecasting method currently available with a significant predictive skill is based on earthquake clustering (see Section 3.3.5).

Two quantitative methods have been developed in recent decades which significantly advance our understanding of earthquake occurrence. (1) Global plate tectonics explains the strain accumulation on and near plate boundaries. In addition, due to recent technological advances, tectonic deformation can be measured by various geodetic and geological methods. Such measurements are becoming available even for intracontinental areas. (2) Linear elasticity theory predicts the static and low-frequency deformation due to earthquakes in the far field. Thus, the release of accumulated strain by earthquakes can be evaluated. The availability of these numerical methods has contributed to the development of several earthquake models such as recurrence models, seismic cycles, characteristic earthquakes and seismic gaps. However, the interface between strain accumulation and its release is complicated due to the stochastic nature of seismicity. Thus, as discussed in the following sections, these simple earthquake recurrence models can be rejected by testing them against empirical evidence.

2.2.2 Temporal classification

Predictions are frequently classified as short-term (up to a few months), intermediate-term (from one year to a decade) or long-term (a few decades or longer), according to the forecast lead time. However, these timescales are defined in varying ways, for example in terms of the technique used (Knopoff 1996) or of available earthquake mitigation measures (Wallace *et al.* 1984). Thus, this classification is not intrinsic. Due to the scale-invariance of seismicity (Section 4), it is impossible to define the 'natural' scale for earthquake temporal features; however, two physical scales may be suggested: one connected to the propagation of elastic waves and earthquake rupture, and the other related to the velocity of tectonic deformation. The size of the earthquake focal area or of the zone of intense shaking suggests that the first scale is of the order of seconds or tens of seconds. Since accumulated strain is released mostly by the largest earthquakes (Section 4.3.1), the second timescale is on the order of decades or even millenia.

2.2.3 Prediction-window specification

The least informative prediction indicates a time–space window where earthquakes are likely to occur without any numerical specification of the probability. Since almost all the putative prediction methods are still very preliminary, this is the usual form of prediction. An example of such a forecast is the $M8$ algorithm (Keilis-Borok *et al.* 1988; Keilis-Borok & Kossobokov 1990), which identifies intervals of 'times of increased probabilities' (TIPs) for large earthquakes ($m \geq 8$ to $m \geq 7.5$), using a pattern-recognition technique. It is not clear whether such TIPs in various regions have the same probability, nor what the level of the increased probability is compared to normal seismicity.

Forecast accuracy can be improved if the probability value is indicated for selected zones of interest. An example would be Nishenko's (1991) map of the circum-Pacific regions with more than 100 seismic zones for which the probability of large earthquakes occurring is calculated using the seismic-gap hypothesis. Verification problems are often encountered at the boundaries of such zones, where a small variation of an earthquake's space–time–magnitude may determine the success or failure of the prediction (Jackson 1996b).

The method which avoids the above difficulties assigns the probability to any infinitesimal seismic region: it formulates the earthquake prediction in terms of probability density or conditional rate of earthquake occurrence (e.g. Rhoades & Evison 1979; Kagan & Knopoff 1987; Evison & Rhoades 1994; Kagan & Jackson 1995). Some combination of the above methods is also possible. For instance, the spatial boundaries of a zone can be specified, whereas the temporal probability is stated as a continuous rate of occurrence per unit time (*cf.* Bakun & Lindh 1985).

2.2.4 Prediction types

In Table 1 we list (in decreasing order of forecast precision) five prediction classes which have been employed or considered recently.

(1) Deterministic prediction: because the input data are the result of measurements, such a prediction is usually given some error bounds. Thus, the forecast should be issued in probabilistic terms. Examples are stress-accumulation models such as the time-predictable and slip-predictable schemes proposed by Shimazaki & Nakata (1980). In principle, if a deterministic algorithm is available, the whole earthquake process including the occurrence of small earthquakes and the details of large-earthquake rupture can be computed.

(2) Specific earthquake prediction: the earthquake to be predicted is regarded as a recognizable entity. Usually only large earthquakes are expected to be predicted; characteristic earthquakes are an example of such models [e.g. the Parkfield prediction, see Bakun & Lindh (1985); or the Tokai prediction, see Mogi (1995)].

(3) Prediction models which specify the magnitude of a future earthquake. A precursor is assumed to supply additional information on the size of future earthquakes. Thus, upon a specific precursor observation, the conditional magnitude distribution of predicted events differs significantly from the standard Gutenberg–Richter (G–R) relation,

$$\log_{10} N(m) = a - bm\,, \tag{1}$$

where $N(m)$ is the number of earthquakes with magnitude $\geq m$, a is the seismic productivity constant and $b \approx 1$. Examples of such methods are given in Rhoades & Evison (1979) and Weimer & Wyss (1994); in most other prospective prediction methods, it is assumed that predicted earthquakes have a special size distribution. An alternative size distribution for predicted earthquakes may specify a b-value for the G–R law which differs significantly from the usual b-value, or a change of the regular G–R relation into a different form.

(4) Time-dependent seismicity: a time-dependent earthquake rate is predicted, the magnitude is assumed to obey the standard G–R law, and the rate or the a-value in (eq. 1) varies in space and time. The rate is defined as the number of earthquakes exceeding a certain magnitude limit, per unit of time–space (the seismic-activity level). Examples include short-term prediction algorithms based on Omori's law (Kagan & Knopoff 1987; Reasenberg & Jones 1989). There are indications that focal mechanisms of earthquakes in a cluster or in aftershock sequences are correlated. Thus, the mechanisms of future events can be better predicted than by using a long-term average.

Table 1. Earthquake prediction types.

No	Prediction Name	Space $\mathbb{R}$	Time T	Focal Mechanism F	Magnitude m
1	Deterministic	+	+	+	+
2	Specific earthquake	+	+	+	+
3	Magnitude specific	+	+	+?	+
4	Time–dependent	+	+	+?	–
5	Time–independent	+	–	–	–

(5) Time-independent seismicity or earthquake potential (Wallace *et al.* 1984): a spatially inhomogeneous earthquake rate is evaluated, assuming that magnitudes follow the G–R distribution. The determination of such earthquake probability is a standard procedure in seismic-hazard analysis (Kagan & Jackson 1994a; Working Group 1995). Although most geophysicists would not consider such hazard evaluation to be a prediction, we discuss it in some detail for two reasons: (a) hazard analysis is the most important practical result that earthquake seismology offers, and (b) static seismicity estimates are necessary in almost any testing of time-varying predictions as a null hypothesis (see below).

In reality, the difference between classes (4) and (5) is not as significant as it seems. As we argue in Section 4, the temporal interaction between earthquakes is governed by power-law distributions: there is no sharp boundary between time-varying and static estimates of seismicity levels. These two prediction classes (4 and 5) can be compared to two methods of weather prediction (Murphy 1996): 'persistence' (assuming that future weather would continue recent history) and 'climatological' (that is taking a long-term average of the available record) forecasts.

There is little doubt that forecasts of the last two types provide some information on earthquake-occurrence probabilities, but it is the first three classes which are usually considered by the general public and by a majority of geophysicists to be 'real' earthquake predictions. For these prediction classes, earthquake precursors such as strain, geochemical, electric or other anomalies are assumed to exist in the 'preparation' zone of an earthquake and to carry information about the time, position and size of an impending earthquake. However, despite long and energetic efforts by researchers in several countries, investigating numerous proposed precursors, not one precursor has been demonstrated to have a statistically significant predictive power in the sense assumed by methods (1)–(3) in Table 1.

3 PREDICTION EVALUATION AND TESTING

Earthquake prediction is intrinsically a statistical problem—the evaluation of predictions should be based on objective statistical tests. Moreover, randomness is the most obvious property of earthquake occurrence. Failure to recognize these facts and to apply modern stochastic theories and statistical methods is at the root of many difficulties in validating prediction schemes. Any proposed earthquake prediction scheme should be testable or *falsifiable* to be of scientific value. Such testing is difficult because geological processes are generally slow; thus, in order to test a proposed method over a reasonable period of time additional assumptions are usually necessary. Following successful testing, prediction algorithms could be optimized. The problem has been discussed at length by Evison & Rhoades (1994) and Kagan & Jackson (1996). The recent VAN debate (Geller 1996b) shows clearly that our geophysical community lacks unambiguous methods for testing predictions.

3.1 Prediction efficiency

There is no consensus in the geophysical community about ways to characterize the efficiency or 'goodness' of earthquake prediction methods. Some ideas can be borrowed from weather-forecasting methods (Murphy 1993, Table 2), but the types of processes used in both disciplines differ. In describing seismicity, we use point stochastic processes as a basic mathematical model (Daley & Vere-Jones 1988; Vere-Jones 1995), whereas weather is described as a continuous process. Therefore, in the former case we predict the occurrence of an event or of a catastrophe, but in the latter case a continuous variable such as temperature is usually predicted (Molchan & Kagan 1992 and references therein).

What criteria can be used to evaluate an earthquake prediction method? A *predictive ratio* (Kagan & Knopoff 1977) is the ratio of the conditional probability of earthquake occurrence (according to some model to be tested and to the predictive data D available) to the same probability according to the Poisson hypothesis:

$$\eta = \mathrm{Prob}(\Delta t, \Delta M, \Delta \mathbf{x}|D)/\mathrm{Prob}(\Delta t, \Delta M, \Delta \mathbf{x}), \quad (2)$$

where Δt, ΔM, $\Delta \mathbf{x}$ are time, seismic moment and space windows, respectively. The predictive ratio has been adopted by Vere-Jones (1978), who called it a 'risk enhancement factor'. Rhoades & Evison (1979) called it a 'risk refinement factor' and Aki (1981) has called a similar quantity a 'probability gain'.

The probabilities for infinitesimal intervals are equivalent to rate intensities, that is $\lambda_1 = \mathrm{Prob}(dt, dM, d\mathbf{x}|D)$ and $\lambda_0 = \mathrm{Prob}(dt, dM, d\mathbf{x})$. For a catalogue of N earthquakes we can test whether the seismicity model yields a better approximation than a Poisson model by calculating the likelihood ratio (Daley & Vere-Jones 1988, Ch. 13):

$$L = \exp\left[-\int_\Omega \lambda_1 dt\, dM\, d\mathbf{x} + \int_\Omega \lambda_0 dt\, dM\, d\mathbf{x}\right] \times \prod_{i=1}^{N} \lambda_1(i)/\lambda_0(i), \quad (3)$$

where $\Omega = \mathbb{R} \times M \times T$, $\mathbb{R}$ is the Euclidian 3-D space, T is time, M is the scalar seismic moment and $\lambda_j(i)$ is evaluated at the point of each earthquake occurrence (see Ogata, Utsu & Katsura 1996). The expression

$$I = \frac{\log_2 L}{N} \quad (4)$$

can be interpreted as the amount of Shannon's information (in bits) about any event that an observer gains through the use of a prediction algorithm (Kagan & Knopoff 1977).

For prediction cases when the probabilities in eq. (2) are not known, two criteria are sufficient to prove the predictive value of a method (Molchan & Kagan 1992; Molchan 1997 and references therein): (a) the ratio of the total volume of alarm zones to the total volume of the region, τ; (b) the ratio of missed qualified earthquakes to the total number of such events, ν. A third criterion is frequently considered: (c) the ratio of the number of unsuccessful alarms to the total number of alarms, μ. For a simplified comparison, we put these prediction aspects into Table 2. Parameters τ, ν and μ can be considered as normalized prediction errors, their values being limited between zero and one. Reasenberg & Matthews (1988) introduced similar parameters for prediction effectiveness, validity (V) and reliability (R), which are $\nu = 1 - R$ and $\mu = 1 - V$. The average predictive ratio $\bar{\eta}$ can be defined in this case as (Molchan 1997, p. 236)

$$\bar{\eta} = (1 - \nu)/\tau\,. \quad (5)$$

Feng *et al.* (1984) proposed a *precision standard*,

$$P = 1 - \nu - \tau, \qquad 0 \le P \le 1, \quad (6)$$

to characterize the efficiency (see also Molchan & Kagan 1992). The above equation ensures that issuing random alarms or using information which is not correlated with earthquake occurrence cannot qualify as a forcasting method, since for such a method $P \approx 0$.

The ideal prediction method should have

$$\tau = \nu = \mu = 0\,. \quad (7)$$

Real (non-ideal) prediction methods cannot have the values of all errors in eq. (7) equal to zero. The best performance one should expect for a prediction technique is that the error values are sufficiently small. The predictive ratio (eq. 5) or the predictive power (eq. 6) do not depend on the μ-value; one should try to reduce the false-alarm rate (Molchan & Kagan 1992) only when it is proven that these parameters are greater than zero for a prospective predictive method. Similarly, discussion of a useful prediction method would be premature, unless its predictive skill is established.

The above discussion of prediction efficiency provides an opportunity for an additional classification of forecast methods useful in our consideration of prediction testing. Let us consider several (limiting-case) methods which have at least one of the errors significantly larger than zero. For an easier comparison, we show these prediction schemes in Table 3. Prediction class (A) corresponds to a partially ideal prediction, since it has a real predictive power ($P \to 1$) (Molchan & Kagan 1992). The short-term earthquake prediction proposed by Kagan & Knopoff (1987) can be considered as an approximation of the class (A) method (Section 3.3.5). Predictions of classes (B)

Table 2. Short definition and mutual relation of prediction characteristics.

Aspect	Symbol	Definition	Relation
Alarm size	τ	Ratio of total volume of alarm zones to total volume of region	$A = 1 - \tau$
Missed events	ν	Ratio of missed qualified earthquakes to total number of such events	$R = 1 - \nu$
False alarms	μ	Ratio of number of unsuccessful alarms to total number of alarms	$V = 1 - \mu$
Accuracy	A	Ratio of total volume of safe zones to total volume of region	$\tau = 1 - A$
Reliability	R	Ratio of successfully predicted qualified earthquakes to total number of such events	$\nu = 1 - R$
Validity	V	Ratio of number of successful alarms to total number of alarms	$\mu = 1 - V$
Predictive ratio	$\bar{\eta}$	Number of successfully predicted events relative to forecast using the null hypothesis	$\bar{\eta} = (1 - \nu)/\tau$
Precision	P	Difference between prediction errors and forecast using the null hypothesis	$P = 1 - \nu - \tau$
Likelihood ratio	L	see eq. (3)	
Information content	I	see eq. (4)	

Table 3. Earthquake prediction methods and their errors.

Prediction error	Ideal prediction	Prediction classes (A)	(B)	(C)
Alarm size (τ)	0	$\downarrow 0$	$\downarrow 0$	$\uparrow 1$
Missed events (ν)	0	$\downarrow 0$	$\uparrow 1$	$\downarrow 0$
False alarms (μ)	0	$\uparrow 1$	$\downarrow 0$	$\downarrow 0$

In the table, $\uparrow 1$ means that the value of the prediction error is approaching 1 from below, and $\downarrow 0$ means that the error is positive and close to zero.

and (C) cannot have a significant predictive power, since either v or τ is close to 1 (see eq. 6).

One of the prediction methods of class (B) is the trivial 'optimist strategy' (Molchan & Kagan 1992) of never declaring an alarm. Another example of (B) can be realized, for example, by a 'prediction' of small earthquakes. According to the G–R relation (eq. 1), the number of earthquakes increases by a factor of 10, as the magnitude decreases by one unit. Thus, one can have a very small false-alarm rate, even when issuing random alarms, since each alarm occupies a small fraction of the total volume. A prediction of class (C) can be realized by using the trivial 'pessimist strategy' (Molchan & Kagan 1992) of alarm declaration over the entire region. In such a case, there are no missed earthquakes and no false alarms. Another prediction of type (C) is to issue a few alarms which collectively occupy most of the volume available for a prediction (see Section 3.3.1 for further discussion).

3.2 Prediction testing—challenges and choices

The above comments call for further discussion of the criteria for accepting or rejecting statistical hypotheses related to earthquake forecasts. Earthquake prediction presents a special problem because no appropriate theory of earthquake occurrence exists, earthquake data is difficult to interpret and the earthquake process has many dimensions. Ideally, earthquake prediction experiments should be organized according to the rules proposed by Jackson (1996a) or Rhoades & Evison (1979, 1996) and Evison & Rhoades (1994).

3.2.1 Null hypothesis

The testing of earthquake predictions must include comparison with the null hypothesis that the claimed prediction successes are due to chance. The null hypothesis should include well-known spatial variations and temporal clustering of seismicity (Kagan & Jackson 1996). The formulation of the null hypothesis for large earthquakes ($m \geq 7$) is relatively easy, since the clustering of these earthquakes is weak, at least on the timescale of a few years. Thus, the Poisson process can often serve as a null hypothesis. However, even in this case, spatial inhomogeneity of earthquake epicentres challenges the verification process (Kagan & Jackson 1995). For moderate earthquakes ($7 \geq m \geq 5$), the simple Poisson process is no longer an acceptable null hypothesis. These earthquakes are strongly clustered, the best-known example of which is aftershock sequences. Thus, the null hypothesis should include clustering as its major feature.

Three models can thus be proposed for the null hypothesis; for all the hypotheses the magnitude is distributed according to the G–R distribution (eq. 1). (a) Zeroth-order model: earthquakes are distributed uniformly randomly in space and time, their focal mechanisms randomly rotated. Thus, earthquakes follow a spatially and temporally homogeneous Poisson process. This scheme may be an appropriate model for seismicity in intracontinental regions where the data on tectonic earthquakes are missing or insufficient. (b) First-order model: space and focal mechanisms are assumed to follow an average seismicity pattern, time is uniformly random, that is we assume that earthquakes follow a spatially inhomogeneous Poisson process. (c) Second-order model: the temporal clustering of earthquakes is modelled. The earthquake process is a spatially inhomogeneous cluster Poisson process (Kagan 1991; Ogata 1998).

A convenient artifice for constructing a null hypothesis is to randomize the times and locations of either the earthquakes or the predictions. Such randomization must be done with care, because both the earthquakes and the predictions have some statistical structure (they are not uniformly distributed in time and space), and a reasonable null hypothesis should preserve this structure. This is an especially important issue since predicted events can be either aftershocks or members of earthquake clusters. Kagan (1996b) proposed two methods for the accounting of earthquake clusters: (1) declustering an earthquake catalogue; and (2) using a null hypothesis which explicitly includes clustering ('alternative prediction').

Stark (1996) argued that since earthquake times and places are difficult to simulate accurately, it is preferable to compare the success rate of a prediction algorithm with the success rate of 'random' algorithms. Aceves, Park & Strauss (1996) randomized the prediction set. However, we argue that the statistical properties of seismicity have been studied extensively, while the predictions may have a non-uniform distribution that is not characterized sufficiently well to be randomized with confidence. Moreover, the randomization of predictions cannot effectively test the schemes which use (not always wittingly) the seismic record to forecast earthquakes (Stark 1997). If, for example, one issues alarms during periods of high seismic activity, the comparison with randomized predictions may show significant advantage to the method. However, such predictive power may be due solely to the non-randomness of the earthquake occurrence, not to the method's intrinsic effectiveness.

Mulargia & Gasperini (1992) tested predictions in forward and reverse time. They argued that if the reverse test shows a better correlation between the predictions and earthquakes, it means that the predictions have been issued preferentially after a strong earthquake has occurred. Therefore, the forecast's success is due to earthquake clustering. In earthquake sequences, a main shock is by definition stronger than aftershocks or foreshocks; thus, for numerous sequences of relatively weak events, only a main shock would have a magnitude larger than the cut-off level. An alarm issued after the first event in a cluster would probably be more 'successful' if tested in reverse than in forward time. However, hypothetically, such a test can be defeated by issuing predictions each time for earthquakes stronger than those that triggered the alarm.

Why is earthquake temporal clustering such an important issue? A prediction technique can either use earthquake-clustering information directly (for example, preferentially issuing alarms during times of high seismic activity), or co-seismic and post-seismic earthquake signals may trigger an alarm. Thus, precursory non-seismic anomalies might appear to have some predictive capability, but this would actually occur purely from earthquake clustering. A null hypothesis that neglects such clustering might perform worse than the precursory-anomalies hypothesis, even though the anomalous phenomena lack real intrinsic predictive power (Stark 1997). Until prediction methods can outperform the prediction schemes which specifically use the clustering algorithm (see Section 3.3.5), or can show that their information gain is statistically independent of earthquake clustering, these methods should be ignored.

Another problem connected with earthquake clustering is whether one can predict the magnitude of a future earthquake, aside from a trivial statement that the magnitude follows the G–R law (Section 2.2.4). As discussed in the previous paragraph, a potential method, claiming to belong to class 3 (Table 1) and showing a statistically significant advantage over a null hypothesis, may have predictive power through direct or indirect use of earthquake clustering. Whether a prospective method yields any specific information on earthquake size distribution needs to be tested.

3.2.2 *Other testing problems*

As we discuss in Section 2.2.3, sharp magnitude–space–time boundaries in the alarm zones can cause instability in statistical tests. According to formal criteria, if an earthquake epicentre falls short of the zone even by a few kilometres such an event is scored as a failure. Similarly, magnitude thresholds in a forecast are often expressed as sharp cut-offs. Small-magnitude variations often result from use of either a preliminary, not final, catalogue or differing magnitude scales. These variations can again make the difference between accepting and rejecting a hypothesis. Real seismicity lacks sharp boundaries in space and magnitude. Sensitivity to various parameters can be significantly reduced by specifying an earthquake probability density for all the magnitude–space–time windows of interest (Section 2.2.3).

What significance level should we seek when testing for potential precursors? The frequently used 5 per cent level signifies that even if the sample selection process is fair, unbiased and without hidden systematic errors, the null hypothesis would be rejected in one case out of 20 due to random fluctuations. Several hundred publications and conference presentations per year explore possible earthquake precursors. Moreover, negative results have a much lower probability of being reported, that is the sample is biased. Thus, the total number of attempts to find precursors is even higher. Although at present very few precursor reports validate their claims statistically, if the tests are carried out under the 5 per cent rule it would mean that many (tens) of precursors would have been 'confirmed' simply by chance. Therefore, the significance level selected for null-hypothesis rejection should be much lower than the usual 5 per cent (Anderson 1992; Kagan 1997b). A level of 0.5 per cent or less would be more appropriate (Aceves *et al.* 1996).

All kinds of subtle and not-so-subtle biases and systematic errors are possible if one searches for 'anomalies' or 'patterns' in a large amount of data. The recent availability of high-speed, low-cost computing makes such extensive searches feasible. Rules for identifying precursors can be adjusted to fit the data (Mulargia 1997); hidden degrees of freedom may be introduced in the testing procedure (that is the selection of region boundaries in space–time–magnitude and the selection of the earthquake database). Studies using posterior adjustment of parameters may be evaluated using statistical tests for hypotheses with *a priori* fixed parameters (Geller 1997).

A comparison with research techniques in medicine may be useful here (*cf.* Lomnitz 1994, pp. 265–266). The complexity of a human organism and the possibility of a patient–doctor–treatment interaction resulting in various biases have led medical researchers since the 1940s to 'double-blind placebo-controlled' experiments as the standard method (see e.g. White, Tursky & Schwartz 1985, in particular Chapters 3 and 5–7). Although some elements of this statistical methodology—such as the use of prediction or earthquake-occurrence randomization (see above)—can be incorporated into earthquake prediction verification, unfortunately such double-blind experiments are impossible in earthquake seismology. However, as in medical science, possible systematic errors and artefacts of various kinds must always be taken into account. At present, we do not see any effective methods for retrospective interpretation of seismic and geological data which would be completely free of possible biases.

3.2.3 *Testing methodology*

It is possible that the only way to avoid the pitfalls described above is to specify the formal rules for the forecast and for its validation in advance of the test (Jackson 1996a; Evison & Rhoades 1994; Rhoades & Evison 1996). Therefore, the following methodology can be proposed for the testing of earthquake prediction schemes.

(1) Case-history investigations: these studies should satisfy the criteria formulated, for example, by the IASPEI group (Wyss & Dmowska 1997, pp. 13–14). However, case histories of 'successful prediction' of one or of several earthquakes do not demonstrate that a method has predictive power, since such success may be due to chance or to a selection bias. Since seismicity is characterized by extreme randomness, it is possible, in principle, to select almost any pattern from large amounts of data (Kagan 1994). Only formal and rigorous tests of the statistical significance of proposed prediction algorithms prove predictive skill (Kagan & Jackson 1996).

(2) Large-scale retrospective rigorous statistical testing using a control sample, that is data which were not considered in formulating the working hypothesis and evaluation of adjustable parameters for the model. The null hypothesis should be formulated and tested against the same data.

(3) Forward prediction testing, during which no adjustment of parameters is allowed and all relevant possible ambiguities in data or the interpretation technique are specified in advance (Evison & Rhoades 1994; Kossobokov *et al.* 1997).

In future, when earthquake prediction models are tested, special attention should be paid to the formal definition of what is predicted, and how the prediction can be tested. For a prediction to be testable, it needs to be issued for as large a region as possible, even if this would entail a less efficient prediction.

3.3 Prediction-testing examples

In this section we discuss several proposed prediction techniques. They have been selected to demonstrate problems and challenges in testing earthquake forecasts.

3.3.1 VAN predictions

Varotsos *et al.* (1996a,b and references therein) claimed to predict earthquakes in Greece using 'seismic electric signals'. Their method is better known as VAN, after the initials of Varotsos, Alexopoulos & Nomicos (1981) who were the authors of the first paper on this scheme. In 1981 VAN claimed to be able to detect electrical signals 7 hr $\pm$ 30 min before every earthquake of $m \geq 2.6$ within 80 km of their observatory. By 1996, Varotsos *et al.* (1996a,b) claimed to be able to predict time–space–magnitude parameters of impending earthquakes with an accuracy in time (Δt) of the order of a few weeks, a space uncertainty (Δr) of 100–120 km and a magnitude error (Δm) of ± 0.7 units. Thus, the VAN method should be classified as being of type (3) in Table 1. The early applications of the VAN model can be classified according to class (B) in Table 3 (practically no false alarms, because the magnitude threshold is very low, see above). Apparently, in an effort to miss fewer large earthquakes, VAN modified their method to increase the time and space window, so the contemporary procedure is close to class (C) in Table 3.

The VAN debate (Geller 1996b) and a recent book (Lighthill 1996) discuss verification of the VAN predictions. Because VAN prediction parameters have not been unambiguously specified, there has been no consensus on the VAN testing results. Kagan & Jackson (1996) point out the reasons why the VAN validation cannot be carried out: (1) the VAN performance is significantly 'improved' because they issue alarms, not at random, but usually after a strong earthquake; (2) VAN 'enhance' the predictive power of the method further by using alarm zones of significant size; (3) the time and magnitude windows are not clearly stated in the prediction announcements; (4) VAN adjust their prediction rules retrospectively; and (5) they simply violate these rules. Moreover, the adjustment of prediction parameters continued even during the publication of the above volumes. For example, in their paper, submitted at the end of 1993, Varotsos *et al.* (1996b) suggested that the time window for almost all of the VAN predictions is 11 and 22 days. However, Varotsos *et al.* (1996b) have claimed success for an earthquake on 1993 June 13 that occurred over two months after the prediction was issued. Another earthquake (1994 April 16) occurred 47 days after their 'prediction'. Thus, large time windows were used even from 1994 until 1996, when the debate papers (Geller 1996b) were published. Varotsos & Lazaridou (1996, p. 1397) suggested that only 70 per cent of VAN predictions should satisfy the strict prediction rules; for the rest of the alarms the limits on Δm, Δt and Δr can be relaxed.

Varotsos *et al.* (1996b) acknowledged issuing 67 'predictions' during the 8.4 years from February 1987 to June 1995 for earthquakes with magnitude $m \geq 5$. Out of 67 forecasts 27 were double 'predictions', giving a total of 94 'predicted' earthquakes. One could completely span an 11.2 years period by issuing 67 predictions at VAN's current maximum time window of 2 months. In addition, VAN predictions are evaluated not in real time but long after the alarms end, when seismicity data are available, making it possible to select the largest earthquake in an alarm window and to adjust the time window to fit the data. Thus, VAN's published statements about their windows are based on *a posteriori* correlations with subsequent seismicity. In effect, VAN's basic procedure is to issue a 'prediction', wait until an earthquake occurs, and then claim 'success', almost regardless of the discrepancies in space, time and magnitude (Geller 1996a). These circumstances make any rigorous testing of VAN forecasts impossible and pointless.

3.3.2 Parkfield prediction

'Characteristic' $m \approx 6$ earthquakes were thought to occur on the San Andreas fault in Parkfield, California, at intervals of approximately 22 years; the last such event occurred in 1966 (Bakun & Lindh 1985). They proposed that there was a 95 per cent probability 'the next characteristic Parkfield earthquake' would occur before 1993, and the US Geological Survey established the 'Parkfield Earthquake Prediction Experiment' (Roeloffs & Langbein 1994). This 'experiment' basically consists of setting up instruments in the hope of recording precursors. Since 1985 there have been no significant earthquakes at Parkfield, but damaging earthquakes (1989 Loma Prieta, 1992 Landers, 1994 Northridge) have occurred elsewhere in California.

Lomnitz (1994) and Kagan (1997b) noted that, assuming a Poisson process with reasonable parameters, there is a high probability that apparently periodic seismicity would occur randomly somewhere in Southern or Central California. Thus, the quasi-periodic sequence of characteristic Parkfield earthquakes can be explained by selection bias. Kagan (1997b) argues that the null hypothesis (the Poisson process plus the G–R relation) is conceptually less complex than other models, and should not be rejected unless there is statistically significant evidence to the contrary. The observed magnitude–frequency curves for small and intermediate earthquakes in the Parkfield area conform to the theoretical distribution computed on the basis of a modified G–R law (modified gamma distribution—see Section 4), using deformation rates for the San Andreas fault (Kagan 1997b). According to the null hypothesis the return time for an $m \geq 6$ earthquake with an epicentre in the Parkfield area is more than 100 years. Thus, the experiment may need to run for several more decades before a moderate or large earthquake occurs in the area. With regard to statistical tests, the Parkfield experiment, by its design alone, cannot answer the question of the validity of the characteristic/quasi-periodic model. Even if an earthquake similar to the one predicted were to occur, it would not be a sufficient basis for drawing statistically significant conclusions (Savage 1993; Kagan 1997b).

The Parkfield and Tokai (Mogi 1995) experiments have not been designed to test the seismic-gap hypothesis, which is their scientific basis. On the contrary, the experiments have been planned under the assumption that the gap model is correct and ready to be implemented. Therefore, when the expected earthquakes failed to occur, the current results of the experiments seem to reject the original model (see also Savage 1993), but they cannot verify or reject the gap hypothesis—the experiments are not falsifiable.

3.3.3 Seismic gaps and characteristic earthquakes

The earthquake-recurrence hypothesis, which goes back to the elastic rebound ideas of Reid (1910), was used extensively by Kelleher, Sykes & Oliver (1973), McCann *et al.* (1979), Nishenko (1991) and others. The recurrence model has been formulated in several variants. The best known of these is the seismic-gap hypothesis adopted by the US Geological Survey as the only official procedure used to predict earthquake probabilities in California and in the circum-Pacific (Agnew & Ellsworth 1991). The seismic-gap hypothesis was developed from a qualitative model in the 1970s and early 1980s (Kelleher *et al.* 1973; McCann *et al.* 1979) to a more quantitative scheme with a numerical estimate of earthquake probabilities in gap zones (Nishenko 1991). The hypothesis uses the following assumptions: (1) earthquakes occur on geologically recognized faults, thus the locations and focal mechanisms are known; (2) large earthquakes are quasi-periodic; after a strong shock the probability of a further large event is low, and it increases as time elapses; (3) the earthquake size distribution on 'individual' faults is described by the 'characteristic earthquake' model, which states that on each fault segment a large earthquake is exclusively characteristic, and that this earthquake ruptures the entire length of the segment.

There are many publications and predictions based on recurrence models of seismicity (Agnew & Ellsworth 1991). Unfortunately, almost all share the same drawbacks as those of the Parkfield prediction discussed above: since the predictions are issued for relatively small regions, only one or a few gap earthquakes are expected to occur during the next few years. Therefore, most predictions are not testable in a reasonable time period. A series of publications by Kelleher *et al.* (1973), McCann *et al.* (1979) and Nishenko (1991) compare favourably by issuing a homogeneous set of predictions for almost the entire circum-Pacific rim. Thus, a few years after the publication of the predictions, it would be possible to make a quantitative comparison of the forecasts with the seismicity record (forward testing) and attempt to validate the gap hypothesis. Kelleher *et al.* (1973, p. 2553) suggested that 'the most realistic test [of the forecasts] will lie in the locations of large earthquakes during the next few decades'. Although the predictions are somewhat vague and ambiguous, and thus disagreement on the testing results is unavoidable (Kagan & Jackson 1991; Nishenko & Sykes 1993; Jackson & Kagan 1993; Kagan & Jackson 1995), some important conclusions can be drawn from these tests.

Probability estimates for earthquakes in zones can be evaluated using three tests (Kagan & Jackson 1995; Jackson 1996b): (1) the total number of zones filled by earthquakes, or the number of earthquakes in various zones; (2) the likelihood that the observed list of filled zones would result from a process with the probabilities specified; and (3) the likelihood ratio to that of a Poissonian null hypothesis. Earlier predictions (Kelleher *et al.* 1973; McCann *et al.* 1979), which specify only a qualitative measure of a seismic-zone hazard, can only be tested by using the first method—comparing the numbers of zones with qualified earthquakes and the number of earthquakes in various zones. Kagan & Jackson (1991) tested the hypothesis that the 'dangerous' gaps are significantly more likely (by a factor of 2 or more) to experience strong earthquakes than the 'safe' zones. We show that the hypothesis can be rejected with more than a 95 per cent confidence level, and find that strong earthquakes have occurred preferentially near the sites of previous, recent, large events. Most plate-boundary segments unruptured in a previous century remain unruptured still.

Kagan & Jackson (1995) tested Nishenko's (1989) seismic-gap model using tests (1)–(3) described above. Nishenko (1991) gave probabilities that each of about 100 zones would be filled by characteristic earthquakes during various periods beginning in 1989. The null hypothesis uses a smoothed version of seismicity since 1977 (see Section 3.3.6 below) and assumes a G–R magnitude distribution (eq. 1). Kagan & Jackson (1995) used both the Harvard centroid moment tensor (Dziewonski *et al.* 1996) and the preliminary determination of epicentres (US Geological Survey 1996) catalogues in testing. Since Nishenko's forecast did not specify a clear relationship between the characteristic earthquake magnitude and the threshold magnitude for a successful prediction, Kagan & Jackson (1995) also used several different magnitude cut-offs in the tests. Using a strict interpretation that only earthquakes equal to or larger than the characteristic magnitude should be counted, the PDE catalogue in 1989–1996 shows only three qualifying earthquakes in the entire area covered by the forecast. For the Harvard catalogue the number is five. The predicted number (Nishenko 1991) for this period is 12.7, and the discrepancy is too large to result from chance at the 98 per cent confidence level. The new (Nishenko 1991) seismic-gap hypothesis predicts too many characteristic earthquakes for three reasons. First, forecasts were made for some zones specifically because they had two or more large earthquakes in the previous centuries, biasing the estimated earthquake rate. Second, open intervals before the first event and after the last event are excluded in calculating the recurrence rate. Third, the forecast assumes that all slip in each zone is released in characteristic earthquakes of the same size, while in fact considerable slip is released by both smaller and larger earthquakes (Kagan 1996a).

The observed size distribution of earthquakes is inconsistent with the characteristic earthquake hypothesis: instead of a deficit of earthquakes above and below the characteristic limit, earthquake numbers are distributed according to the standard G–R relation. Fig. 1 shows the relative magnitude distribution, measured with respect to the estimated characteristic magnitude, summed (stacked) over all forecast zones during two different time periods. The first time period (1968 July 1–1989 January 1) overlaps with the learning period during which the forecast zones were defined, and appears to show a slight 'knee' at the characteristic magnitude. During a later time period (1989 January 1–1996 July 1) this knee disappears. The contrast between the two curves shows that the apparent preference for characteristic earthquakes in the earlier period, slight as it is, comes more from the pre-selection of data rather than a true preference for characteristic-magnitude earthquakes. The dash-dotted and dashed curves in Fig. 1 show the idealized G–R and the characteristic distributions, respectively. Even in the earlier data with the selection bias, the enhancement of characteristic earthquakes relative to the G–R model is only a factor of 2, and this enhancement disappears in the later data. The stronger enhancement predicted by the idealized characteristic model does not find support in either data subset.

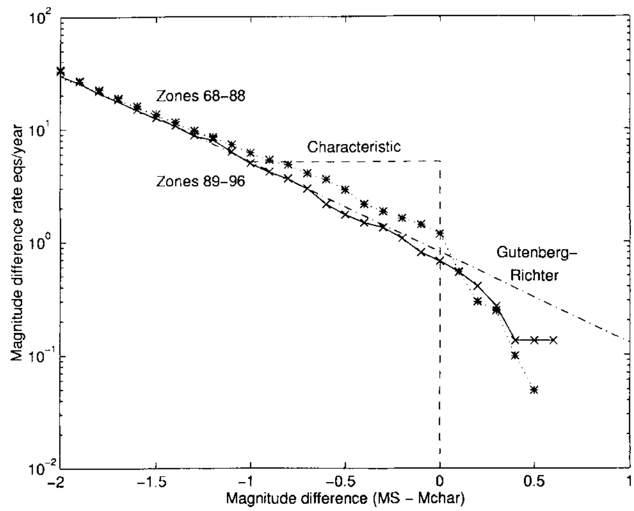

Figure 1. Distribution of PDE magnitude differences for two time periods: 1968 July 1–1989 January 1 and 1989 January 1–1996 July 1, calculated relative to the characteristic magnitude (Nishenko 1991). Asterisks: time interval 1968 July 1–1989 January 1; crosses: time interval 1989 January 1–1996 July 1; dash-dotted line: the G–R relation; dashed line: the predicted distribution of characteristic magnitudes.

By lowering the magnitude threshold for qualifying earthquakes, it is possible to reduce the discrepancy between the observed and predicted number of earthquakes (Nishenko 1991) to an acceptable level. However, for every magnitude threshold Kagan & Jackson (1995) tried, the new seismic-gap model failed the test on the number of filled zones, or the likelihood ratio test, or both, at the 95 per cent confidence level at least. Therefore, no version of the seismic-gap hypothesis has yet shown a significant statistical advantage over a reasonable null hypothesis.

3.3.4 *Forward predictions*

Possibly there are only two kinds of experiments that fully satisfy the criteria proposed in Section 3.2.3: the *M*8 algorithm (Keilis-Borok *et al.* 1988; Keilis-Borok & Kossobokov 1990) and the precursory swarm model of Rhoades & Evison (1979) and Evison & Rhoades (1993).

The *M*8 scheme for predicting large earthquakes is currently being tested. The tests started 1991 July 1 in the circum-Pacific seismic belt, where 147 overlapping circles are specified (Kossobokov *et al.* 1997). The test is assumed to run until 1997 December 31. In the time period ending 1995 July 1, the dangerous intervals (TIPs) occupied 283 half-year intervals out of 1174 possibilities. Five earthquakes $m \geq 7.5$ out of the total of nine events occurred in TIPs. The authors also made a retrospective prediction for the same region in the period 1985 January 1–1991 July 1. The results of the *a posteriori* forecast are significantly better than those of the real-time prediction: eight out of 10 earthquakes have been successfully predicted (Kossobokov *et al.* 1997, p. 228).

The null-hypothesis algorithm, which randomizes earthquake occurrence without taking into account different seismicity levels in seismic zones, '... performed as well as or better than *M*8, 53.30 per cent of 1 000 000 realizations ...' (Kossobokov *et al.* 1997, p. 228). On the other hand, the retrospective prediction performed considerably better than a null hypothesis. As of 1997 May 14, the *M*8 algorithm successfully predicted 10 out of 18 earthquakes with $m \geq 7.5$ (Kossobokov, private communication 1997). These preliminary results suggest that the performance of the *M*8 algorithm is not significantly better than a random guess (the null hypothesis) and again emphasize the important differences between retrospective and forward forecasts. Dieterich (1993) indicated that if one weights the null hypothesis according to '... historical rate of seismicity in each circle', the results of comparison may change significantly in favour of the null hypothesis.

The original precursory swarm model of Rhoades & Evison (1979) performed poorly when tested, and was rejected (Rhoades & Evison 1993). A revised model was then tested, and performed at about the same level as the existing stationary Poisson model (the null hypothesis). Next, a further modification was introduced, and is at present being tested (Rhoades & Evison, private communication, 1997).

3.3.5 *Short-term predictions*

Short-term prediction based on the generalized Omori's law is the only available method for which a predictive power can easily be demonstrated (Kagan & Knopoff 1987; Ogata 1988; Reasenberg & Jones 1989; Kagan 1991; Ogata *et al.* 1996). The earthquake predictive ratio in the wake of even a moderate event rises instantaneously by a factor of many thousand. Most of the following events are weaker than the first one and are thus called aftershocks, but in a small percentage of cases the following earthquake turns out to be larger than the previous one. In such cases the first event is called a foreshock, and the following earthquake a main shock.

In principle, Kagan & Knopoff's (1987) algorithm allows a real-time automatic calculation of earthquake probabilities. The technique used is completely formal and does not require human-operator intervention. Hence, the prediction results can be tested objectively. Proof of the predictive power can be obtained by calculating the likelihood ratio (eq. 3). If the magnitude range of a catalogue is relatively large (more than 1.5–2.0), the information content is of the order of 0.5–2.0 bits per earthquake. The uncertainty in the earthquake rate occurrence can be reduced by a factor of 1.5–4 on average compared to a long-term estimate. Kagan (1991) presented evidence that, for the best available catalogues, the predictability may be close to 10 bits per earthquake (eq. 4), that is the uncertainty can be reduced by a factor of up to 1000 (2^{10}). The reduction of uncertainty represents an average of a very strong predictive ratio in the aftermath of a large earthquake and a practical lack of new information for most earthquakes which are not members of a cluster. Unfortunately, as shown in Table 2 of Kagan (1991), the high likelihood-ratio values are largely due to a near-singularity of the earthquake rate at about the time each event occurred. It means that the 'best' prediction is available in the immediate aftermath of every earthquake. Earthquake catalogues are unreliable at the beginning of large aftershock sequences, when a seismographic network and seismologists are overwhelmed by the number of events. These catalogue deficiencies may also bias our estimate of model parameters significantly. Thus, the model with the maximum likelihood may not be the 'best' one in approximating earthquake occurrence.

The predictability would increase further for real-time prediction (Section 5.3), but the value of this information would decrease since the prediction lead time is very small. However, in this case short-term foreshocks and the initial stages of large-earthquake rupture, which can also be classified as immediate foreshocks, reliably warn of a catastrophic event.

To be effective, short-term alarms should be issued after each earthquake, since any event may be a foreshock. Thus, the number of false alarms is very large for such a method. Since only about 1/4 to 1/3 of earthquakes are preceded by foreshocks sufficiently separated in time from the main shocks (Kagan & Knopoff 1987; Reasenberg & Jones 1989), the reliability of forecasts (see Table 2) is low. Suppose we predict $m \geq 6$ earthquakes using a network with a magnitude cut-off $m = 2$. Then, depending on the prediction threshold, about 10^4 alarms would be generated according to the G–R law for each predicted event. If an alarm is declared after stronger earthquakes, the number of alarms would be smaller, but the reliability of prediction would decrease (Kagan & Knopoff 1987). Thus, only mitigation strategies which have low alarm-initiation costs can benefit from this prediction method (Molchan & Kagan 1992).

3.3.6 *Seismic hazard*

Kagan & Jackson (1994a) estimated long-term worldwide earthquake probabilities by extrapolating the Harvard catalogue (Dziewonski *et al.* 1996) of 1977–1994. The forecast is expressed as a map showing predicted rate densities for earthquake occurrence and for focal-mechanism orientation. In Fig. 2 we display a hazard map for the Northwest Pacific. For temporal prediction we use the Poisson model for the distribution of earthquakes in time. In estimating earthquake probability maps, we use the first half of the catalogue to smooth the seismicity level, and the second half of the catalogue to validate and optimize the prediction. Moreover, the maps can be used as the Poisson null hypothesis for testing by the likelihood method against any other prediction model which shares the same sample space (the same zones, time window and acceptance criteria).

How can we evaluate these predictions? One possibility is to test the internal consistency of the forecast: we simulate earthquake sequences using the forecast maps and compare the likelihood of these synthetic catalogues (Kagan & Jackson 1994a) with those of a control catalogue. In Fig. 3 we display bootstrap distributions for prediction and data, as shown in Fig. 2. We simulate earthquake locations with $R_{\max} = 350$ km (Kagan & Jackson 1994a, p. 13 696), each time calculating the likelihood function and comparing the function value with that obtained for the real catalogue in 1995–1996. Whereas the choice of $R_{\max} = 350$ km was close to optimal for the prediction of the second half of the catalogue (1986–1994), it is clear that for the 1995–1996 seismicity $R_{\max} = 175$ km is a more appropriate smoothing function. Thus, a forecast of all 1995–1996 earthquakes, using the 1994 model with $R_{\max} = 350$ km, would have failed at the 95 per cent confidence level. The reason is apparent from Fig. 2. Earthquakes since 1995 have been much more strongly clustered than before, because of the large number of aftershocks of the strong Kurile Island earthquakes. Even though these aftershocks occurred in an area already red or dangerous in Fig. 2 because of pre-1995 earthquakes, the model with $R_{\max} = 350$ km fails to account for their strong clustering. As Kagan & Jackson (1994a, p. 13 961) explained, the forecast of aftershocks requires the selection of smaller $R_{\max}$. There are two lessons to be learned here. First, like any forecast model, the model of Kagan & Jackson (1994a) needs a quantitative procedure to account for aftershocks. Second, a visual inspection of a colour map like that of Fig. 2 is not adequate: a quantitative hypothesis test is required.

Two destructive earthquakes occurred in areas which are not shown to be especially dangerous on the map (Fig. 2): the Sakhalin event of 1995 May 27 (moment magnitude $m_w = 7.1$) and the Kobe earthquake of 1995 January 16 ($m_w = 6.9$). The lack or paucity of seismicity in the Harvard catalogue in the neighbourhood of both shocks contributed to the low value of the hazard in Fig. 2. These two earthquakes demonstrate two issues related to evaluating earthquake probabilities in regions of low seismicity: (1) a need for new smoothing techniques which would allow the extrapolation of seismicity values from active regions over large distances to areas of low activity; (2) the evaluation of the maximum magnitude.

Kagan and Jackson (1994a) assumed that earthquake size distribution is the same in all regions. To test this hypothesis, Kagan (1997a) investigated the seismic-moment distribution for the Flinn–Engdahl regionalization of the global seismicity using the Harvard data. The maximum moment M_{xg} (see eq. 9 below) can be statistically evaluated only for subduction zones treated as a whole, $M_{xg} = 10^{21}$ to 2×10^{22} N m, which corresponds to a worldwide M_{xg}-value (Kagan 1994; Kagan 1997a). The maximum moment magnitude is 8.0–9.0. For other regions, as well as for single subduction zones, M_{xg} is determined by comparing the number of events in each zone with the seismic moment rate calculated on the basis of the NUVEL-1 model of plate motion (DeMets *et al.* 1990). For subduction zones Kagan (1997a) obtained an estimate of M_{xg} which agrees with the statistical value, providing evidence that most tectonic deformation is released by earthquakes. Kagan (1997a) tested the hypothesis that no statistically significant variations in the b-value and M_{xg} occur in subduction and continental collision zones; the hypothesis cannot be rejected with the available data. These results signify that if we know the deformation rate in continental areas, we can predict the seismic hazard by calculating the seismic-activity rate, since the moment–frequency relation is universal.

4 EARTHQUAKE SCALE INVARIANCE

4.1 Earthquakes as a non-linear dynamic process

Earthquake occurrence distributions exhibit scale-invariant properties: the frequency of an aftershock occurrence decays in time as a power law, and earthquake size distribution is also a power law (Kagan 1994; Vere-Jones 1995; Kagan & Vere-Jones 1996; Main 1996). It has recently been determined that other statistical features of earthquakes, such as the spatial distribution and the rotation of focal mechanisms, are also self-similar (Kagan 1994).

4.1.1 *Earthquake size distribution*

According to the G–R law (eq. 1) the number of earthquakes increases as their sizes decrease. The G–R relation can be transformed into a power-law (Pareto) distribution for the

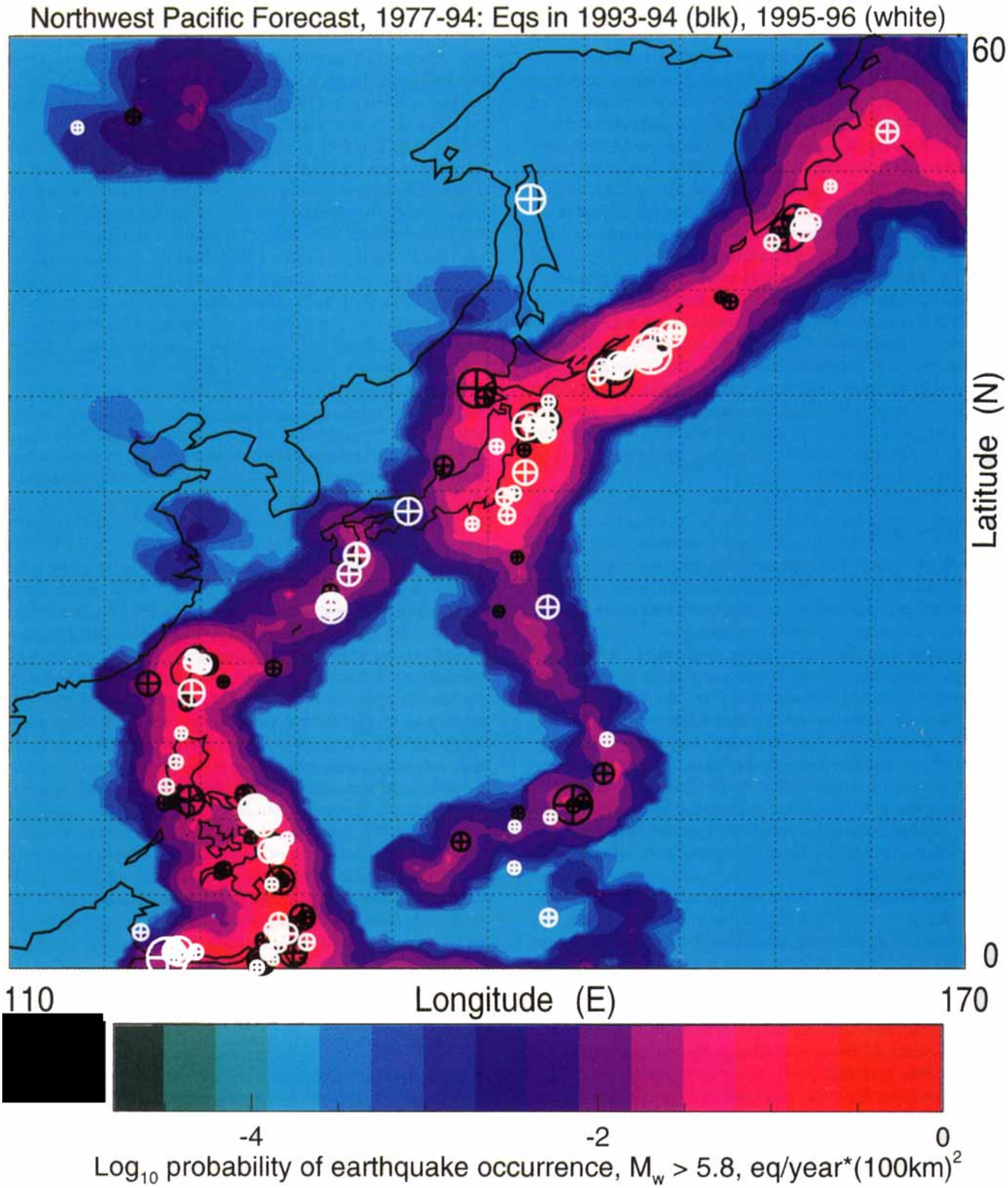

Figure 2. Northwest Pacific seismicity forecast: colour tones show the probability of earthquake occurrence calculated using the Harvard 1977–1994 catalogue: latitude limits 0–60.0°N, longitude limits 110.0–170.0°E. Earthquakes in 1993–1994 are shown in black; earthquakes in 1995–1996 (119 events up to 1996 December 31) are shown in white.

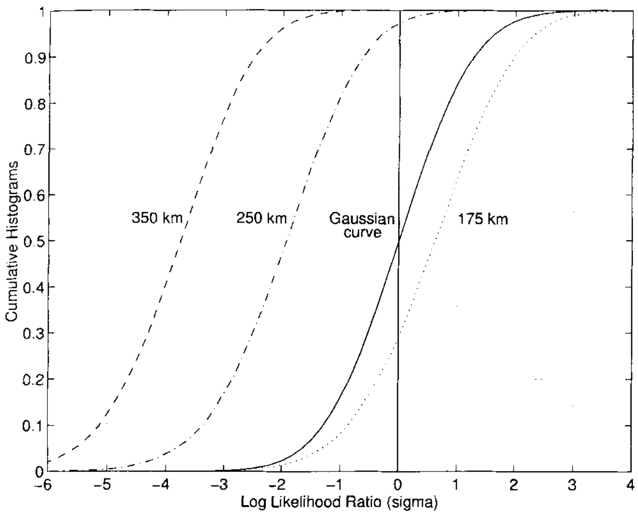

Figure 3. Distribution of scores. Dashed line: simulations with $R_{max} = 350$ km; dash-dotted line: simulations with $R_{max} = 250$ km; dotted line: simulations with $R_{max} = 175$ km; solid line: best Gaussian curve with the same standard deviation as the simulations. Prediction based on seismicity from 1977 to 1994.

scalar seismic moment with the probability density

$$\phi(M) \propto M^{-1-\beta}, \qquad \beta \approx b/1.5. \tag{8}$$

Simple considerations of the finiteness of the seismic moment or deformational energy available to generate earthquakes require that the power-law relation be modified at the maximum size end of the moment scale (Kagan 1994). At a maximum, the probability density tail must have a decay stronger than $M^{-1-\beta}$, with $\beta > 1$. Usually this problem is solved by introducing an additional parameter, a 'maximum moment' (M_{xg}), to the distribution:

$$\phi(M) \propto M^{-1-\beta} \exp(-M/M_{xg}), \tag{9}$$

which is called the modified gamma distribution (Kagan 1994; Main 1996).

To illustrate the gamma-distribution fit to the actual data, Fig. 4 displays a cumulative histogram for the scalar seismic moment of the events in the Harvard catalogue (Dziewonski *et al.* 1996) for shallow, intermediate and deep earthquakes. To ensure data uniformity in time and space, we use events with $M \geq 10^{17.7}$ N m, which correspond to $m_w \geq 5.8$. Investigations of catalogues of smaller earthquakes show that earthquake self-similarity extends to magnitudes as small as zero (that is to a seismic moment of 10^9 N m), and maybe smaller (Abercrombie & Brune 1994). It is likely that, if we disregard source extents of less than a few millimetres (the dimension of microcrystals), there should be practically no lower limit to earthquake size. All the curves in Fig. 4 display a scale-invariant segment (linear in the log–log plot) for small and intermediate values of the seismic moment. At large M, the curves are bent downwards: the lack of very strong earthquakes is the result of the above-mentioned finiteness of the seismic-moment flux. Earthquake-size self-similarity breaks down at a source-size scale of about 500–700 km (Kagan 1997a). Scale invariance of earthquake size for all but the largest earthquakes suggests that the Earth is in a state of self-organized criticality (Bak 1996), where any earthquake has some probability of cascading into a larger one.

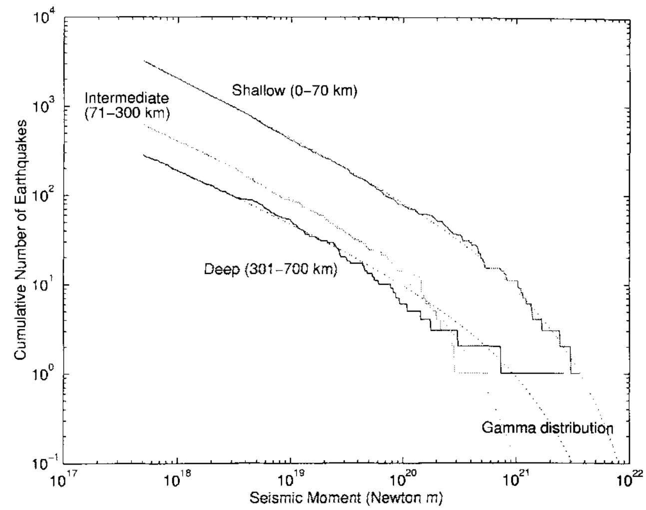

Figure 4. Seismic moment versus cumulative frequency for the 1977 January 1–1996 June 30 Harvard catalogue. The curves show the numbers of events with moment larger than or equal to M. We also show the approximation of curves by the modified gamma distribution, which is the G–R law restricted at large magnitudes by an exponential taper. The slopes of the linear parts of the curves correspond to the β-values (eq. 9) 0.657 ± 0.017, 0.573 ± 0.046, 0.580 ± 0.035, and the maximum moment $M_{xg} = 3.5 \times 10^{21}$, 4.0×10^{20}, 2.2×10^{21} N m for shallow, intermediate and deep earthquakes, respectively. The 95 per cent confidence limits for the maximum magnitude are similarly 8.1–8.7 (8.37), 7.5–8.3 (7.73) and 7.7–∞ (8.23), where the values in parentheses are used in the graph. An extrapolation of the curves to small values of seismic moment shows that the total number of small earthquakes is extremely large.

4.1.2 Earthquake-process self-similarity in different magnitude ranges

Several classes of stochastic multidimensional models have been applied to the statistical analysis of earthquake catalogues using likelihood methods (Kagan 1991; Ogata 1998). The results of these studies suggest that most distributions controlling earthquake interaction have a fractal or scale-invariant form (Mandelbrot 1983). The parameters of earthquake occurrence are shown to be similar for shallow earthquakes of different magnitude ranges and seismogenic regions, confirming self-similarity for the earthquake process. Since micro-earthquakes in rock specimens and in mines seem to follow similar distributions, the self-similarity extends over many orders of magnitude (Kagan 1994).

4.1.3 Temporal fractal pattern

Statistical studies of earthquake occurrence (Kagan 1991; Ogata 1998) and the results of computer simulations (Kagan & Vere-Jones 1996) suggest that the temporal behaviour of earthquake sequences is governed by power laws. For more than 100 years (Utsu, Ogata & Matsu'ura 1995), it has been known that for shallow earthquakes the rate of aftershock occurrence (Omori's law) has a power-law decay:

$$\phi(t) \propto t^{-1-\theta}, \tag{10}$$

where $\theta = 0$ to 0.5. Foreshocks are also shown to obey a similar power-law increase before a main shock (Kagan 1991). Kagan (1982, 1994) suggested that this dependence is due to stress diffusion: if we assume that stresses at the end of an earthquake rupture are below the critical value and thereafter change randomly according to 1-D Brownian motion, then the level-set of this motion is a fractal set with a dimension 0.5 (Mandelbrot 1983). When the random stress reaches the critical level, a new rupture starts. Therefore, the time intervals between the end of fracture and the beginning of a new earthquake are distributed according to a power law. Thus, the short-term clustering of earthquakes is due to stress diffusion.

4.1.4 *Fractal pattern of earthquake hypocentres*

Using several earthquake catalogues, we analyse the distribution of distances between pairs of earthquake hypocentres to determine the spatial fractal correlation dimension (δ) of an earthquake fracture (Kagan 1994 and references therein). As the time span of the catalogue increases, the correlation dimension δ asymptotically reaches a value 2.1–2.2 for shallow earthquakes. The spatial scale invariance breaks down for distances between hypocentres of 2000–3000 km; this distance corresponds to the average size of continents or the total thickness of the mantle.

4.1.5 *Cauchy distribution of focal-mechanism rotations*

Kagan (1982) introduced the rotational Cauchy distribution to represent the rotations of the focal mechanisms of microdislocations which comprise the earthquake focal zone. The Cauchy distribution is especially important for earthquake geometry, since theoretical arguments and simulations show that a stress tensor in a medium with defects follows this distribution (Kagan 1994). The 3-D rotation of earthquake focal mechanisms can also be approximated by the Cauchy distribution (Kagan 1994). The Cauchy law is a stable distribution with a power-law tail (it is fractal and should yield fractal fault geometrical patterns).

4.1.6 *Random stress model*

On the basis of the results presented above, we offer a model of random defect interaction in a critical stress environment, which, without additional assumptions, seems to explain most of the available empirical results (Kagan 1994; Kagan & Vere-Jones 1996). In the time domain, Omori's law of foreshock/aftershock occurrence and, in general, the time clustering of earthquake events, are a consequence of the Brownian-motion-like behaviour of random stresses due to defect dynamics. These results justify the short-term prediction proposed by Kagan & Knopoff (1987).

Similarly, the presence, evolution and self-organized aggregation of defects in the rock medium are responsible for fractal spatial patterns of earthquake faults and rotation of earthquake focal mechanisms. The Cauchy distribution governs the stresses caused by these defects, as well as the rotation of focal mechanisms (Kagan 1994). These considerations indicate that, if we know the geometry of the defects in a medium, future deformation patterns can be predicted.

The results obtained force one to question the suitability of some concepts and models commonly used in the theory of earthquake sources. In particular, standard models of the source are based on the mechanics of man-made objects, thus they introduce such concepts as 'an individual earthquake' (see Section 5.1), 'a fault-plane' or 'a fault-surface', 'a crack-tip' or 'a fault-tip' and 'friction', which in a scale-invariant model lack an unambiguous definition (Kagan 1994).

4.2 Earthquake modelling

Efforts to develop mechanical or computer models of the earthquake process (Ben-Menahem 1995; Knopoff 1996; Knopoff *et al.* 1996) have not yet achieved real predictive power. The earthquake models used in simulations are usually autonomous, isolated, closed systems, but in nature tectonic earthquakes result from global mantle convection. Thus, no region can be considered isolated.

It is difficult to judge computer simulations of earthquake occurrence: similar calculations of fluid dynamics can be directly compared to the actual velocity field (Ruelle 1991; Frisch 1995). Synthetic earthquake catalogues, however, need to be matched with real catalogues. Due to the highly random nature of seismicity, earthquake occurrence can be discussed only in statistical terms. If synthetic sequences are to be exploited in modelling seismicity, we must show that they have the same statistical characteristics as real earthquake sequences with respect to their distribution in size, time and space; it is here that problems develop.

(1) Earthquake size distribution is not sufficiently specific to distinguish between competing models of seismogenesis. The power-law seismic-moment distribution (which is equivalent to the G–R relation, see Section 4.1.1) seems to be an outcome of many models with a hierarchical structure. The power laws may also be the result of critical self-organization (Bak 1996).

(2) To a large degree, simulated spatial earthquake distribution is controlled by the fixed geometry of faults designated at the start of the computations. An actual geometrical fault pattern may be the result of self-organization in which stress redistribution plays a major role. Only a few mechanical models, such as that of Cowie, Sornette & Vanneste (1995), simulate the evolving geometry and mechanical properties of a fault system, but this is done for a very simple mechanical model.

(3) Such major features of earthquake occurrence as aftershock sequences are not reproduced by most models, casting doubt on the applicability of their results to real earthquakes (Kagan 1994). Since the simulations do not generate foreshocks and aftershocks, only the temporal interaction of main events can be investigated. Main-shock sequences can be fairly well approximated by a Poisson process, which shows that these earthquakes are statistically independent. There is little information in sequences of independent events; such time-series can be produced by a great variety of mechanisms. One should expect that, on closer inspection, main shock occurrence will turn out to be non-Poissonian. However, it is debatable whether large earthquakes in nature are clustered or quasi-periodic (Kagan 1994). On the other hand, the temporal pattern of foreshock/main-shock/aftershock sequences is reasonably well known. Thus, if a computer model could generate such sequences, one could compare them with observational results.

4.3 Recurrence models and long-term earthquake prediction

4.3.1 *Scalar models*

The recurrence hypothesis implies that earthquake hazard is small immediately following a previous large earthquake and increases with time. The usual explanation is that a large earthquake releases most of the stress in a given fault segment and that further earthquakes would be unlikely until the stress is restored by plate motion. These simple ideas form a basis for most long-term earthquake prediction efforts (see Section 3.3.3). On the other hand, earthquake catalogues testify to the fact there there is no quiescence after a strong earthquake—on the contrary, the earthquake rate decays according to Omori's law. The aftershocks of the Nobi 1891 earthquake, described by Omori (1894), still obey the eponymous law more than 100 years after the beginning of the sequence (Utsu *et al.* 1995).

The earthquake-clustering hypothesis has been offered to explain the spatial and temporal clustering of events (Kagan 1994). This hypothesis, which is the opposite of the recurrence model, seems to contradict the steady accumulation of strain due to plate motion, and is thus considered counter-intuitive and even paradoxical. However, if the earthquake size–frequency relation is described by the G–R or by the modified gamma distribution, the strongest earthquakes release the major part of the tectonic deformation (Kagan 1996a). The arguments on strain accumulation and release are thus applicable only to these large earthquakes.

To illustrate this, we display the probability density for the total moment rate released by earthquakes in Fig. 5(a). The density is displayed for three statistical distributions commonly used to approximate earthquake size (Kagan 1996a). All these distributions reproduce the G–R law for small and intermediate events. For large earthquakes, the distribution densities have either a delta function at the maximum moment, a density truncation or exponentially smooth decay—the gamma distribution. In the figure we have adjusted the maximum moment, $M_{\max}$, so that the moment rate is identical for all three distributions: $M_{\max} = 10^{21}$, 3.38×10^{21} and 4.74×10^{21} N m, respectively (see Fig. 4). These values of the moment correspond approximately to magnitude values $M_{\max} = 8.0$, 8.35 and 8.45. For the truncated distributions, the maximum moment rate release is at the moment of the maximum earthquake. However, for the gamma distribution, the maximum release is at $M_{\max}/3$ (corresponding to magnitude 8.13 in the graph). About 65 per cent of the total moment is due to earthquakes smaller than $M_{\max}/3$; the rest (35 per cent) is released by larger events. The cumulative distributions shown in Fig. 5(b) demonstrate that earthquakes with $m > 7.3$ (10^{20} N m) account for about 70 per cent of the total seismic moment.

Although tectonic stress accumulates steadily, it may not be released on the same fault in a quasi-periodic fashion: very large displacements can be stored and then released in a few giant earthquakes. These giant earthquakes are so rare (Kagan 1997a) that the effect of significant stress release by these events may have very little predictive value. Moreover, the stress in the seismogenic zones may always be close to a critical state, even after very strong earthquakes, allowing new large events to occur very soon thereafter. Earthquakes that occur near newly filled dams testify to the near-criticality of stress even in stable continental areas (Scholz 1991). Additionally, a system of faults can accommodate large-scale deformation by releasing it in clusters of large earthquakes occurring on different, possibly closely related faults, by rotation of microblocks, and by a host of other mechanisms.

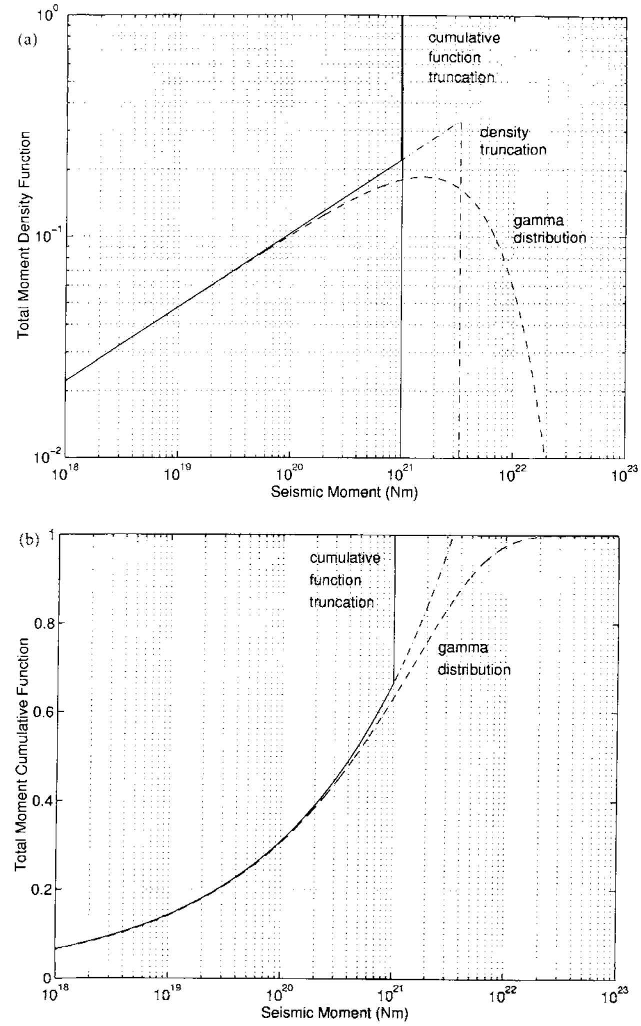

Figure 5. Probability function for distribution of seismic-moment rate. Solid line: truncated G–R cumulative distribution; dash-dotted line: truncated G–R distribution density; dashed line: modified gamma distribution. All distributions have a *b*-value for small earthquakes equal to 1. The maximum moment parameter is adjusted so that the total moment is the same for all three distributions. (a) Distribution density; (b) cumulative distribution.

The discussion above assumes that stress and strain can be approximated as scalar quantities. Tectonic deformation is effected through a fault system which forms very complicated 3-D geometrical patterns. Thus, scalar representation is clearly insufficient.

4.3.2 Stress tensor and earthquakes: fundamental assumptions and problems

The advent of high-speed computers and mass determination of seismic moment tensor parameters allows us to calculate the stress tensor and its relation to earthquake triggering. Several recent publications (King, Stein & Lin 1994; Harris & Simpson 1996; Jaumé & Sykes 1996, and references therein) explore the inter-relation of stress and earthquakes. Many interesting results have been obtained thus far, but their interpretation in the framework of stress accumulation and release using the recurrence models encounters serious difficulties.

If a large earthquake occurs when the stress exceeds the strength of rocks, why do small earthquakes occur over the seismogenic zone all the time? If the stress value is close to the critical level over a large area, should strong earthquakes occur more frequently than usual, leading to a smaller b-value in the region—this feature has not been observed unambiguously—or does the increased stress level simply trigger more earthquakes without regard to size? The magnitude–frequency relation for aftershocks does not seem to vary from that for all earthquakes. If there is a stress shadow after a large earthquake in the focal zone and nearby, how can one explain the occurence of aftershocks? There is no clear spatial, temporal or magnitude boundary between the aftershocks and other earthquakes. If the increased stress triggers earthquakes and a large earthquake releases stress, why are there more aftershocks than foreshocks? The models of stress triggering earthquakes assume that the Coulomb fracture criterion is valid for the Earth's interior, as has been established in laboratory testing of rock specimens. However, numerous attempts to evaluate the friction coefficient *in situ* have been inconclusive, often suggesting that the coefficient is close to zero (Kagan 1994). The tensor stress field in a fault zone is a complex, self-similar, 3-D mosaic, thus determining where an earthquake starts and stops becomes a formidable problem (Kagan 1994). These facts challenge the conventional paradigm for earthquake generation.

It is possible that ideas about stress accumulation and its release in earthquakes will be useful for long-term forecasts of earthquake occurrence. As we have seen above, previous attempts to apply these models have not been successful. The fundamental problem in their implementation has been that basic and essential procedures of scientific research were neglected: in almost all cases the prediction scenarios were not stated as testable, falsifiable hypotheses.

5 DISCUSSION: EARTHQUAKE PREDICTABILITY

5.1 Specific-earthquake prediction

Is it possible to predict a specific earthquake, even in principle? The concept of precursors for an isolated specific earthquake seems to be based on an intuition that preparing for a large earthquake is similar to the life of a biological organism: its stages of life can be observed by appropriate techniques and approprite precursors can be identified. The 'characteristic earthquake' model demonstrates these ideas (for instance, one discusses *the next* Parkfield earthquake). Aki (1995, p. 244) said that 'Once a characteristic earthquake is identified for a given fault segment, it becomes an individual, like a human being, for which life expectancy at a certain age can be evaluated and used for determining the premium for life insurance.' Contrary to that, in continuum physical systems there are no individuals. We can, for example, subdivide a mountain range into separate mountains, but this is a purely human way of naming things. There is no physical meaning behind it, and no computer algorithm can be created to make such a subdivision unambiguously. As we mentioned in Section 4.1.6, the definition of an individual earthquake or an individual fault is questionable in view of the scale-invariant nature of the earthquake process. It is the process of measurement and its interpretation which isolates the clusters of elementary rupture events and gives them a separate identity as individual earthquakes (Kagan & Vere-Jones 1996).

Most geophysicists would agree that our present knowledge of the geology and tectonics of seismogenic regions is clearly insufficient to predict a specific individual earthquake, since the possibility of an earthquake similar to the predicted one must be ruled out. If earthquake size, space and magnitude form a continuous distribution, a specific event cannot reliably be identified. Moreover, the scale invariance of an earthquake process signifies a highly sensitive non-linear dependence of its evolution on the initial conditions. Exactly when and where earthquakes occur and how large they grow after they start depend on a myriad of fine and unmeasureable details of the physical state of the Earth throughout a large volume, not just near the fault.

The new developments in the non-linear mechanics of earthquake generation, scale invariance and self-organization of the earthquake process, and statistical analysis of earthquake sequences (Kagan 1994) strongly suggest that the earthquake process is unstable: the size of each earthquake is determined only during the process of rupture (Brune 1979). Many small earthquakes occur throughout any seismic zone, demonstrating that the critical conditions for earthquake nucleation are satisfied almost everywhere. Apparently, any small shock could grow into a large event. This would mean that no long-term earthquake precursors are possible that can yield precise information on the size of a future event (that is classes 1–3 in Table 1). The reliable prediction of a specific strong earthquake can be reached only at the timescale when inertial effects are strong, that is at the timescales of seconds and at most tens of seconds for great earthquakes (see Section 5.3 and 5.4).

Sornette & Sammis (1995), Varnes & Bufe (1996) and Johansen *et al.* (1996) found log-periodic fluctuations of seismicity and ion concentrations in groundwater before a few earthquakes. According to these authors, these log-periodic modulations are second-order effects superimposed over a hyperbolic long-term increase of seismicity before a main shock. They propose that both of these effects result from the fact that a large earthquake is a critical phenomenon, in the sense of statistical physics. The long-term earthquake clustering with a power-law temporal dependence is a characteristic feature of seismicity (Kagan 1994 and references therein). The log-periodicity would allow a more exact prediction of time for a main shock occurrence and would forecast its magnitude approximately. However, the existence of the log-periodic scale needs to be confirmed by rigorous testing. These log-periodic patterns have been observed only for a few earthquake sequences. Johansen *et al.* (1996, p. 1401) remarked that '... there are presumably other sorts of fits with as many parameters that would work as well'. As we discussed above, the data available for examination comprise many hundreds and thousands of sequences. Thus, one needs to show that the properties of the sequences that have been analysed are representative of earthquake occurrence as a whole, in other words the selection of the sequences has been unbiased. Gross & Rundle (1995) investigated the log-periodic patterns of seismicity in a more systematic manner, and their preliminary results are negative. We note that some precursory patterns that showed interesting and promising results in preliminary *ex post facto* tests were disappointing when tested in the forward mode (Section 3.3.4).

However, is it possible to predict a specific earthquake in a laboratory or in computer simulations? In a laboratory and simulations, a crack develops instabilities which make its propagation highly chaotic and unpredictable (Abraham 1996; Marder 1996; Marder & Fineberg 1996; Abraham *et al.* 1997). These instabilities are due to crack propagation, especially at a speed close to the elastic-wave velocity. Stress and fracture conditions in laboratory specimens differ significantly from those in earthquake fault zones: the boundary effects are controlled by an experimenter in the lab. Therefore, fracture can self-organize only at spatial scales much smaller than that of the specimen. In fault zones, stress, rock mechanical properties and fault geometry are self-organized with the development of large-scale self-similar patterns.

Researchers working in the computer modelling of earthquake occurrence are aware of the chaotic development of synthetic seismicity (Knopoff *et al.* 1996). A more important point here is whether one can predict—in theory or by computer models—when an earthquake rupture propagation will stop and whether it is possible to predict possible branching earthquake faults and 3-D rotations of earthquake focal mechanisms. Thus, earthquake size depends on the conditions in widely separated regions of the seismogenic zone, and on our definition of the end of faulting. If we are not able to predict this deterministically, then predictions 1 and 2 (Table 1) cannot be achieved, even in ideal circumstances.

Currently, the exact conditions and constitutive equations for earthquake rupture are unknown. Because of this and significant computational difficulties, most simulations are carried out in a quasi-static regime. However, even if these equations were known, there are still fundamental difficulties in predicting when rupture stops. Even if we knew the position of all of the atoms in our model, quantum-mechanical effects would prevent full knowledge of fault-rupture propagation. Kagan (1982) proposed modelling fault propagation during an earthquake as a continuous-state, critical, branching random process. A new fault system can branch starting from any continuum point and develop into a large earthquake. Classical continuum mechanics is the foundation of earthquake models, but the continuum real numbers are not computable (see e.g. Svozil 1995; Casti & Karlqvist 1996), making deterministic computer modelling impossible.

Due to defects, the theoretical strength of materials is two to three orders of magnitude higher than the real strength (Marder & Fineberg 1996). In natural systems, the largest defects (earthquake faults, for example) are comparable in size with the size of the system. The geometry of defects is scale invariant and fractal, and the defects form Cantor-set-type temporal and spatial patterns (Kagan 1994; Kagan & Vere-Jones 1996). Geometry and other branches of mathematics have undergone a fundamental change since Cantor's discoveries at the end of the 19th century (Mandelbrot 1983; Vilenkin 1995). The present continuum mechanics is still pre-Cantorian in outlook.

5.2 Precursors

To date, the empirical search for earthquake precursors that yield information on the size of an impending earthquake (that is classes 1–3 in Table 1) has been futile. Precursors can be subdivided into two categories: strong and weak. Strong precursors are phenomena which yield a predictive ratio either $\eta \geq 10$ or $\eta \leq 0.1$, thus they predict either very high seismic activity, or low activity (quiescence). Weak precursors, in our definition, have $\eta \neq 1$, but $0.1 \leq \eta \leq 10$. The strong precursors would undoubtedly be seen even without a sophisticated statistical analysis. The efforts of the last 100 years to find these strong precursors have failed. In all probability, they do not exist (*cf.* Turcotte 1991).

An opinion often expressed (see e.g. Lomnitz 1994; Ben-Menahem 1995) is that by combining information from several unspecified, weak precursors, one can reliably predict earthquakes. Many publications claim to see precursors to earthquakes (Mogi 1995; Lomnitz 1994 and references therein). Space considerations do not allow us to discuss most of these in detail. Almost all the precursors have been found after an earthquake, i.e. retrospectively (Geller 1997; Mulargia 1997; Geller *et al.* 1997), and almost all the reports are case histories, without any attempt to confirm the precursor's validity statistically. There are usually no objective definitions of 'precursory anomalies', no consistency in appearance of the reported precursors, no quantitative physical mechanism links the alleged precursors to earthquakes, and natural or artificial causes unrelated to the earthquakes have not been compellingly excluded. One often observes an inverse correlation between the quality of the data and the number of 'surprising' precursor results based on these data.

Several statistical tests have been carried out in the forward mode (Sections 3.3.3 and 3.3.4), the most rigorous examination of a prediction method's validity (Engelhardt & Zimmermann 1988, p. 230). The techniques that have been tested are the results of long-term prediction efforts by highly qualified scientists—they perhaps represent the 'cream of the crop' in precursor analysis. The results of the tests have shown that even the best prediction methods have possibly no predictive skill. The null hypothesis could not be rejected as an explanation of the test results.

Is it possible that some statistical indications of future earthquake size exist, that is can we predict that the distribution of future earthquake size would differ from the G–R relation for a certain time–space interval? In other words, is the 'magnitude-specific' prediction (Item 3 in Table 1) possible? It is difficult to answer this question conclusively. Many schemes claim to predict the size of future earthquakes, and some statistical tests seem to indicate that a proposed method indeed has a predictive power. For example, Molchan & Rotwain (1985) and Molchan *et al.* (1990) tested 'bursts of aftershocks' as a predictor of strong earthquakes. The 'bursts of aftershocks' are used as one of the patterns in the *M*8 algorithm (Section 3.3.4). Whereas the first attempt was unsuccessful, the accumulation of earthquake data later allowed Molchan *et al.* (1990) to confirm, albeit marginally, the statistical significance of this premonitory pattern. The question arises whether the predictive power of this method is due to long-term earthquake clustering or to some other features of seismicity. Similar problems arise in analysing other prediction claims. As discussed earlier (Sections 3.2.1, 3.3.5 and 4.1), clustering has a very strong predictive power: the future occurrence rate may increase by several orders of magnitude after a strong earthquake. Unfortunately, all attempts to derive additional information on the size distribution of future events have been fruitless; the distribution cannot be shown to differ from the standard G–R law in a statistically significant manner.

If the size distribution depends on details of stress patterns, then one would expect that the b-value in eq. (1) would change significantly in the wake of a very strong earthquake which releases and redistributes stress over large regions. However, despite many attempts to find the b-value variation before and after a large event, no such pattern has been unambiguously found.

The general scientific methodology is to assume that the null hypothesis (no precursors) is correct, until convincing evidence to the contrary is provided. In any case, the burden of proof is on the proponents of precursors. We thus conjecture that there are no precursors which forecast the size of a future earthquake (class 3 in Table 1) and there is no preparatory stage for an earthquake.

5.3 Real-time warning systems

Several near-real-time earthquake warning systems are now in operation in a few countries (Lomnitz 1994; Malone 1996). By providing timely information on source parameters (location, depth, origin time, magnitude, focal mechanism, etc.) to government officials and the public, these systems can facilitate relief efforts after large earthquakes. The NRC Panel (1991) discussed a system of almost-real-time warning and automatic response which is in effect a real-time prediction.

Modern technology makes it possible to obtain earthquake parameters during the actual faulting process (see e.g. Scrivner & Helmberger 1995). The results of such real-time inversion could be used to calculate the strong motion (Olsen, Archuleta & Matarese 1995) of an earthquake during the process of rupture and provide this information to potential users in real-time. If seismometers are situated close to a rupturing fault and the computation is almost simultaneous with the arrival of seismic waves, a warning can be issued for facilities far enough from the fault, with a lead time from a few seconds up to 1–2 minutes (NRC Panel 1991, p. 27). During this time (the propagation of destructive shear and surface waves), the extent of the faulting which has occurred can, in principle, be determined. Thus, the forecast has the properties of a deterministic prediction (Item 1 in Table 1) or of the 'ideal prediction' in Table 3. However, we cannot predict whether an earthquake rupture would stop propagating at the time of its registration, thus the prediction will be less accurate after the interval corresponding to the wave-propagation time. The results of a real-time or near-real-time seismic system can be used by short-term algorithms (Section 3.3.5) to evaluate earthquake probabilities (Kagan & Knopoff 1987; Molchan & Kagan 1992) at longer time intervals (minutes to weeks).

5.4 Earthquake and weather prediction

The highly heterogeneous state of the Earth, the absence of a quantitative theory for earthquake generation and the inaccessibility of fault zones to direct measurements impose serious difficulties on earthquake predictability. However, we argue that the fundamental reason for the difficulties lies elsewhere. In the turbulent motion of fluids, the fluid properties and basic laws governing displacement (the Navier–Stokes equations) are known (Ruelle 1991; Frisch 1995). However, the reliable prediction of air motion in a room can be calculated only for a few minutes and Earth atmosphere circulation can be predicted for a few days, when inertial effects are strong. The sensitive dependence on initial conditions signifies that longer-term weather prediction is impossible (Ruelle 1991; Frisch 1995). For earthquakes, the inertial effects are strong only during earthquake rupture and the subsequent seismic-wave propagation, that is on the timescale of seconds.

There is another difference between weather and earthquake prediction. The earthquake process is strongly asymmetric in time: there are fewer foreshocks, if any, than aftershocks in an earthquake sequence. The time asymmetry of seismicity, which is so different from the turbulent flow of fluids, explains why earthquake prediction, unlike weather forecasting, is unreliable even at very small lead times. In more than 50 per cent of earthquake sequences (Section 3.3.5), the first event is the largest (a main shock). Whereas earthquake initiation is usually sharp and abrupt, its stopping phase consists of many rupture events, which are either classified as a late phase of faulting or as immediate aftershocks. While the most violent manifestation of atmosphere turbulence—a tornado—can be predicted with a lead time on the order of half an hour with few errors (Desrochers & Donaldson 1992, p. 382), a catastrophic earthquake may occur practically without warning. However, the prediction of a tornado's path, as well as that of a tropical hurricane, faces serious difficulties, since small variations of initial conditions may drastically change its trajectory.

5.5 Are earthquakes predictable?

It is much more difficult to establish that a solution for a certain scientific problem is impossible than the opposite. Such findings are common in mathematics. In other sciences, long, unsuccessful attempts to resolve problems such as '*perpetuum mobile*' in physics or the transformation of elements in chemistry have led scientists to believe that these problems are insolvable, even before the laws that prove the impossibility of the solution have been formulated. The history of unsuccessful attempts over the last 100 years to find a method for predicting future earthquake size suggests that this problem may also be insolvable. Of course, it cannot be proven at present; therefore, our conclusions can only be provisional.

Is prediction as popularly defined (Items 1–3 in Table 1) inherently impossible or just extremely difficult? Scientifically, the question can be addressed using a Bayesian approach (Anderson 1992; Geller *et al.* 1997). Each failed prediction attempt lowers the *a priori* probability for the next attempt. Based on the record to date, the current probability of successful prediction is extremely low. The obvious ideas have been tried and rejected during the last 100 years of modern earthquake seismology. The task of systematically observing subtle precursory phenomena (if they exist), formulating hypotheses and testing them rigorously against future earthquakes would require an immense effort over several decades. A rigorous proof of reliable forecasting of future earthquake size would qualify as a major breakthrough.

From a theoretical point of view, earthquake prediction and, in general, the understanding of the large-scale deformation of brittle solids, are extremely difficult scientific problems, perhaps even more difficult than the study of turbulence—the major unsolved problem of science. The solution may require the development of completely new mathematical and theoretical tools. We should not expect significant progress in this direction in the near future.

This sceptical view of current earthquake prediction efforts should not be interpreted as a statement that attempts to mitigate the destructive effects of earthquakes are futile. Evaluating seismic hazard as well as time-dependent earthquake rates (Items 4 and 5 in Table 1) should significantly increase our ability to predict future earthquakes and adopt appropriate strategies for earthquake counter-measures. Real-time seismology can facilitate relief efforts after large earthquakes and eventually provide an immediate warning.

6 CONCLUSIONS

(1) An empirical search for earthquake precursors which forecast the size of an impending earthquake has been fruitless. Rigorous statistical attempts to verify proposed precursors have either been negative or inconclusive. The lack of consistency and systematics in the reported cases of precursors suggests that the precursors are caused by random noise or result from chance coincidence.

(2) Recent developments in the non-linear dynamics of earthquake generation make it questionable that any precursors exist. The long history of unsuccessful attempts to find precursors and the lack of rigorous statistical confirmation of their existence suggest that no preparatory stage exists for an earthquake. This would mean that any small earthquake could grow into a large event.

(3) Although stress accumulation and release models may be the best vehicles for understanding earthquake processes, applying simple scalar models and simple geometries of earthquake faults leads to contradictions and paradoxes, indicating that our present understanding is significantly deficient.

(4) The seismic-moment conservation principle and new geodetic deformation data offer a new way to evaluate seismic hazard, not only for tectonic plate boundaries, but also for continental interiors.

(5) Earthquake clustering with power-law temporal decay can be used to estimate the future earthquake occurrence rate. Such schemes exist for short-time prediction and can be developed for longer-time forecasts. Quantitative multidimensional stochastic models of earthquake clustering need to be significantly improved, both for statistical testing of predictions and for short-term earthquake forecasts.

(6) For very short time intervals, real-time seismology can provide relatively reliable information on source parameters during earthquake rupture and on shaking amplitude before destructive seismic waves arrive.

ACKNOWLEDGMENTS

This research was supported in part by the National Science Foundation through Cooperative Agreement EAR-8920136 and USGS Cooperative Agreements 14-08-0001-A0899 and 1434-HQ-97AG01718 to the Southern California Earthquake Center (SCEC). I am grateful for useful discussions held with D. D. Jackson, L. Knopoff and D. Sornette of UCLA, R. Geller of Tokyo University, F. Mulargia of Bologna University, D. Vere-Jones and F. Evison of Victoria University in Wellington, D. Rhoades of New Zealand Institute for Industrial Research and Development, P. B. Stark of UC Berkeley, and V. I. Keilis-Borok, G. M. Molchan and V. G. Kossobokov of the Russian Academy of Science. Publication 367, SCEC. Publication 4886, Institute of Geophysical and Planetary Physics, University of California, Los Angeles.

REFERENCES

Abercrombie, R.E. & Brune, J.N., 1994. Evidence for a constant *b*-value above magnitude 0 in the southern San Andreas, San Jacinto and San Miguel fault zones, and at the Long Valley caldera, California, *Geophys. Res. Lett.*, **21**, 1647–1650.

Abraham, F.F., 1996. Parallel simulations of rapid fracture, in *Fracture—Instability Dynamics, Scaling and Ductile/Brittle Behavior,* pp. 311–320, eds Selinger, B.R., Mecholsky, J.J., Carlsson, A.E. & Fuller, E.R., Jr, Mater. Res. Soc. Pittsburgh, PA.

Abraham, F.F., Marder, M., Griffiths, J.R. & Fineberg, J., 1997. MAAD scientists and others do numerical fracture studies, *Physics Today,* **50** (2), 15, 89–90.

Aceves, R.L., Park, S.K. & Strauss, D.J., 1996. Statistical evaluation of the VAN method using the historic earthquake catalog in Greece, *Geophys. Res. Lett.*, **23**, 1425–1428.

Agnew, D.C. & Ellsworth, W.L., 1991. Earthquake prediction and long-term hazard assessment, *Rev. Geophys. Suppl.*, **29,** 877–889.

Aki, K., 1981. A probabilistic synthesis of precursory phenomena, in *Earthquake Prediction, An International Review,* Maurice Ewing Vol. 4, pp. 566–574, eds Simpson, D.W. & Richards, P.G., AGU, Washington, DC.

Aki, K., 1995. Earthquake prediction, societal implications, *Rev. Geoph., Suppl.*, **33,** 243–247.

Anderson, P.W., 1992. The Reverend Thomas Bayes, needles in haystacks, and the fifth force, *Physics Today,* **45** (1), 9–11.

Bak, P., 1996. *How Nature Works: the Science of Self-Organized Criticality,* Copernicus, New York, NY.

Bakun, W.H. & Lindh, A.G., 1985. The Parkfield, California, earthquake prediction experiment, *Science,* **229,** 619–624.

Ben-Menahem, A., 1995. A concise history of mainstream seismology–origins, legacy, and perspectives, *Bull. seism. Soc. Am.*, **85,** 1202–1225.

Brune, J.N., 1979. Implications of earthquake triggering and rupture propagation for earthquake prediction based on premonitory phenomena, *J. geophys. Res.*, **84,** 2195–2198.

Casti, J.L. & Karlqvist, A., eds., 1996. *Boundaries and Barriers: on the Limits to Scientific Knowledge,* Addison-Wesley, Reading, MA.

Cowie, P.A., Sornette, D. & Vanneste, C., 1995. Multifracture scaling properties of a growing fault population, *Geophys. J. Int.*, **122,** 457–469.

Daley, D.J. & Vere-Jones, D., 1988. *An Introduction to the Theory of Point Processes,* Springer-Verlag, New York. NY.

DeMets, C., Gordon, R.G., Argus, D.F. & Stein, S., 1990. Current plate motions, *Geophys. J. Int.*, **101,** 425–478.

Desrochers, P.R. & Donaldson, R.J., 1992. Automatic tornado prediction with an improved mesocyclone-detection algorithm, *Weather Forecasting,* **7,** 373–388.

Dieterich, J.H., 1993. Comparison of *M*8 test results with 'Poisson' model, *USGS Open-file rept* **93-333,** Appendix I.

Dziewonski, A.M., Ekström, G. & Salganik, M.P., 1996. Centroid-moment tensor solutions for July–September 1995, *Phys. Earth planet. Inter.*, **97,** 3–13.

Engelhardt, W., von & Zimmermann, J., 1988. *Theory of Earth Science,* University Press, Cambridge.

Evison, F.F. & Rhoades, D.A., 1993. The precursory earthquake swarm in New Zealand: hypothesis tests, *New Zealand J. Geol. Geophys.*, **36,** 51–60; Correction, 267.

Evison, F.F. & Rhoades, D.A., 1994. On the testing of earthquake precursors, in *Electromagnetic Phenomena Related to Earthquake Prediction,* pp. 1–11, eds Hayakawa M. & Fujinawa, Y., Terrapub, Tokyo.

Feng, D.Y., Gu, J.P., Lin, M.Z., Xu, S.X. & Yu, X.J., 1984. Assessment of earthquake hazard by simultaneous use of the statistical method and the method of fuzzy mathematics, *Pure appl. Geoph.*, **122,** 982–997.

Frisch, U., 1995. *Turbulence: Legacy of A. N. Komogorov,* Cambridge University Press, Cambridge.

Geller, R.J., 1991. Shake-up for earthquake prediction, *Nature,* **352,** 275–276.

Geller, R.J., 1996a. VAN: A critical evaluation, in *Critical Review of VAN, Earthquake Prediction from Seismic Electrical Signals,* pp. 155–238, ed. Lighthill, J., World Scientific, Singapore.

Geller, R.J., ed., 1996b. Debate on 'VAN', *Geophys. Res. Lett.,* **23,** 1291–1452.

Geller, R.J., 1997. Earthquake prediction: a critical review, *Geophys. J. Int.,* **131,** 425–450 (this issue).

Geller, R.J., Jackson, D.D., Kagan, Y.Y. & Mulargia, F., 1997. Earthquakes cannot be predicted, *Science,* **275,** 1616–1617.

Gross, S.J. & Rundle, J.B., 1995. A systematic test of time-to-failure analysis, *EOS, Trans. Am. geophys. Un.,* **76(46),** Suppl., F409.

Harris, R.A. & Simpson, R.W., 1996. In the shadow of 1857—the effect of the great Ft Tejon earthquake on subsequent earthquakes in southern California, *Geophys. Res. Lett.,* **23,** 229–232.

Jackson, D.D., 1996a. Earthquake prediction evaluation standards applied to the VAN method, *Geophys. Res. Lett.,* **23,** 1363–1366.

Jackson, D.D., 1996b. Hypothesis testing and earthquake prediction, *Proc. Nat. Acad. Sci. USA,* **93,** 3772–3775.

Jackson, D.D. & Kagan, Y.Y., 1993. Reply to Nishenko & Sykes, *J. geophys. Res.,* **98,** 9917–9920.

Jaumé, S.C. & Sykes, L.R., 1996. Evolution of moderate seismicity in the San Francisco Bay region, 1850 to 1993: seismicity changes related to the occurrence of large and great earthquakes, *J. geophys. Res.,* **101,** 765–789.

Johansen, A., Sornette, D., Wakita, H., Tsunogai, U., Newman, W.I. & Saleur, H., 1996. Discrete scaling in earthquake precursory phenomena: evidence in the Kobe earthquake, Japan, *J. Phys. I. France,* **6,** 1391–1402.

Jordan, T.H., 1997. Is the study of earthquakes a basic science?, *Seism. Res. Lett.,* **68,** 259–261.

Kagan, Y.Y., 1982. Stochastic model of earthquake fault geometry, *Geophys. J. R. astr. Soc.,* **71,** 659–691.

Kagan, Y.Y., 1991. Likelihood analysis of earthquake catalogue, *Geophys. J. Int.,* **106,** 135–148.

Kagan, Y.Y., 1994. Observational evidence for earthquakes as a nonlinear dynamic process, *Physica D,* **77,** 160–192.

Kagan, Y.Y., 1996a. Comment on 'The Gutenberg-Richter or characteristic earthquake distribution, which is it?' by S. G. Wesnousky, *Bull. seism. Soc. Am.,* **86,** 274–285.

Kagan, Y.Y., 1996b. VAN earthquake predictions—an attempt at statistical evaluation, *Geophys. Res. Lett.,* **23,** 1315–1318.

Kagan, Y.Y., 1997a. Seismic moment-frequency relation for shallow earthquakes: regional comparison, *J. geophys. Res.,* **102,** 2835–2852.

Kagan, Y.Y., 1997b. Statistical aspects of Parkfield earthquake sequence and Parkfield prediction experiment, *Tectonophysics,* **270,** 207–219.

Kagan, Y.Y. & Jackson, D.D., 1991. Seismic gap hypothesis: ten years after, *J. geophys. Res.,* **96,** 21 419–21 431.

Kagan, Y.Y. & Jackson, D.D., 1994a. Long-term probabilistic forecasting of earthquakes, *J. geophys. Res.,* **99,** 13 685–13 700.

Kagan, Y.Y. & Jackson, D.D., 1994b. Earthquake prediction: A sorrowful tale, *EOS, Trans. Am. geophys. Un., Suppl.,* **75(25),** 57 (abstract).

Kagan, Y.Y. & Jackson, D.D., 1995. New seismic gap hypothesis: five years after, *J. geophys. Res.,* **100,** 3943–3959.

Kagan, Y.Y. & Jackson, D.D., 1996. Statistical tests of VAN earthquake predictions: comments and reflections, *Geophys. Res. Lett.,* **23,** 1433–1436.

Kagan, Y.Y. & Knopoff, L., 1977. Earthquake risk prediction as a stochastic process, *Phys. Earth planet. Inter.,* **14,** 97–108.

Kagan, Y.Y. & Knopoff, L., 1987. Statistical short-term earthquake prediction, *Science,* **236,** 1563–1567.

Kagan, Y.Y. & Vere-Jones, D., 1996. Problems in the modelling and statistical analysis of earthquakes, in *Lecture Notes in Statistics,* **114,** pp. 398–425, eds Heyde, C.C., Prohorov, Yu.V., Pyke, R. & Racher, S.T., Springer, New York, NY.

Keilis-Borok, V.I. & Kossobokov, V.G., 1990. Premonitory activation of an earthquake flow: algorithm *M8*, *Phys. Earth planet. Inter.,* **61,** 73–83.

Keilis-Borok, V.I., Knopoff, L., Rotwain, I.M. & Allen, C.R., 1988. Intermediate-term prediction of occurrence times of strong earthquakes, *Nature,* **335,** 690–694.

Kelleher, J.A., Sykes, L.R. & Oliver, J., 1973. Possible criteria for predicting earthquake locations and their applications to major plate boundaries of the Pacific and Caribbean, *J. geophys. Res.,* **78,** 2547–2585.

King, G.C.P., Stein, R.S. & Lin, J., 1994. Static stress changes and the triggering of earthquakes, *Bull. seism. Soc. Am.,* **84,** 935–953.

Knopoff, L., 1996. Earthquake prediction—the scientific challenge, *Proc. Nat. Acad. Sci. USA,* **93,** 3719–3720.

Knopoff, L., Aki, K., Allen, C.R., Rice, J.R. & Sykes, L., eds., 1996. *Earthquake Prediction: The Scientific Challenge,* Colloquium Proceedings, *Proc. Nat. Acad. Sci. USA,* **93,** 3719–3837.

Kossobokov, V.G., Healy, J.H. & Dewey, J.W., 1997. Testing an earthquake prediction algorithm, *Pure appl. Geophys.,* **149,** 219–232.

Lighthill, J., ed., 1996. *Critical Review of VAN, Earthquake Prediction from Seismic Electrical Signals,* World Scientific, Singapore.

Lomnitz, C., 1994. *Fundamentals of Earthquake Prediction,* Wiley, New York, NY.

Ma, Z., Fu, Z., Zhang, Y., Wang, C., Zhang, G. & Liu, D., eds., 1990. *Earthquake Prediction: Nine Major Earthquakes in China (1966–1976),* Springer-Verlag, New York, NY.

Main, I.G., 1996. Statistical physics, Seismogenesis, and seismic hazard, *Rev. Geophys.,* **34,** 433–462.

Malone, S., 1996. 'Near' realtime seismology, *Seism. Res. Lett.,* **67,** 52–54.

Mandelbrot, B.B., 1983. *The Fractal Geometry of Nature,* 2nd edn, W. H. Freeman, San Francisco, CA.

Marder, M., 1996. Energetic developments in fracture, *Nature,* **381,** 275–276.

Marder, M. & Fineberg, J., 1996. How things break, *Physics Today,* **49(9),** 24–29.

McCann, W.R., Nishenko, S.P., Sykes, L.R. & Krause, J., 1979. Seismic gap and plate tectonics: seismic potential for major boundaries, *Pure appl. Geophys.,* **117,** 1082–1147.

Mogi, K., 1995. Earthquake prediction research in Japan, *J. Phys. Earth,* **43,** 533–561.

Molchan, G.M., 1997. Earthquake prediction as a decision-making problem, *Pure appl. Geophys.,* **149,** 233–247.

Molchan, G.M. & Kagan, Y.Y., 1992. Earthquake prediction and its optimization, *J. geophys. Res.,* **97,** 4823–4838.

Molchan, G.M. & Rotwain, I.M., 1985. Statistical analysis of long-range precursors of strong earthquakes, in *Computational Seismology,* **15,** pp. 56–70, eds Keilis-Borok, V.I. & Levshin, A.L., Allerton Press, New York, NY.

Molchan, G.M., Dmitrieva, O.E., Rotwain, I.M. & Dewey, J., 1990. Statistical analysis of the results of earthquake prediction, based on bursts of aftershocks, *Phys. Earth planet. Inter.,* **61,** 128–139.

Morowitz, H.J., 1996. Complexity and epistemology, in *Boundaries and Barriers: on the Limits to Scientific Knowledge,* pp. 188–198, eds Casti, J.L. & Karlqvist, A., Addison-Wesley, Reading, MA.

Mulargia, F., 1997. Retrospective validation of time association, *Geophys. J. Int.,* **131,** 500–504 (this issue).

Mulargia, F. & Gasperini, P., 1992. Evaluating the statistical validity beyond chance of VAN earthquake precursors, *Geophys. J. Int.,* **111,** 32–44.

Murphy, A.H., 1993. What is a good forecast—an essay on the nature of goodness in weather forecasting, *Weather Forecasting,* **8,** 281–293.

Murphy, A.H., 1996. General decompositions of MSE-based skill scores—measures of some basic aspects of forecast quality, *Mon. Weather Rev.,* **124,** 2353–2369.

Nishenko, S.P., 1989. Earthquakes: hazards and predictions, in *The Encyclopedia of Solid Earth Geophysics,* pp. 260–268, ed. James, D.E., Van Nostrand Reinhold Co., New York, NY.

Nishenko, S.P., 1991. Circum-Pacific seismic potential: 1989–1999. *Pure appl. Geophys.,* **135,** 169–259.

Nishenko, S.P. & Sykes, L.R., 1993. Comment on 'Seismic gap hypothesis: Ten years after' by Y. Y. Kagan & D. D. Jackson, *J. geophys. Res.,* **98,** 9909–9916.

NRC (National Research Council) Panel, 1991. *Real-time Earthquake Monitoring: Early Warning and Rapid Response,* National Academy Press, Washington, DC.

Ogata, Y., 1988. Statistical models for earthquake occurrence and residual analysis for point processes, *J. Am. stat. Assoc.,* **83,** 9–27.

Ogata, Y., 1998. Space-time point-process models for earthquake occurrences, *Ann. Inst. stat. Math.,* in press.

Ogata, Y., Utsu, T. & Katsura, K., 1996. Statistical discrimination of foreshocks from other earthquake clusters, *Geophys. J. Int.,* **127,** 17–30.

Olsen, K.B., Archuleta, R.J. & Matarese, J.R., 1995. 3-Dimensional simulation of a magnitude-7.75 earthquake on the San-Andreas fault, *Science,* **270,** 1628–1632.

Omori, F., 1894. On the after-shocks of earthquakes, *J. College Sci., Imp. Univ. Tokyo,* **7,** 111–200.

Popper, K.R., 1980. *Logic of Scientific Discovery,* Hutchinson, London.

Reasenberg, P.A. & Jones, L.M., 1989. Earthquake hazard after a mainshock in California, *Science,* **243,** 1173–1176.

Reasenberg, P.A. & Matthews, M.V., 1988. Precursory seismic quiescence: a preliminary assessment of the hypothesis, *Pure appl. Geophys.,* **126,** 373–406.

Reid, H.F., 1910. Elastic rebound theory, *Univ. Calif. Publ. Bull. Dept. Geol. Sci.,* **6,** 92–120.

Rhoades, D.A. & Evison, F.F., 1979. Long-range earthquake forecasting based on a single predictor, *Geophys. J. R. astr. Soc.,* **59,** 43–56.

Rhoades, D.A. & Evison, F.F., 1989. On the reliability of precursors, *Phys. Earth planet. Inter.,* **58,** 137–140.

Rhoades, D.A. & Evison, F.F., 1993. Long-range earthquake forecasting based on a single predictor with clustering, *Geophys. J. Int.,* **113,** 371–381.

Rhoades, D.A. & Evison, F.F., 1996. The VAN earthquake predictions, *Geophys. Res. Lett.,* **23,** 1371–1374.

Roeloffs, E. & Langbein, J., 1994. The earthquake prediction experiment at Parkfield, California, *Rev. Geophys.,* **32,** 315–336.

Ruelle, D., 1991. *Chance and Chaos,* Princeton University Press, Princeton, NJ.

Savage, J.C., 1993. The Parkfield prediction fallacy, *Bull. seism. Soc. Am.,* **83,** 1–6.

Scholz, C., 1997. Whatever happened to earthquake prediction?, *Geotimes,* **42,** 16–19.

Scholz, C.H., 1991. Earthquakes and faulting: self-organized critical phenomena with a characteristic dimension, in *Spontaneous Formation of Space-time Structures and Criticality,* pp. 41–56, eds Riste, T. & Sherrington, D., Kluwer Academic, Dordrecht.

Scrivner, C.W. & Helmberger, D.V., 1995. Preliminary work on an early warning and rapid response program for moderate earthquakes, *Bull. seism. Soc. Am.,* **85,** 1257–1265.

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

Sornette, D. & Sammis, C.G., 1995. Complex critical exponents from renormalization group theory of earthquakes: implications for earthquake predictions, *J. Phys. I. France,* **5,** 607–619.

Stark, P.B., 1996. A few statistical considerations for ascribing statistical significance to earthquake predictions, *Geophys. Res. Lett.,* **23,** 1399–1402.

Stark, P.B., 1997. Earthquake prediction: the null hypothesis, *Geophys. J. Int.,* **131,** 495–499 (this issue).

Svozil, K., 1995. Set theory and physics, *Foundations Phys.,* **25,** 1514–1560.

Turcotte, D.L., 1991. Earthquake prediction, *Ann. Rev. Earth planet. Sci.,* **19,** 263–281.

US Geological Survey, 1996. *Preliminary Determination of Epicentres (PDE), Monthly Listings,* US Dept Interior, National Earthquake Information Centre, Denver, CO.

Utsu, T., Ogata, Y. & Matsu'ura, R.S., 1995. The centenary of the Omori formula for a decay law of aftershock activity, *J. Phys. Earth,* **43,** 1–33.

Varnes, D.J. & Bufe, C.G., 1996. The cyclic and fractal seismic series preceding an m_b 4.8 earthquake on 1980 February 14 near the Virgin Islands, *Geophys. J. Int.,* **124,** 149–158.

Varotsos, P. & Lazaridou, M., 1996. Reply to 'Difficulty of statistical evaluation of an earthquake prediction method', by Utada, *Geophys. Res. Lett.,* **23,** 1395–1398.

Varotsos, P., Alexopoulos, K. & Nomicos, K., 1981. Seven-hour precursors to earthquakes determined from telluric currents, *Praktika Acad. Athens,* **56,** 417–433.

Varotsos, P., Eftaxias, K., Vallianatos, F. & Lazaridou, M., 1996a. Basic principles for evaluating and earthquake prediction method, *Geophys. Res. Lett.,* **23,** 1295–1298.

Varotsos, P., Lazaridou, M., Eftaxias, K., Antonopoulos, G., Makris, J. & Kopanas, J., 1996b. Short-term earthquake prediction in Greece by seismic electric signals, in *A Critical Review of VAN,* pp. 29–76, ed. Lighthill, J., World Scientific, Singapore.

Vere-Jones, D., 1978. Earthquake prediction—a statistician's view, *J. Phys. Earth,* **26,** 129–146.

Vere-Jones, D., 1995. Forecasting earthquakes and earthquake risk, *Int. J. Forecasting,* **11,** 503–538.

Vilenkin, N.Y., 1995. *In Search of Infinity,* Birkhauser, Boston, MA.

Wallace, R.E., Davis, J.F. & McNally, K.C., 1984. Terms for expressing earthquake potential, prediction and probability, *Bull. seism. Soc. Am.,* **74,** 1819–1825.

Weimer, S. & Wyss, M., 1994. Seismic quiescence before the Landers (M = 7.5) and Big Bear (M = 6.5) 1992 earthquakes, *Bull. seism. Soc. Am.,* **84,** 900–916.

White, L., Tursky, B. & Schwartz, G.E. eds., 1985. *Placebo: Theory, Research and Mechanisms,* Guilford Press, New York, NY.

Working Group on the probabilities of future large earthquakes in southern California, 1995. Seismic hazards in southern California: Probable earthquakes, 1994–2024, *Bull. seism. Soc. Am.,* **85,** 379–439.

Wyss, M. ed., 1991. *Evaluation of Proposed Earthquake Precursors,* American Geophysical Union, Washington, DC.

Wyss, M. & Dmowska, R. eds., 1997. Earthquake Prediction—State of the Art, *Pure appl. Geophys.,* **149,** 1–264.