Pure appl. geophys. 155 (1999) 233–258
0033–4553/99/040233–26 $ 1.50 + 0.20/0

© Birkhäuser Verlag, Basel, 1999

Pure and Applied Geophysics

Is Earthquake Seismology a Hard, Quantitative Science?

Y. Y. KAGAN[1]

Abstract—Our purpose is to analyze the causes of recent failures in earthquake forecasting, as well as the difficulties in earthquake investigation. We then propose that more rigorous methods are necessary in earthquake seismology research. First, we discuss the failures of direct earthquake forecasts and the poor quantitative predictive power of theoretical and computer simulation methods in explaining earthquakes. These failures are due to the immense complexity of earthquake rupture phenomena and lack of rigor in the empirical analysis of seismicity. Given such conditions, neither "holistic," interdisciplinary analysis of geophysical data nor greater reliance on the currently available results of earthquake physics is likely to work without revising scientific methodology. We need to develop more rigorous procedures for testing proposed patterns of earthquake occurrence and comparing them to predictions of theoretical and computer modeling. These procedures should use methods designed in physics and other sciences to formulate hypotheses and carry out objective validation. Since earth sciences study a unique object, new methods should be designed to obtain reliable and reproducible results. It is likely that the application of sophisticated statistical methods will be needed.

Key words: Crisis in earthquake seismology, earthquake occurrence, statistical methods, hypotheses testing, fractals, earthquake source models.

1. Introduction

We define earthquake seismology as the study of earthquake source and occurrence. Thus, the problems of elastic wave propagation and the study of earth structure are outside the scope of this paper. It is widely accepted that failure has dogged the extensive efforts of the last 30 years to find 'reliable' earthquake prediction methods, the efforts which culminated in the Parkfield prediction experiment (ROELOFFS and LANGBEIN, 1994, and references therein) in the U.S.A. and the Tokai experiment in Japan (MOGI, 1995). EVANS (1997), GELLER *et al.* (1997), JORDAN (1997), SCHOLZ (1997), SNIEDER and VAN ECK (1997), and WYSS *et al.* (1997) discuss various aspects of earthquake prediction and its lack of success. Jordan (1997) commented that "The collapse of earthquake prediction as a unifying theme and driving force behind earthquake science has caused a deep crisis."

[1] Institute of Geophysics and Planetary Physics, University of California, Los Angeles, CA 90095-1567, U.S.A. E-mail: ykagan@ucla.edu; WEB: http://scec.ess.ucla.edu/ykagan.html

Similar difficulties and failures exist in other earth sciences, however earthquake seismology is a special case because its predictions are so important to seismic hazard reduction. Moreover, in other disciplines, empirical models or a common sense approach seem at least to be weakly predictive in a qualitative way for observations. Hurricanes, flooding, volcanic eruptions and other geophysical catastrophes can be forecast with increasing accuracy as prediction lead time diminishes. By contrast, in earthquake seismology the question remains whether current models have even qualitative predictive power which is assumed to be a property of even 'soft' sciences. For example, for long-term predictions based on the seismic gap hypothesis, it is not clear whether the predictions are better or worse than those made by the null hypothesis, i.e., by random choice (NISHENKO and SYKES, 1993; JACKSON and KAGAN, 1993).

The difference in predictive capability is due to the fact that other geophysical catastrophes involve large mass transport of fluids, which is a relatively slow process. Manifestations of such a transport can be monitored and a warning can be issued, usually a few days or hours, in advance. Earthquakes represent a propagation of fracture with a velocity of km/s, thus if their preparation stage cannot be monitored or, as some evidence suggests (KAGAN, 1997b), is absent, the prediction can only be a statistical one.

As we argue below, current research in earthquake seismology is in a fact-gathering, phenomenological stage, accompanied by attempts to find empirical correlations between the observed parameters of earthquakes. Theoretical models have not been successful in explaining, let alone quantitatively predicting, these phenomena. MARGENAU (1950, p. 28) calls this stage of scientific development "*correlation science*." In other words, earthquake seismology is not a *normal* science in KUHN's (1970) sense of the word: there is no widely accepted *paradigm* explaining the basic properties of earthquake occurrence.

In a recent paper (KAGAN, 1997b) temporal earthquake prediction, interpreted more traditionally, has been reviewed. Therein, I argue that, earthquake prediction as understood by the majority of geophysicists and the general public—the prediction of specific individual events—is not possible. However, earthquake rates can be predicted either in a time-independent mode (seismic hazard) or as a temporally dependent rate. The major problem, presently unknown, is whether we can achieve magnitude-dependent prediction, i.e., a prediction of magnitude distribution of future earthquakes statistically different from the GUTENBERG-RICHTER (1944) law.

Here we do not focus on the question of earthquake predictability in a narrow sense (KAGAN, 1997b), but rather on a broader question, why recent research in earthquake seismology has not achieved the success of other sciences. Despite an increase in the quality and quantity of data collected, there has been no major breakthrough; no significant progress has been made in understanding seismicity and earthquake occurrence.

Two approaches are often proposed as a solution to this crisis: (1) "holistic" analysis of geophysical data (JORDAN, 1997), (2) further reliance on earthquake physics (BEN-ZION *et al.*, 1999). As stated below, these approaches alone are unlikely to produce fruitful results since the major problem is methodological: earthquake research needs a significantly higher level of rigor. If we replace one paradigm by another and continue to conduct our research as before, relatively little will change. Thus, we argue that the major challenge facing earthquake seismology is not a paradigm shift but the application of more rigorous methods of analysis. Moreover, new methods for hypotheses verification need to be developed. The development of many sciences such as medical, biological, agricultural research, etc., especially in the second half of the 20th century, has been characterized by their transformation from qualitative "soft" disciplines to quantitative ones. This transformation has been accompanied by the application of increasingly sophisticated statistical methods.

The above distinction between mathematically based hard sciences and descriptive soft sciences can be traced back to the beginning of the scientific method in Western civilization (SANTILLANA, 1961, pp. 214–222). Whereas astronomy, and physics in general, used the Pythagorean-Platonic paradigm of an abstract mathematical model applied to observed phenomena, the Aristotelian scheme of descriptive categorization and teleological understanding dominated biological sciences. It is interesting to note that whereas biology is now moving towards mathematization, geology is still largely a descriptive science, plate tectonics being an analog of the 19th century Darwinian theory.

Below, we discuss a broader definition of prediction, summarize the failure of present research methods, and attempt to explain their causes. In the final section we consider more rigorous methods that can be applied to verify hypotheses of earthquake patterns.

2. *Prediction*

Quantitative prediction is the aim of every science. BEN-MENAHEM (1995, p. 1217) says: "[T]he ultimate test of every scientific theory worthy of its name, is its ability to predict the behavior of a system governed by the laws of said discipline." By quantitative prediction, in a broad sense, we mean any new numerical information on earthquake occurrence, be that the distribution of size, location, time, or other properties. ENGELHARDT and ZIMMERMANN (1988, p. 203) define *retrodiction* as a "prediction" of a past state, and *codiction* as a "prediction" of a simultaneous state. Thus, on the basis of a given hypothesis, we can "predict" past earthquake occurrence or the present state of earthquake-related variables in a certain region. Because of the stochastic nature of the earthquake process (KAGAN, 1992, 1994), the prediction must be statistical.

"The practical demand for the increased use of factually and hypothetically founded prognoses requires a complete turn around in the geologist's accustomed point of view. This exposes him to a danger he does not encounter in making retrodictions: each prognosis will at some time be proven right or wrong, and he who made it must reckon with the possibility that the course of events or the very measures he proposed may "refute" it. The backward-looking geologist is safe from such dangers, for the ambiguity of abductive inference and the impossibility of "direct" proof make it impossible to disprove 'once and for all' certain retrodictions or abductive codictions, possibly setting long years of research at nought" (ENGELHARDT and ZIMMERMANN, 1988, p. 230). By *abduction* the authors mean a logical determination of an initial state of a natural system, if the final state and the controlling laws are known.

The citation above points out the challenges that exist in earthquake seismology. In other scientific disciplines the failure of a model to explain and predict phenomena can be established by carrying out an appropriate experiment. The long-term recurrence time for large earthquakes makes it difficult to evaluate earthquake prediction methods. At the same time, occurrence of an unpredicted earthquake, especially in regions covered by dense observational networks, demonstrates our deficient understanding of earthquake process. Some of these unpredicted events are destructive, thus dramatizing for the general public the fundamental failure of the science.

3. Failures

3.1. Earthquake Prediction: Failures and Difficulties

3.1.1. Seismic hazard—limited success. Earthquake hazard estimates are presently based on earthquake geology, i.e., a description of active faults (COBURN and SPENCE, 1992; YEATS *et al.*, 1997). Failures are obvious—most large earthquakes in California in the 20th century (1952 Kern County, 1971 San Fernando, 1989 Loma Prieta, 1992 Landers, 1994 Northridge) occurred on faults that had not been properly identified earlier. Similarly, several earthquakes of unexpectedly large magnitude occurred recently in Spitak, Armenia (1988), Netherlands (1992), Latur, India (1993), and North Sakhalin (1995).

Although geology, paleoseismology in particular, has achieved significant progress in assessing past seismicity, the problems of the non-uniqueness of its estimates are still very serious (WELDON, 1996). While estimates for seismic hazard due to intermediate (M 5–M 7) earthquakes in seismically active regions may differ by a factor of about 2–3, the difference may be as large as a few orders of magnitude for large (M 7–M 8) and very large ($M > 8$) earthquakes. The uncertainty in the maximum magnitude size is especially pronounced at intracontinental areas with a low level of contemporary seismicity (KAGAN, 1997b).

Standard estimates of earthquake size distribution are based on the characteristic earthquake model (KAGAN, 1997a,b). The model assumes that a maximum earthquake depends on the length of a causal tectonic fault or fault segment (SCHWARTZ and COPPERSMITH, 1984). Although no quantitative model has been proposed to identify faults and no critical testing of the hypothesis has been carried out by its proponents, the characteristic model is generally accepted in hazard estimates. KAGAN (1997a,b) demonstrates that this model greatly underestimates the size of the largest earthquakes, and the prospective (forward) testing of the model reveals its failure to describe regional magnitude-frequency relation (KAGAN, 1997b).

Several recent investigations (TANIOKA and GONZALEZ, 1998; SCHWARTZ, 1999) obtained maps of slip distribution for large earthquake pairs. These maps suggest that the slip pattern is highly non-uniform, and the slip distribution for subsequent earthquakes significantly differs from a previous event. This pattern contradicts the central assumption of the characteristic earthquake model (SCHWARTZ and COPPERSMITH, 1984): that slip distribution during characteristic events is nearly identical.

Recently, repeating microearthquakes have been studied in the Parkfield area (NADEAU and MCEVILLY, 1997; NADEAU and JOHNSON, 1998, and references therein). These event sequences exhibit certain properties of characteristic earthquakes: regularity of occurrence, nearly identical wave forms and illustrate many interesting features of earthquake occurrence. However, these microearthquakes are not characteristic earthquakes in a strict sense: these events do not release most of the tectonic deformation on a fault segment as the real characteristic earthquakes are assumed to do (SCHWARTZ and COPPERSMITH, 1984). As with large characteristic earthquakes, attempts to predict these microearthquakes using their quasi-periodicity have not yet succeeded (R. Nadeau, private communication, 1997).

3.1.2. Long-term prospective predictions—unsuccessful or inconclusive results. Recently several experiments were carried out to forecast long-term earthquake potential. Most of these attempts were based on the seismic gap model which was first proposed by FEDOTOV (1968). SYKES (1971) reformulated it on the basis of plate tectonics and MCCANN *et al*. (1979) developed a map of circum-Pacific seismic zones, including a forecast for each zone. NISHENKO (1991) based his prediction of seismic activity in the Pacific zones on a new seismic gap model which considers the recurrence time and characteristic earthquake magnitude specific to each plate boundary segment. These authors apparently did not anticipate the necessity for rigorously testing the model performance. However, it was a great service to science that they produced a prediction that can (admittedly with some difficulty) be tested, i.e., the seismic gap model was formulated as a falsifiable hypothesis.

The testing of the McCann *et al.* (1979) and Nishenko (1991) forecasts by Kagan and Jackson (1991, 1995) showed that no version of the seismic gap hypothesis has yet evidenced a significant statistical advantage over a reasonable null hypothesis (Kagan, 1997b). There is a possibility that due to the long-term clustering property of earthquakes, the gap model performs significantly worse than the Poisson (null) hypothesis, although Nishenko and Sykes (1993) dispute this conclusion.

Two prediction experiments which generated immense public interest—Tokai, Japan (Ishibashi, 1981; Mogi, 1995) and Parkfield, California (Bakun and Lindh, 1985)—are also based on the seismic gap model. However, in contrast to the investigations mentioned above, these experiments were not planned as falsifiable statements. Despite the significant cost of these prediction experiments, the seismic gap hypothesis which was their basis was not tested rigorously either before or during the experiments. As Kagan (1997a) indicates, even if an earthquake similar to the expected Parkfield event were to occur, no statistically significant conclusions could be drawn. Meanwhile, damaging earthquakes for which no specific long-term forecasts were issued, occurred elsewhere in Japan (1993 Okushiri Island, 1995 Kobe) and California (1992 Landers, 1994 Northridge). Although one may expect that unanticipated earthquakes may shake some remote areas, these events occurred in places covered by one of the densest observational networks, and caused significant damage. Thus the absence of any reasonable estimate of seismic hazard for these regions seems especially embarrassing.

The M 8 algorithm was proposed about 15 years ago by using a pattern-recognition technique (Kossobokov *et al.*, 1997) to predict long-term earthquake probabilities. The prospective testing of the model was carried out in 1992–1997 for large ($M \geq 7.5$) earthquakes in the Pacific area. Preliminary test results indicated that the algorithm performs no better than the null Poisson hypothesis (*ibid.*). Although final test results show improvement in the hypothesis' performance (Kossobokov *et al.*, 1998, 1999), the model's predictive skill for $M \geq 7.5$ earthquakes is apparently low.

The only prediction experiment planned from the beginning as a technically rigorous test of the hypothesis is the precursory swarm model formulated by Evison and Rhoades (1993, 1997). In these experiments all necessary parameters of the earthquake forecast and statistical criteria for acceptance/rejection of the hypothesis were specified before the tests were started. Both tests revealed that the research hypothesis (the precursory swarm model) does not exhibit a statistically significant advantage over the null hypothesis, i.e., the test results are negative. Tests of the modified model are ongoing.

An often expressed opinion (see, for example, Gusev, 1998) is that since the tectonic deformation responsible for earthquakes is slow, we need decades or even centuries to accumulate sufficient data to test earthquake occurrence hypotheses. This may be true for certain aspects of models, however as discussion in this subsection testifies, it is possible to design experiments so that testing of long-term forecasts is accomplished on a considerably shorter timescale, i.e., in a few years.

3.1.3. Short-term prediction. Several algorithms have been proposed to use foreshock-mainshock-aftershock patterns for short-term earthquake prediction (KAGAN and KNOPOFF, 1987; REASENBERG and JONES, 1989; MICHAEL and JONES, 1998). Appropriate models for short-term clustering are very important even for studying long-term effects. Seismicity rate increases by a factor of many thousands in the wake of earlier earthquakes, whereas long-term rate changes are less than one order of magnitude (KAGAN, 1994). Thus we cannot see these long-term signals in the presence of stronger short-term effects.

However, geophysicists are primarily concerned with deterministic short-term earthquake prediction—the search for immediate earthquake precursors that would allow reliable forecasting on a time scale of days and hours. A recent debate on the VAN predictions (GELLER, 1996; KAGAN and JACKSON, 1996) shows clearly that the geophysical community lacks unambiguous methods to test even short-term predictions. Although there is no consensus on this issue, an extensive search for deterministic short-term precursors has been successful (GELLER *et al*., 1997; WYSS *et al*., 1997). Failure to find any reliable earthquake "precursors" (KAGAN, 1997b) may indicate that predicting individual earthquakes is inherently impossible.

3.2. Earthquake Phenomenology

A significant number of publications concerning earthquake phenomenology exists: for example, according to the INSPEC database, a paper on earthquake size distribution (magnitude-frequency relation and its variants) is published about once a week. A vast diversity exists in the description of earthquake size distributions, spatial patterns of earthquake focal zones, temporal regularities of earthquake occurrence, and temporal interrelation between earthquakes (KAGAN, 1994, 1997b).

In my opinion, most (90% or more) of the seismicity patterns published, even in the best geophysical journals, are partly or wholly artifacts. It is impossible, of course, to substantiate this definitely, but the discussion in the previous subsections (see also more in KAGAN, 1994, 1997b) demonstrates that efforts and predictions of many experienced researchers proved fruitless when tested against a future earthquake record. These investigations were carried out in the "mainstream" of earthquake seismology, thus subject to special attention and criticism. Because no unifying theoretical principle exists, a significant part of the phenomenological research explores specific problems, and thus is outside the normal cross-validation and result duplication which is normative in other quantitative sciences. Such diversity of research directions and lack of requirements for objective model testing results in lower standards.

Are large earthquakes quasi-periodic or clustered in time? What is the size distribution of the largest earthquakes in a particular area? Is the distribution characteristic or Gutenberg–Richter? Is the geometry of earthquake faults or short-term earthquake temporal behavior scale-invariant or are there several different length and timescales (see, for example, OUILLON *et al*., 1996, and references therein)? The differences in opinion are so great to date that no common ground has been found and, most importantly, the debate has continued for many years without resolution, i.e., the research seems largely ineffective.

3.3. Theory

The term *earthquake physics* usually denotes the physical processes and conditions that govern the occurrence of sudden slip on earthquake faults, e.g. a slip initiation and arrest, dynamic rupture, as well as all supporting processes and conditions which lead to an earthquake, including fault zone structure, tectonic loading, stress evolution and redistribution within the crust, and the operative boundary conditions. No comprehensive theory of earthquake generation process or fundamental equations exist. Thus, earthquake physics represents a set of ideas, some of them simple, based on observations of earthquake deformation and common sense taken from engineering disciplines such as rock mechanics, friction physics, fracture mechanics, and materials science, and from recent developments in statistical physics and nonlinear sciences.

Since there are no current comprehensive reviews of earthquake physics, for a more specific discussion see SCHOLZ (1990); NEWMAN *et al*. (1994, in particular, GABRIELOV and NEWMAN, 1994); KAGAN (1994); KNOPOFF *et al*. (1996); MAIN (1996); SCHOLZ (1998); BEN-ZION *et al*. (1999). Existing physical models of earthquakes are often mutually contradictory. Lack of agreement on the fundamental principles of earthquake rupture and seismicity patterns modeling is demonstrated by the papers cited above.

3.3.1. Elastic rebound theory. The elastic rebound theory was first formulated by REID (1910) following his analysis of the 1906 San Francisco earthquake. SCHOLZ (1997, p. 18) states with reference to the rebound theory: "... the next earthquake is likely to happen when the strain released in the previous one has been restored." In recent years the rebound model has become the basis for several hypotheses, such as earthquake recurrence, earthquake cycle, and seismic gap models (EVANS, 1997). These models, in turn, have been used in long-term prediction efforts. Tests of the seismic gap hypothesis (section 3.1.2) show that its performance is not better than the Poisson model. Researchers who feel that the recurrence hypothesis is correct and that the failure of the seismic gap models can be explained by inappropriate applications of the seismic cycle model, will do well to follow the example of MCCANN *et al*. (1979) and NISHENKO (1991) and propose a testable, falsifiable prediction of future seismicity. Such a forecast should be made for a sufficiently large area, to be verifiable in a reasonable time span.

The problem with the Reid model is whether stress concentration and release is local, i.e., involving 10–15 km along a fault trace, or, as California data suggest (JACKSON *et al.*, 1997) there is no "strain concentration" around faults expected to slip in the near future, and the old concentrations of strain rate are after-slip of events (including those long past). This finding would suggest that strain builds up on a broad regional scale and is randomly (stochastically) released somewhere. The presence of large foreshocks or aftershocks (some of them comparable in size with a mainshock) indicates that stress is not fully released even by a very large earthquake. Geodetic measurements of deformation along known tectonic faults should help to resolve this problem.

3.3.2. Dilatancy-diffusion model. The dilatancy-diffusion hypothesis (an increase in rock volume prior to failure) was proposed in the late 1960s–early 1970s (SCHOLZ *et al.*, 1973; GELLER, 1997; SCHOLZ, 1997, and references therein) to explain the variation in elastic wave velocities, as well as other geophysical anomalies observed before earthquakes. Subsequent measurements indicated that the original observations may have been artifacts. Thus the dilatancy model was largely abandoned by the late 1970s. SCHOLZ (1997) suggests that this rejection may have been premature: "[T]hese results were accepted by consensus as deeming the failure of both the theory and of that particular form of earthquake precursor, neither of which has been seriously investigated since." We argue below, that such lack of critical testing and rigorous verification characterizes not only the dilatancy hypothesis but all others as well.

3.3.3. Rock friction in laboratory and earthquake models. Various forms of rock friction law are offered to explain and predict earthquake rupture and occurrence (SCHOLZ, 1998, and references therein). The original friction law was formulated from laboratory experiments and invoked to explain earthquakes *in situ*. However, the conditions in the laboratory experiments are radically different from the conditions *in situ*. The laboratory experiments are conducted in a block with (usually) a single polished fault, contained in an apparatus. Thus the rock friction law is based on the mechanics of man-made, engineered objects. Three clearly defined, highly different characteristic geometrical scales can be distinguished in such objects: (1) exterior or macroscopic, which applies to an apparatus and the sample size; (2) interior or microscopic, which describes the defects and intergranular boundaries of the object material; and (3) the irregularities of the object surface.

The factors connected with the latter scale are usually taken into account by the introduction of frictional forces (SCHOLZ, 1998). The analysis of fault geometry (KAGAN, 1994) indicates that earthquakes do not occur on a single (possibly fractally wrinkled) surface, but on a fractal structure of many closely correlated faults (GABRIELOV and NEWMAN, 1994). The total number of these small faults can be very large, practically infinite.

Similarly, scales (1) and (2) are usually separated in materials science and in engineering applications by introducing effective properties of the material (LEARY, 1997). In earthquake studies we consider the propagation of a rupture through rock material which, during millions of years of its tectonic history, has been subjected to repeated deformation. Material properties of rocks are scale-invariant (CHELIDZE, 1993; LEARY, 1997; TURCOTTE, 1997). Even if natural rock specimens are used in laboratory experiments, these specimens are carefully selected for internal defects to be small compared to the size of the sample. Thus the internal structure of a specimen resembles the uniform material properties characteristic of engineered materials.

Most of the mechanical models of earthquake occurrence are based on Euclidean geometrical forms that are plane surfaces of rupture, whereas the geometry of natural tectonic forms is fractal (CHELIDZE, 1993; MAIN, 1996; KAGAN, 1994). This is only one reason to doubt the relevance of laboratory experiments for earthquake prediction.

3.3.4. Computer seismicity simulation. Ideas from the engineering practice of fracture mechanics and rock friction are used in many computer models of earthquake occurrence, like various modifications of block-friction models (KNOPOFF *et al*., 1996). Hundreds of papers treating computer simulation of earthquakes have been published in physics and geophysics journals in the last decade. As was mentioned above, these models are not based on the fundamental laws of physics, and the mechanical and geometrical properties of these constructs differ radically from natural fault systems. These models must be justified and confirmed by carefully comparing their results with earthquake data.

Several concepts taken from solid state physics (critical phenomena) or self-organized criticality have been proposed to model earthquake phenomena, although as in mechanical models, their relevance to earthquake problems is unclear, partly because a comparison of the results drawn from these theories with the experimental data is inconclusive. Although the ideas borrowed from nonlinear mechanics, critical point phenomena, and self-organized criticality explain in general the scaling properties of earthquakes and provide new tools with which to describe seismicity, these methods are still of limited value (KAGAN, 1994; MAIN, 1995, 1996).

Two reasons explain the ineffectiveness of such research: (1) considerable theoretical work is conducted in a corroborative vacuum: there is very little confrontation between experimental facts and results of simulations; (2) as discussed in section 3.2, the seismological observational evidence is inadequate and insufficient: the fundamental regularities of earthquake occurrence are still in doubt. Such a diversity of opinion makes it relatively easy to "confirm" simulation results, using one or another of the contradictory empirical "observations." Thus the major difficulty we face in modeling earthquake occurrence arises not only because of deficient models, but also from the lack of reliable observational results to verify the models' predictions.

The results of synthetic models are usually compared to the power-law distribution for event size and the value of the distribution's exponent. However, earthquake size distribution is the least informative of all power laws that govern earthquake occurrence: the power-law size distribution can be obtained from a variety of models (KAGAN, 1994); the fact that events in a model follow a fractal distribution, even with a correct exponent value, does not prove that the model can describe seismicity.

Short-term earthquake clustering, the major feature of a shallow earthquake occurrence, is not reproduced by most models, casting doubt on the applicability of their results to real earthquakes. Some models generate either aftershocks or foreshocks, although evidence exists (KAGAN, 1994; MICHAEL and JONES, 1998) that both aftershocks and foreshocks manifest the same clustering process. Therefore they should be explained by the same mechanism. In addition, during a rupture, each earthquake can be regarded as a dynamic fusion of many infinitesimal events (KAGAN, 1982); a model which generates a clustered earthquake sequence should explain a complex internal structure of each earthquake as well.

There is increasing information regarding the complexity of earthquake source time release; currently an earthquake catalog with source time functions is available on the WEB (RUFF and MILLER, 1994). Such data may soon allow one to obtain a statistical description of the source function complexity. When such results become available, quantitative testing of theoretical models can be carried out pertaining to whether these models correctly predict statistics of seismic moment release complexity, multiplicity of earthquake sources, etc.

Since most simulations do not generate foreshocks and aftershocks, only temporal interaction for main events can be investigated. Mainshock sequences can be fairly well approximated by a Poisson process which indicates that these earthquakes are largely statistically independent. There is little information in sequences of independent events; as in the power-law size distribution, such time series can be produced by a variety of mechanisms.

3.3.5. Laboratory fracture experiments. In a laboratory, a crack develops instabilities which makes its propagation highly chaotic and unpredictable (XU *et al.*, 1997; KALIA *et al.*, 1997; MARDER, 1998; SHARON and FINEBERG, 1999). The instabilities and sensitive dependence on initial conditions are due to crack propagation, especially at a speed close to elastic wave velocity. Stress and fracture conditions in laboratory specimens differ significantly from those in earthquake fault zones: the boundary effects are controlled by the laboratory researcher. Therefore, fracture can self-organize only at spatial scales substantially smaller than those of the specimen. In the fault zones stress, rock mechanical properties and fault geometry are self-organized as large-scale self-similar patterns develop.

Consequently, even if a good predictive capability can be achieved in modeling laboratory ruptures, transferring this capability to real earthquakes is completely different. As explained, in the laboratory: (1) the spatial domain of fracture is well defined; (2) the boundary conditions are well known and controlled. Conversely, in the earth's crust: (1) fracture can occur virtually anywhere; (2) the boundary conditions are unknown primarily because the boundaries themselves cannot be defined. Therefore, the relevance of laboratory results to actual earthquakes and the predictive power of these results must be demonstrated.

3.3.6. Stress triggering. Stress in the earth's interior where earthquakes occur is not practically measurable. Moreover, stress is a tensor, not a scalar, with significant complications ensuing. In stress diagrams (Kagan, 1994; King *et al.*, 1994; Deng and Sykes, 1997) stress tensor invariants (or components) form a complicated mosaic, even when only moderate and large earthquakes are accounted for. This mosaic should become increasingly more complex if we calculate stress for smaller earthquakes.

Stress, earthquake size, and other parameters of seismicity have a power law or a fractal distribution with an exponent value of less than one (Kagan, 1994). This results in extreme stochastic variability of seismicity parameters. The statistical variability can easily be confused with regional or temporal variations of seismicity. The statistical properties of power-law variables are not well known and are the subject of current intensive research (Samorodnitsky and Taqqu, 1994; Adler *et al.*, 1998). These properties significantly differ from more familiar stochastic quantities with a finite second moment, such as Gaussian variables. Neither the average nor the correlations are defined for such variables: one should use quantiles/percentiles and codifferences, respectively (*ibid.*). One may argue that in the real earth the stress singularities are smoothed out one way or another, and as soon as the stress is finite, all statistical moments would also be finite. However we do not yet know how to handle these singularities. Thus the results would strongly depend on how the stress infinities near the earthquake focal zone are treated. These stresses in focal zones of earlier earthquakes trigger the next events and define the way earthquake rupture propagates and stops.

To verify the ability of the stress triggering patterns inferred from observations to predict future earthquake activity, it is important to make forward predictions. These predictions must be compared with the actual seismicity record. The predictions should also be tested against a null hypothesis which uses known statistical properties of earthquake occurrence to specify time, space, and focal mechanism for future events (Kagan and Jackson, 1994). Only when the stress triggering method outperforms the empirical based algorithm can the predictive ability of the method be considered verified (cf. Hardebeck *et al.*, 1998).

3.3.7. Theory—critical review. The major problem with earthquake physics is that—given the lack of appropriate testing—it is not known which models (if any) can reproduce the fundamental properties of earthquake occurrence quantitatively. Presently we are uncertain whether theoretical or computer models of rupture are even approximately correct, the predictions of these models are tested only qualitatively and never in an actual prospective mode. Even if some models are successful in engineering applications, i.e., within the narrow range of laboratory experiments, we still must show that these constructs are useful in fundamentally understanding the phenomena.

At the present time, numerical earthquake models have shown no predictive capability exceeding or comparable to the empirical prediction based on earthquake statistics. Even if a theoretical or physical model exhibits some predictive skill (DIETERICH, 1994), we should always question whether the predictive power derives from a deeper theoretical understanding, or from the earthquake statistics which are imbedded in the model. A model may have a large number of adjustable parameters both obvious and implied to simulate a complicated pattern of seismic activity successfully. However, the model may not have a theoretical predictive capability. Thus we should request that model's prediction must outperform an empirically based one.

Earthquake physics as practiced now employs mathematical tools that were designed before 1870. Substantial progress has been made in mathematics subsequently with the appearance of set theory, topology, group theory, and the theory of stochastic processes, to mention just some new mathematical disciplines. The necessity of new mathematical approaches to earthquake seismology problems is highlighted by scale-invariant, fractal properties of seismicity, discussed above. These properties do not only mean that empirical distributions can be approximated by a straight line in a log–log plot. The scale-invariant geometry of earthquake faults signifies, for example, that the geometrical and mechanical characteristics of these objects would radically differ from those of solid bodies with more familiar Euclidean forms (MANDELBROT, 1983).

PENROSE (1989, p. 125) asks whether the Mandelbrot set (MANDELBROT, 1983, pp. 188–189) is computable: is there any computer procedure that in a finite number of steps would decide that an arbitrary point in a complex plane belongs to the set? The answer proves to be "*no*" (BLUM *et al.*, 1998, p. 55); even if we use the *real-number* arithmetic operations, no algorithm can decide whether a point is in the set. BLUM *et al.* (1998) determine that the reason for the "undecidability" of the Mandelbrot set and many similar complex mathematical objects, is that their boundary has a fractal Hausdorff dimension.

As discussed earlier in this section, earthquake occurrence in all variables is characterized by the scale-invariance and fractal dimensions. Earthquake fault system, for example, is a fractal object (KAGAN, 1982, 1994), thus its boundary cannot be computed in a continuum limit (BLUM *et al.*, 1998). Moreover, the stress at the fractal boundary should be nowhere differentiable function (see subsection 3.3.6), thus it is possible that the calculation of earthquake rupture criteria for points close to a "fault-tip" cannot effectively be carried out. It seems probable, therefore, that the application of new mathematical tools would be necessary to create a comprehensive physical theory of earthquake occurrence.

The diversity of earthquake models testifies to the lack of a common disciplinary matrix (KUHN, 1970). These theoretical models have not yielded conclusions that can be verified or rejected by comparing them to actual earthquake data, i.e., the hypotheses are not falsifiable (POPPER, 1980). Currently, these theoretical studies have contributed no identifiable new knowledge regarding actual earthquakes. Hence earthquake physics fails to function as a *theoretical science* (MARGENAU, 1950). One can argue that the physics of earthquakes does not yet exist. As a result, we have no clear idea which research directions are worth pursuing.

4. Causes of Failures: Lack of Rigor, Hypotheses are not Falsifiable

POPPER (1980) introduced the notion of falsifiability for two essential reasons: (1) to reject the hypotheses which contradict observational evidence and (2) for demarcation—to identify models and hypotheses which are unscientific. As we discussed above, the models, hypotheses, and experiments employed in earthquake seismology are often not falsifiable even in principle. For example, the earthquake recurrence hypothesis (section 3.3.1) has never been formulated in terms amenable to rigorous testing. How do we test the elastic rebound theory and its derivative models? Does it predict that earthquake rupture is not repeated on the same fault plane at short time intervals? What are the formal tolerance limits? How is the rupture zone defined? Since slip varies along the fault surface (if the surface can be unambiguously defined), how do we draw boundaries? To verify the earthquake recurrence model we should formulate a formal hypothesis and a null hypothesis and test both of them against independent data.

Below we briefly list several common deficiencies of earthquake occurrence models and data processing methods (some of these are proposed for case studies by WYSS, 1991 and WYSS and DMOWSKA, 1997):

(1) Non-uniqueness of earthquake occurrence models. In the previous section we mentioned that theoretical models of seismicity are not specific enough to be tested against real data.

(2) Drawing conclusions from extremely small samples (e.g., a few earthquakes in one or a few small regions). Moreover, most studies and predictions are retrospective. Such sampling obviously introduces biases, as the selection process searches for patterns, and almost any pattern can be found if enough data are sifted. The mathematical discipline called Ramsey theory (GRAHAM and SPENCER, 1990) states that a surprisingly broad range of patterns can be found in even a small amount of data. Retrospective prediction "successes" may also be due to selection bias.

(3) The number of adjustable parameters in a statistical model is often comparable to the effective number of data points, i.e., data are overinterpreted or overfitted. Hidden degrees of freedom and systematic effects must be carefully

investigated before the results of statistical studies can be presented. A probably apocryphal statement attributed to Enrico Fermi holds, that with four adjustable parameters one can approximate an elephant.

(4) *Ad hoc* and *ex post facto* adjustment of prediction hypotheses, after they failed in original tests.

(5) Failure to employ statistical methods properly, if at all. Retroactive adjustment of parameters to produce the best correlation must be taken into account when evaluating the statistical significance of retrospective tests (MULARGIA, 1997).

(6) Use of inappropriate null hypotheses in statistical tests (KAGAN, 1997b; STARK, 1997). Earthquakes are known to cluster in space and time (foreshock-mainshock-aftershock sequences). Inappropriate "strawman" null hypotheses (e.g., uniform randomness in space and time) can be rejected with an apparently high level of confidence even when the working hypothesis has no merit (KAGAN and JACKSON, 1996).

(7) Non-reproducibility of results. A case history is the usual method for investigating seismicity; the tendency is to collect the maximum possible information from different scientific disciplines and integrate this information into a complete, holistic picture of earthquake occurrence. The problem with these descriptive results is poor reproducibility. Even if a few studies seem to produce similar results, it is not clear whether this replication is caused by similar prior assumptions and selection bias (i.e., the reproducibility is illusory), or an underlying physical cause. Many papers have been published, for instance, to confirm the seismic gap model using particular case histories (NISHENKO and SYKES, 1993; JACKSON and KAGAN, 1993, and references therein). However, forward tests of this model (section 3.1.2) demonstrate that it lacks predictive power.

The difficulties described above can be ignored if applying current methodology such as an extensive descriptive, phenomenological analysis of earthquake case histories would result in success. In the Parkfield and Tokai prediction experiments major efforts were made to accumulate and integrate geologic, geodetic, and seismic information for an in-depth analysis of the data. However, both experiments have apparently failed. One may argue that both experiments were designed to detect short-term precursors, and since no large earthquake occurred in the respective areas, the experiments are still continuing. Nonetheless, for what duration should monitoring of putative short-term precursors be continued to be considered productive? KAGAN (1997a) argues that the return time for a $M \approx 6$ earthquake in the Parkfield area may exceed 100 years. For the seismic gap hypothesis or the $M8$ method, the results of the test are obtained in a few years (subsection 3.1.2). Thus experiments can be designed to provide an answer much sooner than the Tokai and Parkfield efforts. Nevertheless, despite failure in the Parkfield and Tokai tests, the seismological community still accepts the premises on which these experiments are based, with the exception of a few investigators. A likely failure of the seismic gap models (section 3.1.2) has not generally been acknowledged, and has not led to critical large-scale tests of the earthquake rebound models (section 3.3.1).

5. Recommendations and Conclusions

The failures described in the previous sections imply that a review of the foundations of earthquake science is necessary. Whereas valid scientific reasons for disagreement on earthquake predictability may exist (WYSS *et al.*, 1997), most researchers agree that the rules for seismicity pattern investigations as well as reporting of the results must be significantly tightened. However, implementing more rigorous rules for data analysis is likely to be difficult and contentious. Let us consider a few methodological criteria relevant to the problems of earthquake seismology.

5.1. Simple versus Complex Models

Two explanatory approaches for seismic patterns can be proposed: (1) Introduce more *ad hoc* factors influencing seismicity, in effect introducing as many degrees of freedom in a model as the number of features the hypothesis seeks to explain. As shown in the previous section, the success of this approach is small to non-existent. (2) Treat most of the observed features as a random phenomenon. Only major features should be analyzed by using extensive statistical data. These simple patterns of earthquake occurrence may characterize the universal properties of seismicity, and such investigations are reproducible.

KIRCHHERR *et al.* (1997, see also LI and VITANYI, 1997) consider how one should treat simple versus complex models, accounting for the Occam's razor principle and Bayesian reasoning, from the point of view of Kolmogorov's complexity. They suggest that each model should be assigned a prior probability equal to its complexity, with these probabilities used in the Bayesian posterior evaluation of observational evidence. According to this rule, simple models have a strong initial advantage in explaining phenomena. Such a methodology corresponds to usual scientific practice: simple hypotheses are not rejected unless there is compelling evidence against them. Thus, although the earth is very complex (ORESKES *et al.*, 1994), unless a complicated hypothesis is based on a properly verified understanding of basic relations, a complex model would lack predictive power and would remain a descriptive, phenomenological tool.

5.2. Reductionist versus Emergent Phenomena Approach

JORDAN (1997) argues that earthquake seismology may not succeed using the reductionist program, which explains the complexity of natural phenomena by a few fundamental equations. Following ANDERSON's (1972) arguments, Jordan suggests that due to the complexity of geosystems, new "emergent" laws must be found as explanatory constructs for earthquake science. There is no doubt that the earth is an extremely complicated system; however, some observational regularities of earthquake rupture and occurrence, such as the power-law size distribution, are seen in relatively simple systems, such as a rock specimen or man-made objects (MAIN, 1996).

Calculations of molecular dynamics (KALIA *et al*., 1997; MARDER, 1998; SHARON and FINEBERG, 1999) demonstrate that the basic properties of tensile fracture can effectively be derived from simple laws. Similarly, precise laboratory measurements of fault propagation demonstrate multiple branching of fault surfaces. These simulations reproduce the fractal character of a fracture. Moreover, calculating the total energy balance in laboratory fracture experiments (SHARON and FINEBERG, 1996, 1999) establishes that most elastic energy goes into creating new surface. Although the conditions during tensile fracture differ from those of shear failure in earthquakes, the above result may bear on the problem of the heat paradox for earthquake faults (SCHOLZ, 1996). One should anticipate that these computer simulations, even if extended to shear fracture, would need to be combined with continuum-style models of material rupture. The new, yet unknown, emergent effects may appear during a transition from one to another model.

5.3. Seismicity and Turbulence

Tectonic earthquakes and rupture in a laboratory exhibit scale-invariance over a very broad range of distances 10^{-3}–10^5 m. Similar scaling, even over an extended distance range, is observed for another large-scale deformation of matter—turbulence of fluids (KAGAN, 1992, 1994). Both phenomena share stochastic scale-invariance, have hierarchical space-time structures, and multiple power-law dependencies. Turbulence is characterized by an energy transport cascade from large-scale to smaller structures (MANDELBROT, 1983; MOLCHAN, 1997). KAGAN (1973) proposes a stochastic model where time-space-magnitude features of seismicity are represented as a cascade in energy dimension from the largest to the smallest earthquakes. This model, using very few assumptions, reproduces major statistical properties of seismicity, suggesting that the dissipation of energy in earthquakes follows the pattern similar to turbulent cascades: large-scale tectonic deformation energy is dissipated on creating new fractal surfaces of earthquake faults (see the previous subsection).

Although the exact mechanism of both seismicity and turbulent motion is unknown, the ubiquity of their manifestations makes one question whether this complexity is due to the deep topological properties of space-time itself (see also discussion on computability of the Mandelbrot set and other fractals in subsection 3.3.7). Thus, although its fundamental equations are still unknown, earthquake seismology may be a branch of mathematical physics, like turbulence of fluids.

Comparing seismicity to turbulence indicates that earthquake science may have a profound problem: the Navier–Stokes formula in fluid dynamics is well known, but has not facilitated our understanding of turbulence. Thus, we cannot construct theoretical solutions from the first principle. The best hypotheses which fit the turbulence data are essentially statistical hierarchical (fractal) models. Although dynamical systems and fractal notions have provided new tools with which to describe the turbulence (as well as for seismicity characterization), these tools are still of limited use (LORENZ, 1993; NEWMAN *et al*., 1994; TURCOTTE, 1997).

GLEICK (1987) provides an eloquent and entertaining account of the failure of theoretical physics to explain turbulence. His story regarding Heisenberg (GLEICK, 1987, pp. 121, 329) reflects a general feeling among mathematical physicists that turbulence may be insolvable. The systems which theoretical physics analyzes are isolated and closed, whereas in earth sciences we must deal with systems which are widely open (ORESKES *et al*., 1994). If the analogy of seismicity to turbulence is even partially true, this suggests that prospects of predicting earthquake dynamics and seismicity patterns purely by theoretical insight are low.

5.4. Methodology in Medical Research

We argue above that for earthquake research to succeed, we need more rigorous methods of data analysis. Research in the health sciences could be used as an example of such a methodological transition, since a human organism is complex and many types of experiments are not possible with human subjects. This is similar to the circumstances in the geosciences. Because significant effort has been made to develop scientific methods in medicine, we may benefit from their experience and insight. During the second half of the 20th century, medical researchers realized that without carefully controlled trials, experimental results would be of limited use. SHAPIRO and SHAPIRO (1997, p. 73) even claim that the effect of the medications of a prescientific era was largely psychological: "Despite the problem of bias and the hazards of interpretation, we propose that the available data support the overall hypothesis that the history of medical treatment up to the era of scientific medicine is largely the history of the placebo effect" (however, see also STERNBERG, 1998, who partially disagrees with this view).

Medical researchers found that single-blind experiments (when patients do not know who is getting a medication and who is on a placebo) are insufficient to exclude a bias. The experimenters should not know this either. Such double-blind experiments have been found to be reproducible and since the 1970s became a standard research method. In addition to the double-blind method, SHAPIRO and SHAPIRO (1997, p. 210) propose other general rules for the statistical analysis of medical data. Most of these rules are applicable to analyzing earthquake data as well (cf. our list in section 4): "Well-controlled studies usually include . . . *a priori* specification of hypotheses; independent and dependent variables; specification of statistical procedures to be used; adequate sample size; the use of measures with demonstrated reliability and validity, to control for errors in measurement; assessment of statistical power; clear differentiation between *a priori* hypotheses and *post* *hoc* hypotheses; the use of adjusted probability values based on the number of variables tested, to avoid errors; and subsequent replication of results."

Reviews of statistical trials in medical research (such as NOWAK, 1994; TAUBES, 1998) make for fascinating reading: a seemingly slight change in methodology or data selection may lead to completely different conclusions. As the above publications testify, the development of a new scientific methodology which began around the 1930s (SHAPIRO and SHAPIRO, 1997, p. 153), continues today in the health sciences. Such history testifies that the major source of errors and biases is usually not technically insufficient methods, but neglect of the basic statistical rules for gathering and processing data. We cannot adopt all the critical methods employed in medical and biological research, since many such methods are specific to these disciplines. No experiments are possible with real earthquakes, and the earth is a unique object of study; thus our ability to make controlled tests is even more limited than in medicine. However, we need to develop methods which yield reproducible, objective results.

5.5. Discussion

The need for sophisticated statistical treatment of data emerged as medical research moved from strong, easily observable effects to more subtle ones (SHAPIRO and SHAPIRO, 1997). Similarly, in earthquake seismology short-term clustering is an obvious feature of shallow seismicity, whereas long-term effects (long-term clustering or quasi-periodicity) must be studied in the presence of a much stronger signal. Thus without an efficient and faithful model of short-term clustering, in most cases no practical result can be obtained for long-term effects. Such a model is needed to explain the short-term, time-space-focal mechanism regularities of earthquake sequences.

With the exception of a few problems such as wave propagation and its influence on rupture, no conclusions can be obtained from the fundamental equations, i.e., no deductive conclusion is yet possible to explain real earthquake patterns. Thus the problems in earthquake seismology must be solved inductively, and each proposed seismicity pattern should be rigorously tested for its predictive skill. A starting point might be to require that all claims of successful earthquake predictions outperform the following two "trivial" (null) hypotheses: (1) predictions assuming regular earthquake occurrence, based on the long-term average of past seismicity, and (2) predictions based on extrapolating the recent seismicity record, such as aftershock sequences of strong earthquakes or the earthquake history of recent years. These null hypotheses correspond to so-called "climatological" and "persistence" forecasts in meteorology (MURPHY, 1996). Seismic gap model validation (KAGAN and JACKSON, 1991, 1995) and testing of VAN predictions (KAGAN and JACKSON, 1996) offer examples of such tests applied to seismicity.

Therefore, the following methodology can be proposed for testing of earthquake prediction schemes (KAGAN, 1997b):

1. Case history investigations: these studies should satisfy the criteria formulated, for example, by the IASPEI group (WYSS, 1991; WYSS and DMOWSKA, 1997, pp. 13–15; WYSS and BOOTH, 1997). However, case histories of "successful prediction" of one or several earthquakes do not demonstrate a method has predictive skill, since such success may be due to chance or a selection bias. Since seismicity is characterized by extreme randomness, it is possible in principle to select almost any pattern from abundant data. For example, the lack of agreement over whether the 1989 Loma-Prieta earthquake was predicted meaningfully (U.S. GEOLOGICAL SURVEY STAFF, 1990; HARRIS, 1998) demonstrates (1) a forecast must be formally and rigorously specified in advance, and (2) a track record of a prediction method is needed—no reliable conclusion can be reached on the basis of one or even several events.

2. Rigorous large-scale retrospective statistical testing, using a control sample, i.e., the data that were not considered in formulating the working hypothesis and evaluating adjustable parameters for the model. The null hypothesis should be formulated and tested against the same data. However, tectonic and geological conditions always vary and the long-term, long-range clustering property of seismicity makes it difficult, if not impossible, to use an earthquake catalog in different time intervals or in different regions as a control sample. As experience in medical research (section 5.4) suggests, even meticulous efforts to control the bias often fail. Test of the $M8$ algorithm (KOSSOBOKOV *et al.*, 1997, p. 228; 1999) again demonstrates that the results of the *a posteriori* forecast could be significantly better than those of the real-time prediction.

3. Forward, prospective prediction testing, during which no adjustment of parameters is allowed and all relevant possible ambiguities in the data or the interpretation technique are specified in advance (EVISON and RHOADES, 1993, 1997; KOSSOBOKOV *et al.*, 1997). Such testing should be carried out for empirical and theoretical predictions of earthquake occurrence. A track record of the predictions should be established to verify the method's predictive power. Only the results of properly executed forward testing are to be accepted as a final verdict of a method performance.

Because of practical and theoretical considerations, many such tests involve the occurrence of large earthquakes. To carry out these tests in a reasonably short time, the test space should be extended to large seismic regions such as circum-Pacific (see section 3.1.2), or global seismicity. For smaller earthquakes, seismicity in one region is usually dominated by a few major earthquakes. Thus we cannot separate the effects of random fluctuations in effectively small samples from possible regularities of earthquake occurrence. This would necessitate that even for small events, prediction tests must be carried out in several different regions.

5.6. Conclusions

The above discussion suggests the insufficiency of two commonly proposed solutions for the crisis in earthquake seismology: (1) the holistic, interdisciplinary integration of geophysical data and the search for new "emergent" regularities, or (2) the application of earthquake physics methods. These methods may be necessary to make progress in the understanding of earthquakes, however they will remain unsuccessful if their performance is not properly validated. Without critical testing, the holistic approach erodes into descriptive phenomenology, its constructs deprived of predictive power. Similarly, without confronting real data and controlled verification of its performance, theoretical or computer modeling breaks down into speculation and computer manipulation.

The "fifth force" controversy (FRANKLIN, 1993) demonstrates the difference in attitudes between geophysicists and theoretical physicists in establishing scientific truth. Whereas the latter have a critical and rigorous approach to formulating and testing new hypotheses, and enforce a strong conformance to these norms, the earth science community lacks effective rules to verify a hypothesis. A method to enforce discipline in formulating and justifying hypotheses in geophysics needs to be found and discussed. Higher standards for research in earthquake seismology must be enforced. Authors should adopt a more rigorous style of scientific investigation, and reviewers and editors of geophysical journals should reject manuscripts which do not satisfy the above requirements.

Summarizing the discussion, we conclude that the only quantitative, reproducible knowledge pertinent to earthquake occurrence available currently is statistical. The regularities, discovered by OMORI (1984), GUTENBERG and RICHTER (1944) and others (see more in KAGAN, 1994), have withstood the test of time and are confirmed, albeit qualitatively, by recent developments in nonlinear dynamics and critical phenomena. The advances in statistical analysis of seismological data, and new understanding of the scaling properties of seismicity, including possible universality of major properties of earthquake occurrence (KAGAN, 1994, 1997b; MAIN, 1995, 1996), provide a unique opportunity to evaluate seismic hazard and to estimate the short- and long-term rate of future earthquake occurrence, i.e., to predict earthquakes statistically.

Stress accumulation and release models may yield new understanding of the earthquake processes and eventually allow us to predict earthquakes more reliably. The advances in space geodetic measurements—Global Positioning System (GPS), Synthetic Aperture Radar Interferometry (InSAR)—should greatly increase our understanding of the earthquake process (JACKSON *et al.*, 1997; MEADE and SANDWELL, 1996). New powerful observational tools and interpretive techniques allow us to obtain time-space distribution of moment release for large earthquakes

(Ruff and Miller, 1994; Tanioka and Gonzalez, 1998; Schwartz, 1999—see additional discussion in subsections 3.1.1 and 3.3.4). Seismic moment tensor solutions for small ($m \geq 3.5$) and moderate size earthquakes (Zhu and Helmberger, 1996, and references therein) are becoming routinely available. These and other new methods and techniques may soon change a perspective in earthquake research.

Finally, I would like to answer the question in my title: "Is earthquake seismology a hard, quantitative science?" It is clear from the above discussion, that the answer is "not yet," although with application of more rigorous methods for proposing and validating scientific hypotheses, we may reach this goal.

Acknowledgments

I appreciate partial support from the Southern California Earthquake Center (SCEC). SCEC is funded by NSF Cooperative Agreement EAR-8920136 and USGS Cooperative Agreements 14-08-0001-A0899 and 1434-HQ-97AG01718. This work has greatly benefited from stimulating discussions with D. D. Jackson and D. Sornette of UCLA, Y. Ben-Zion of USC, R. Geller of Tokyo University, F. Mulargia of Bologna University, P. B. Stark of UC Berkeley, M. Wyss of Univ. of Alaska, I. Main of Univ. of Edinburgh, M. Ohtake of Tohoku Univ., V. I. Keilis-Borok and V. G. Kossobokov of Russian Academy of Science. I am grateful to three anonymous reviewers and to the editor M. Wyss for their very useful remarks. The SCEC contribution number is 454.

Note Added in Proof

The recent *Nature* WEB debate on earthquake prediction (see http://helix.nature.com/debates/earthquake/equake_frameset.html) confirmed several of the ideas and conclusions discussed in this paper. Scarcely anyone proposed that deterministic type reliable prediction of individual earthquakes is feasible in the near future. It is especially interesting to note the opinions of the debators on seismic gap and earthquake recurrence models. Scholz (week 6 contribution) in effect claims that none of the results based on the seismic gap hypothesis are ready to be tested. This statement is surprising: according to the GEOREF database more than 375 publications, many of them published in the last few years, cite as their subject the *seismic gap model*. If this model, despite its claimed great potential value for evaluation of the intermediate- and long-term seismic hazard, cannot be tested after nearly 30 years of development, general scientific methods in earthquake seismology need to be significantly revised.

REFERENCES

ADLER, A., FELDMAN, R., and Taqqu, M. S. (eds), *A Practical Guide to Heavytails: Statistical Techniques for Analyzing Heavy Tailed Distributions* (Birkhäuser, Boston 1998).

ANDERSON, P. W. (1972), *More is Different*, Science *177*, 393–396.

BAKUN, W. H., and LINDH, A. G. (1985), *The Parkfield, California, Earthquake Prediction Experiment*, Science *229*, 619–24.

BEN-MENAHEM, A. (1995), *A Concise History of Mainstream Seismology—Origins, Legacy, and Perspectives*, Bull. Seismol. Soc. Am. *85*, 1202–1225.

BEN-ZION, Y., SAMMIS, C., and HENYEY, T. (1999), *Perspectives on the Field of Physics of Earthquakes*, Seismol. Res. Lett., in press.

BLUM, L., CUCKER, F., Shub, M., and Smale, S., *Complexity and Real Computation* (Springer, New York 1998).

CHELIDZE, T. (1993), *Fractal Damage Mechanics of Geomaterials*, Terra Nova *5*, 421–437.

COBURN, A. W., and SPENCE, R., *Earthquake Protection* (Wiley, New York 1992).

DENG, J. S., and SYKES, L. R. (1997), *Stress Evolution in Southern California and Triggering of Moderate-, Small-, and Micro-size Earthquakes*, J. Geophys. Res. *102*, 24,411–24,435.

DIETERICH, J. (1994), *A Constitutive Law for Rate of Earthquake Production and its Application to Earthquake Clustering*, J. Geophys. Res. *99*, 2601–2618.

ENGELHARDT, VON W., and ZIMMERMAN, J., *Theory of Earth Science* (Cambridge University Press, Cambridge 1988).

EVANS, R. (1997), *Assessment of Schemes for Earthquake Prediction: Editor's Introduction*, Geophys. J. Int. *131*, 413–420.

EVISON, F. F., and RHOADES, D. A. (1993), *The Precursory Earthquake Swarm in New Zealan*d: *Hypothesis Tests*, New Zealand J. Geophys. *36*, 51–60 (Correction p. 267).

EVISON, F. F., and RHOADES, D. A. (1997), *The Precursory Earthquake Swarm in New Zealand: Hypothesis Tests II*, New Zealand J. Geol. Geophys. *40*, 537–547.

FEDOTOV, S. A. (1968), *On the Seismic Cycle, Feasibility of Quantitative Seismic Zoning and Long-term Seismic Prediction*. In *Seismic Zoning of the USSR*, pp. 121–150, Nauka, Moscow (in Russian); English translation: Israel program for scientific translations, Jerusalem, 1976.

FRANKLIN, A., *The Rise and Fall of the "Fifth Force": Discovery, Pursuit, and Justification in Modern Physics* (American Inst. Physics, New York 1993).

GABRIELOV, A. M., and NEWMAN, W. I., *Seismicity modeling and earthquake prediction: A review*. In *Nonlinear Dynamics and Predictability of Geophysical Phenomena* (Newman, W. I., Gabrielov, A., and Turcotte, D. L., eds.), Geoph. Monogr. *83* (Washington, American Geophysical Union 1994) pp. 7–13.

GELLER, R. J. (ed.) (1996), *Debate on "VAN*," Geophys. Res. Lett. *23*(11), 1291–1452.

GELLER, R. J. (1997), *Earthquake Prediction: A Critical Review*, Geophys. J. Int. *131*, 425–450.

GELLER, R. J., JACKSON, D. D., KAGAN, Y. Y., and MULARGIA, F. (1997), *Earthquakes Cannot be Predicted*, Science *275*, 1616–1617.

GLEICK, J., *Chaos, Making a New Science* (Viking, New York 1987).

GRAHAM, R. L., and SPENCER, J. H. (1990), *Ramsey Theory*, Scientific American *263*(7), 112–117.

GUSEV, A. A. (1998), *Earthquake Precursors: Banished Forever?* EOS Trans. AGU *79*, 71–72.

GUTENBERG, B., and RICHTER, C. F. (1944), *Frequency of Earthquakes in California*, Bull. Seismol. Soc. Am. *34*, 185–188.

HARDEBECK, J. L., NAZARETH, J. J., and HAUKSSON, E. (1998), *The Static Stress Change Triggering Model: Constraints from two Southern California Aftershock Sequences*, J. Geophys. Res. *103*, 24,427–24,437.

HARRIS, R. A. (1998), *Forecasts of the Loma-Prieta, California, Earthquake*, Bull. Seismol. Soc. Am. *88*, 898–916.

ISHIBASHI, K., *Specification of a soon-to-occur seismic faulting in the Tokai District, Central Japan, based upon seismotectonics*. In *Earthquake Prediction: An International Review* (Simpson, D. W. and Richards, P. G., eds.), Ewing Monograph Series, *4*, pp. 297–332 (Am. Geophys. Union, Washington 1981).

Jackson, D. D., and Kagan, Y. Y. (1993), *Reply [to Nishenko and Sykes]*, J. Geophys. Res. *98*, 9917–9920.

Jackson, D. D., Shen, Z.-K., Potter, D., Ge, B.-X., and Sung, L.-Y. (1997), *Southern California Deformation*, Science *277*, 1621–1622.

Jordan, T. H. (1997), *Is the Study of Earthquakes a Basic Science?* Seismol. Res. Lett. *68*(2), 259–261.

Kagan, Y. Y. (1973), *A Probabilistic Description of the Seismic Regime*, Izv. Acad. Sci. USSR, Phys. Solid Earth, 213–219 (English translation).

Kagan, Y. Y. (1982), *Stochastic Model of Earthquake Fault Geometry*, Geophys. J. Roy. Astr. Soc. *71*, 659–691.

Kagan, Y. Y. (1992), *Seismicity: Turbulence of Solids*, Nonlinear Sci. Today *2*, 1–13.

Kagan, Y. Y. (1994), *Observational Evidence for Earthquakes as a Nonlinear Dynamic Process*, Physica D *77*, 160–192.

Kagan, Y. Y. (1997a), *Statistical Aspects of Parkfield Earthquake Sequence and Parkfield Prediction Experiment*, Tectonophysics *270*, 207–219.

Kagan, Y. Y. (1997b), *Are Earthquakes Predictable?* Geophys. J. Int. *131*, 505–525.

Kagan, Y. Y., and Jackson, D. D. (1991), *Seismic Gap Hypothesis: Ten Years After*, J. Geophys. Res. *96*, 21,419–21,431.

Kagan, Y. Y., and Jackson, D. D. (1994), *Long-term Probabilistic Forecasting of Earthquakes*, J. Geophys. Res. *99*, 13,685–13,700.

Kagan, Y. Y., and Jackson, D. D. (1995), *New Seismic Gap Hypothesis: Five Years After*, J. Geophys. Res. *100*, 3943–3959.

Kagan, Y. Y., and Jackson, D. D. (1996), *Statistical Tests of VAN Earthquake Predictions: Comments and Reflections*, Geophys. Res. Lett. *23*, 1433–1436.

Kagan, Y. Y., and Knopoff, L. (1987), *Statistical Short-term Earthquake Prediction*, Science *236*, 1563–1567.

Kalia, R. K., Nakano, A., Omeltchenko, A., Tsuruta, K., and Vashishta, P. (1997), *Role of Ultrafine Microstructures in Dynamic Fracture in Nanophase Silicon Nitride*, Phys. Rev. Lett. *78*, 2144–2147.

King, G. C. P., Stein, R. S., and Lin, J. (1994), *Static Stress Changes and the Triggering of Earthquakes*, Bull. Seismol. Soc. Am. *84*, 935–953.

Kirchherr, W., Li, M., and Vitanyi, P. (1997), *The Miraculous Universal Distribution*, Math. Intelligencer *19*, 7–15.

Knopoff, L., Aki, K., Allen, C. R., Rice, J. R., and 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., and Dewey, J. W. (1997), *Testing an Earthquake Prediction Algorithm*, Pure appl. geophys. *149*, 219–232.

Kossobokov, V. G., Healy, J. H., and Dewey, J. W. (1998), *Testing an Earthquake Prediction Algorithm: M 7.5+Earthquakes in Circum-Pacific*, *1992–1997*, XXVI General Assembly European Seismol. Commis., Abstracts, Tel Aviv, Israel, p. 42.

Kossobokov, V. G., Romashkova, L. L., Keilis-Borok, V. I., and Healy, J. H. (1999), *Testing Earthquake Prediction Algorithms*: *Statistically Significant Advance Prediction of the Largest Earthquakes in the Circum-Pacific*, *1992–1997*, Phys. Earth Planet. Inter. *111*, 187–196.

Kuhn, T. S., *The Structure of Scientific Revolutions*, 2nd ed. (Univ. Chicago Press, Chicago 1970).

Leary, P. C. (1997), *Rock as a Critical-point System and the Inherent Implausibility of Reliable Earthquake Prediction*, Geophys. J. Int. *131*, 451–466.

Li, M., and Vitanyi, P., *An Introduction to Kolmogorov Complexity and its Applications*, 2nd ed. (Springer, New York 1997).

Lorenz, E. N., *The Essence of Chaos* (Univ. Washington Press, Seattle 1993).

Main, I. G. (1995), *Earthquakes as Critical Phenomena—Implications for Probabilistic Seismic Hazard Analysis*, Bull. Seismol. Soc. Am. *85*, 1299–1308.

Main, I. G. (1996), *Statistical Physics*, *Seismogenesis*, *and Seismic Hazard*, Rev. Geophys. *34*, 433–462.

Mandelbrot, B. B., *The Fractal Geometry of Nature*, 2nd ed. (W. H. Freeman, San Francisco 1983).

Marder, M. (1998), *Computational Science—Unlocking Dislocation Secrets*, Nature *391*, 637–638.

Margenau, H., *The Nature of Physical Reality: A Philosophy of Modern Physics* (McGraw-Hill, New York 1950).

McCann, W. R., Nishenko, S. P., Sykes, L. R., and Krause, J. (1979), *Seismic Gaps and Plate Tectonics: Seismic Potential for Major Boundaries*, Pure appl. geophys. *117*, 1082–1147.

Meade, C., and Sandwell, D. T. (1996), *Synthetic Aperture Radar for Geodesy*, Science *273*, 1181–1182.

Michael, A. J., and Jones, L. M. (1998), *Seismicity Alert Probabilities at Parkfield, California, Revisited*, Bull. Seismol. Soc. Am. *88*, 117–130.

Mogi, K. (1995), *Earthquake Prediction Research in Japan*, J. Phys. Earth *43*, 533–561.

Molchan, G. M. (1997), *Turbulent Cascades: Limitations and a Statistical Test of the Lognormal Hypothesis*, Physics of Fluids *9*, 2387–2396.

Mulargia, F. (1997), *Retrospective Validation of Time Association*, Geophys. J. Int. *131*, 500–504.

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.

Nadeau, R. M., and McEvilly, T. V. (1997), *Seismological Studies at Parkfield. V. Characteristic Microearthquake Sequences as Fault-zone Drilling Targets*, Bull. Seismol. Soc. Am. *87*, 1463–1472.

Nadeau, R. M., and Johnson, L. R. (1998), *Seismological Studies at Parkfield VI: Moment Release Rates and Estimates of Source Parameters for Small Repeating Earthquakes*, Bull. Seismol. Soc. Am. *88*, 790–814.

Newman, W. I., Gabrielov, A., and Turcotte, D. L., eds., *Nonlinear Dynamics and Predictability of, Geophysical Phenomena*, Geoph. Monogr. *83* (American Geophys. Union, Washington 1994).

Nishenko, S. P. (1991), *Circum-Pacific Seismic Potential: 1989–1999*, Pure appl. geophys. *135*, 169–259.

Nishenko, S. P., and Sykes, L. R. (1993), *Comment on "Seismic Gap Hypothesis: Ten Years After" by Y. Y. Kagan and D. D. Jackson*, J. Geophys. Res. *98*, 9909–9916.

Nowak, R. (1994), *Problems in Clinical Trials go far Beyond Misconduct*, Science *264*, 1538–1541.

Omori, F. (1894), *On the After-shocks of Earthquakes*, J. College Sci., Imp. Univ. Tokyo *7*, 111–200 (with Plates IV–XIX).

Oreskes, N., Shrader-Frechette, K., and Belitz, K. (1994), *Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences*, Science *263*, 641–646, also see Science *264*, 331.

Ouillon, G., Castaing, C., and Sornette, D. (1996), *Hierarchical Scaling of Faulting*, J. Geophys. Res. *101*, 5477–5487.

Penrose, R., *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics* (Oxford Univ. Press, New York 1989).

Popper, K. R., *Logic of Scientific Discovery*, 2nd ed. (Hutchinson, London 1980).

Reasenberg, P. A., and Jones, L. M. (1989), *Earthquake Hazard After a Mainshock in California*, Science *243*, 1173–1176.

Reid, H. F. (1910), *The California Earthquake of April 18, 1906, Vol. 2: The Mechanics of the Earthquake*, Carnegie Institution of Washington, Washington, D.C.

Roeloffs, E., and Langbein, J. (1994), *The Earthquake Prediction Experiment at Parkfield, California*, Rev. Geophys. *32*, 315–336.

Ruff, L. J., and Miller, A. D. (1994), *Rupture Process of Large Earthquakes in the Northern Mexico Subduction Zone*, Pure appl. geophys. *142*, 101–172 (see also WEB site http://www.geo.lsa.umich.edu/SeismoObs/STF.html).

Samorodnitsky, G., and Taqqu, M. S., *Stable non-Gaussian Random Processes: Stochastic Models with Infinite Variance* (Chapman and Hall, New York 1994).

Santillana de, G., *The Origins of Scientific Thought; from Anaximander to Proclus, 600 B.C. to 500 A.D.* (Univ. Chicago Press, Chicago 1961).

Scholz, C. H., *The Mechanics of Earthquakes and Faulting* (Cambridge Univ. Press, Cambridge 1990).

Scholz, C. H. (1996), *Faults Without Friction?* Nature *381*, 556–557.

Scholz, C. (1997), *Whatever Happened to Earthquake Prediction?* Geotimes *42*(3), 16–19.

Scholz, C. H. (1998), *Earthquakes and Friction Laws*, Nature *391*, 37–42.

Scholz, C. H., Sykes, L. R., and Aggarwal, Y. P. (1973), *Earthquake Prediction: A Physical Basis*, Science *181*, 803–810.

Schwartz, D. P., and Coppersmith, K. J. (1984), *Fault Behavior and Characteristic Earthquakes: Examples from Wasatch and San Andreas Fault Zones*, J. Geophys. Res. *89*, 5681–5698.

SCHWARTZ, S. Y. (1999), *Non-characteristic Behavior and Complex Recurrence of Large Subduction Zone Earthquakes*, J. Geophys. Res., in press.

SHAPIRO, A. K., and SHAPIRO, E., *The Powerful Placebo: From Ancient Priest to Modern Physician* (J. Hopkins Univ. Press, Baltimore 1997).

SHARON, E., and FINEBERG, J. (1996), *Microbranching Instability and the Dynamic Fracture of Brittle Materials*, Physical Review B *54*, 7128–7139.

SHARON, E., and FINEBERG, J. (1999), *Confirming the Continuum Theory of Dynamic Brittle Fracture for Fast Cracks*, Nature *397*, 333–335.

SNIEDER, R., and VAN ECK, T. (1997), *Earthquake Prediction: a Political Problem?* Geologische Rundschau *86*, 446–463.

STARK, P. B. (1997), *Earthquake Prediction: The Null Hypothesis*, Geophys. J. Int. *131*, 495–499.

STERNBERG, E. M. (1998), *Unexpected Effects*, Science *280*, 1901–1902.

SYKES, L. R. (1971), *Aftershock Zones of Great Earthquakes, Seismicity Gaps, and Earthquake Prediction for Alaska and the Aleutians*, J. Geophys. Res. *76*, 8021–8041.

TAUBES, G. (1998), *The (Political) Science of Salt*, Science *281*, 898–907, see also Science *281*, 1961–1963.

TANIOKA, Y., and GONZALEZ, F. I. (1998), *The Aleutian Earthquake of June 10, 1996 (M_w 7.9) Ruptured Parts of both the Andreanof and Delarof Segments*, Geophys. Res. Lett. *25*, 2245–2248.

TURCOTTE, D. L., *Fractals and Chaos in Geology and Geophysics*, 2nd ed. (Cambridge Univ. Press, Cambridge 1997).

U.S. GEOLOGICAL SURVEY STAFF (1990), *The Loma-Prieta, California, Earthquake—An Anticipated Event*, Science *247*, 286–293.

WELDON, R. J. (1996), *Why We Cannot Predict Earthquakes with Paleoseismology, and What We (Paleoseismologists) Really Should Be Doing*, EOS Trans. AGU *77*(46), Fall AGU Meet. Suppl., p. F53.

WYSS, M., ed., *Evaluation of Proposed Earthquake Precursors* (American Geophys. Union, Washington 1991).

WYSS, M., and BOOTH, D. C. (1997), *The IASPEI Procedure for the Evaluation of Earthquake Precursors*, Geophys. J. Int. *131*, 423–424.

WYSS, M., ACEVES, R. L., and PARK, S. K., GELLER, R. J., JACKSON, D. D., KAGAN, Y. Y., and MULARGIA, F. (1997), *Cannot Earthquakes be Predicted? (Technical comments)*, Science *278*, 487–490.

WYSS, M., and DMOWSKA, R., eds. (1997), *Earthquake Prediction—State of the Art*, Pure appl. geophys. *149*, 1–264.

XU, X.-P., NEEDLEMAN, A., and ABRAHAM, F. F. (1997), *Effect of Inhomogeneities on Dynamic Crack Growth in an Elastic Solid*, Modelling and Simulation in Materials Science and Engineering *5*, 489–516.

YEATS, R. S., SIEH, K., and ALLEN, C. R., *The Geology of Earthquakes* (Oxford Univ. Press, New York 1997).

ZHU, L. P., and HELMBERGER, D. V. (1996), *Advancement in Source Estimation Techniques Using Broadband Regional Seismograms*, Bull. Seismol. Soc. Am. *86*, 1634–1641.

(Received August 19, 1998, revised January 18, 1999, accepted January 22, 1999)