Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Tectonophysics 413 (2006) 25–31

TECTONOPHYSICS

www.elsevier.com/locate/tecto

Testing earthquake prediction methods: «The West Pacific short-term forecast of earthquakes with magnitude MwHRV ≥ 5.8»

V.G. Kossobokov *

International Institute of Earthquake Prediction Theory and Mathematical Geophysics, Russian Academy of Sciences,
79-2 Warshavskoye Shosse, Moscow 117556, Russia
Institute de Physique du Globe de Paris, 4 Place Jussieu, 75252 Paris, Cedex 05, France

Received 25 February 2005; received in revised form 10 May 2005; accepted 4 October 2005
Available online 28 November 2005

Abstract

Analyzing the tables and probability maps posted by Yan Y. Kagan and David D. Jackson in April 2002–September 2004 at http://scec.ess.ucla.edu/~ykagan/predictions_index.html and the catalog of earthquakes for the same period, the conclusion is drawn that the underlying method could be used for prediction of aftershocks, while it does not outscore random guessing when main shocks are considered.

Keywords: Probabilistic forecasting; Earthquake prediction; Null-hypothesis; Confidence level; Prediction efficiency

1. Introduction

The mature wisdom of any science is determined by its ability to predict phenomena under study. The abruptness along with apparent irregularity and infrequency of earthquake occurrences facilitate formation of a common perception that earthquakes are in fact random unpredictable phenomena. The earthquake prediction controversy got fertilized from numerous discussions and debates (e.g. Geller, 1997; Wyss, 1997a; Nature Debates, 1999; Cyranoski, 2004), supported, on one side, with a surprisingly small number of basic systematic studies and, on the other side, with a multitude of post-the-fact numerological exercises. As a result, although hundreds, if not thousands of observed phenomena have been claimed "to precede systematically" large earthquakes, there are almost no reproducible quantitative definitions of "precursors". The IASPEI Call for precursor nominees came out with 31 candidates (Wyss, 1991), none of which was found to fully satisfy its Guidelines, mainly due to the eventual inability of authors to provide a precise definition of the observed phenomenon. The situation practically did not change in the second round of the initiative (Wyss, 1997b): only five out of the forty candidates submitted, seemed to deserve further study as possibly related to earthquake prediction. None of them could be considered yet as a validated precursor. Hopefully, "A seismic shift in thinking" (Cyranoski, 2004) towards basic science will result a renaissance of strict definitions and systematic experiments in the field of earthquake prediction, which just a few years ago seemed to dissolve in whispering and condemning "the p-word". This article intends to remind the basics, which were developed eventually and used persistently

---

* International Institute of Earthquake Prediction Theory and Mathematical Geophysics, Russian Academy of Sciences, 79-2 Warshavskoye Shosse, Moscow 117556, Russia.

E-mail addresses: volodya@mitp.ru, volodya@ipgp.jussieu.fr.

0040-1951/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.tecto.2005.10.006

by Vladimir I. Keilis-Borok and his group (Keilis-Borok and Soloviev, 2003) since pioneering “Seismology and logics” (Keilis-Borok, 1964) and “One regularity in the occurrence of strong earthquakes” (Keilis-Borok and Malinovskaya, 1964), in application to independent analysis of «The West Pacific short-term forecast of earthquakes with magnitude MwHRV ≥ 5.8» by Kagan and Jackson (2000) (http://scec.ess.ucla.edu/~ykagan/predictions_index.html).

2. Definitions

No scientific prediction is possible without exact definition of the anticipated phenomenon and the rules, which define clearly in advance of it whether the prediction is confirmed or not. The problem of earthquake prediction allows several definitions that are loosely used by different authors. Among those one can distinguish deterministic statements on the future occurrence (Keilis-Borok, 1982; Kossobokov and Shebalin, 2003) and estimations of conditional intensity, i.e., estimates of conditional probability of the forthcoming earthquake (Jackson and Kagan, 1999). Perhaps, the most simple consensus formulation of the first one was worked out in 1976 by the distinguished Panel on Earthquake Prediction (Allen et al., 1976). According to it,

> An earthquake prediction must specify the expected magnitude range, the geographical area within which it will occur, and the time interval within which it will happen with sufficient precision so that the ultimate success or failure of the prediction can readily be judged. Only by careful recording and analysis of failures as well as successes can the eventual success of the total effort be evaluated and future directions charted. Moreover, scientists should also assign a confidence level to each prediction.

Jackson and Kagan (1999) point out that their “testable earthquake forecasts” give no predictions because no specific area is defined within the active seismic belts on the west of Pacific Ocean. Nevertheless, the theory of earthquake prediction strategies (Molchan and Kagan, 1992; Molchan, 1997, 2003) provides the unique optimal rules for using probabilistic forecasting to define deterministic predictions of earthquakes in case the loss function is defined. Specifically, given the daily updated tables and maps of “the short-term probability p” and “the long-term probability P”, one should define the two predictions arising from setting either a threshold probability p_0 or a threshold probability ratio $(p/P)_0$: in either case, at a given time the locations with the currently observed value equal or above the threshold form the alarm area of prediction. The theory expects p/P being more effective than p in course such a forecast-induced deterministic prediction (Molchan, 2003).

The effectiveness and reliability of an earthquake prediction method could be attributed from the comparison with the null-hypothesis that presumes random distribution of seismic events. The dual levels of statistical significance and confidence of the prediction results equal $1-B(n-1,\ N,\ \mu)$ and $B(n-1,\ N,\ \mu)$, correspondingly, where B is the cumulative binomial distribution function, N counts the total number of target earthquakes that happened, n counts how many of them are predicted, and μ is the space–time volume of alarm in percentage to the total considered in the course of the test.

Specifically, let T and S be the total time and territory considered; $\delta(t,\ s)$ equals 1, if a target earthquake happen at time t and location s, or 0, otherwise; A_t—the territory covered by the alarms at time t; $\tau\times\mu$—the direct product measure on $T\times S$ (we reserve the general case of a potentially time–space dependent measure for future more sophisticated null-hypotheses like “the short-term probability p”); $N=\int_{T\times S}\delta(t,s)$—the total number of target earthquakes within $T\times S$ and $n=\int_A\delta(t,s)$—the number of target earthquakes within $A=\bigcup_T A_t$. The time–space occupied by alarms in percentage to the total space–time considered equals $\mu=\int_A \mathrm{d}(\tau\times\mu)/\int_{T\times S}\mathrm{d}(\tau\times\mu)$.

Due to evidently confined location of earthquake sources in seismically active areas (i.e., belts, fault zones, cracks, etc.) the literal estimation of μ from measuring of prediction area in square km is inadequate, since it equalizes the areas of high and low seismic activity and, at the extreme, areas where earthquake happen and do not happen. It is of particular importance, when claiming prediction of time. Therefore, for earthquake prediction purposes it is natural to assume τ being the uniform measure of time, which corresponds to the Poisson random recurrence of earthquakes, while the measure μ being proportional to spatial density of epicenters, e.g., the empirical spatial frequency of earthquakes. To avoid arbitrariness of the assumptions of smoothing methods we compute the measure μ of an area being proportional to the number of epicenters of earthquakes from a representative sample catalog. As it was demonstrated earlier (Kossobokov et al., 2000), this spatial measure by itself is useful in estimation of seismic hazard. Thus, the natural null-hypothesis measure that we use (Kossobokov and Shebalin, 2003) combines random recurrence in time with the original observed heterogeneity of the empirical spatial distribution of earthquakes.

The null-hypothesis we use here has a nice analogy, so-called "seismic roulette", which might be helpful for general understanding by readers of different background:

> Consider a roulette wheel with as many sectors as the number of events in "the best sample" catalog, a sector for each event. Make your bet according to prediction—(1) determine, which events are inside area of alarm, and (2) put one chip in each of the corresponding sectors. Nature turns the wheel.

If seismic roulette is not perfect and you are smart enough, win systematically and outscore random guessing.

Although some tests require demonstrating more restricted criteria, in most applications the confidence levels of 95% and 99% are used to denote "significant" and "very significant" deviation from random guessing based on the null-hypothesis.

Besides the statistical significance, the most important characteristic of a prediction algorithm is its effectiveness characterized by the error diagram (Molchan, 2003). The ν versus μ diagram, where $\nu=(N-n)/N$ is the percentage of failures-to-predict and μ is the percentage of alarm, demonstrates how far from a random guessing are predictions that result from the algorithm. For example, the performance of random guessing is associated with the diagonal that connects "optimist's" $\{0, 1\}$ and "pessimist's" $\{1, 0\}$ strategies, while "perfect knowledge" reference $\{0, 0\}$ indicates the aim of the ideal success. In the figures below the diagonal of "random guessing" is outlined from below with the curves Γ^{α} associated with the confidence levels $(1-\alpha)=95\%$ and 99%. Specifically, in each case $\Gamma^{\alpha}=(\mu,\ \nu^{\alpha}(\mu))$ where $\nu^{\alpha}(\mu)=1-\min\{n|B(n-1,\ N,\ \mu)>1-\alpha\}/N$; $\alpha=5\%$ or 1%.

The lower envelope Γ of the error point set $\{\nu(\xi),\ \mu(\xi)\}$, produced by different choices of adjustable parameters ξ, is the basic signature of the prediction algorithm, which permits to apply different prediction strategies. For example, if for some ξ, $\nu(\xi)+\mu(\xi)>1$, one can switch to a more efficient "opposite prediction algorithm" (i.e., to declare an alarm when the algorithm defines no alarm and vice versa). Given a loss function $\gamma(\nu,\ \mu)$, the theory permits estimation of the optimal strategy. Specifically, the point where the γ-isoline and Γ touch one another determines both the minimum achievable loss and the optimal set of adjustable parameters in the prediction algorithm. More details on earthquake prediction strategies and their applications are presented in Molchan (2003).

3. Data

Jackson and Kagan (1999) consider the territory of West Pacific short-term forecast being coarse-grained into cells, 0.5° by 0.5° each. Every day using the Harvard CMT data available they compute separately, in the real-time forecasting mode, the two tables of parameters characterizing North-Western (14641 nodes between 0–60°N and 110–170°E) and South-Western (19481 nodes between 0–60°S and 110°E–170°W) regions. (Each of the 121 nodes on the intersection appears twice and may have different characterization in NW and SW tables.) We consider the last two columns of these tables denoted "the short-term probability p" of earthquakes with magnitude MwHRV ≥ 5.8 and "probability ratio p/P". The archive of the tables computed in the real-time forecasting mode for the period from April 10, 2002 made available due to the daily updates kindly E-mailed to me by Yan Y. Kagan.

In the test area of West Pacific short-term forecast experiment and the period from April 10, 2002 to September 13, 2004 the Harvard CMT data reports 218 earthquakes of magnitude MwHRV ≥ 5.8 and hypocenter depth less or equal to 70 km. By definition (Kagan and Jackson, 2000; http://scec.ess.ucla.edu/~ykagan/predictions_index.html), these earthquakes were the target events of the forecast experiment. According to the definition from Keilis-Borok et al. (1980), there are 67 aftershocks and 151 main shocks.

For estimation of μ, the catalog of all earthquakes of magnitude 4 or larger from the NEIC Global Hypocenters Data Base in 1964–1984 was used. In the practice of the on-going test of the M8-MSc algorithms (Kossobokov et al., 1999) this sample catalog proved being an adequate representation of the global distribution of earthquakes (in our experience the estimates obtained by using different global representative seismic catalogs, e.g., the Harvard CMT data set, are naturally very close to each other).

4. Characterizing deterministic predictions

Each target earthquake E defines the threshold values $p(E)$ and $p/P(E)$ being the values of "short-term probability p" and "probability ratio p/P" determined in advance for the day of the earthquake. The threshold $p(E)$ defines the unique area of alarm of the minimal value of μ required for successful prediction of the target earthquake, $\mu_p(E)$, which is the union of all cells, where $p \geq p(E)$. In a similar way, one can define the unique area of alarm associated with deterministic prediction based on p/P statistic and a target earthquake E and the value $\mu_{p/P}(E)$. For example, the largest September 25, 2003, MwHRV = 8.3 earthquake off shore Hokkaido requires $\mu_p(E) = 70.6\%$ and $\mu_{p/P}(E) = 79.1\%$ for its successful prediction, while the October 10, 2002, NW of New Guinea, MwHRV = 7.6 is predicted if the alarm $\mu_p(E) = 1.2\%$ and $\mu_{p/P}(E) = 0.7\%$ of the total space–time volume considered.

It should be noted that the attribute "short-term" in (Jackson and Kagan, 1999; Kagan and Jackson, 2000) appears misleading: it actually mixes up the frequency of the probabilistic forecast updates with a common perception of the duration of alarm. Although the temporal accuracy of the "short-term forecasts" has not been investigated in full detail, it was noted that, even for reasonably small values of μ (i.e., for the high values of thresholds generating them), some places stay in the alerted areas for years.

An archive of the experiment provides all the target earthquakes $\{E\}$ that happened in its course, as well as the two samples of the minimal space–time volumes $\{\mu_p(E)\}$ and $\{\mu_{p/P}(E)\}$ associated with it. Since a given threshold (not necessarily the one associated with a target earthquake) uniquely defines the outcome of the deterministic prediction in course the experiment and the percentage of failures-to-predict v, in particular, one can define the empirical performance curves from $\{\mu_p(E), v_p(E)\}$ and $\{\mu_{p/P}(E), v_{p/P}(E)\}$ by plotting their lower envelopes Γ_p and $\Gamma_{p/P}$. To simplify the figures below we show on them the generating sets $\{\mu_p(E), v_p(E)\}$ and $\{\mu_{p/P}(E), v_{p/P}(E)\}$, but not Γ_p and $\Gamma_{p/P}$.

5. The prediction results

Fig. 1 displays the (μ, v)-diagrams that characterize the efficiency of deterministic predictions of the 218 target earthquakes arising from "the short-term probability p" and "probability ratio p/P" forecasts. It is rather evident that the achieved statistics are much better than random guessing through wide ranges of μ and, accordingly, either of the corresponding threshold values. On the contrary, when the aftershocks are ex-

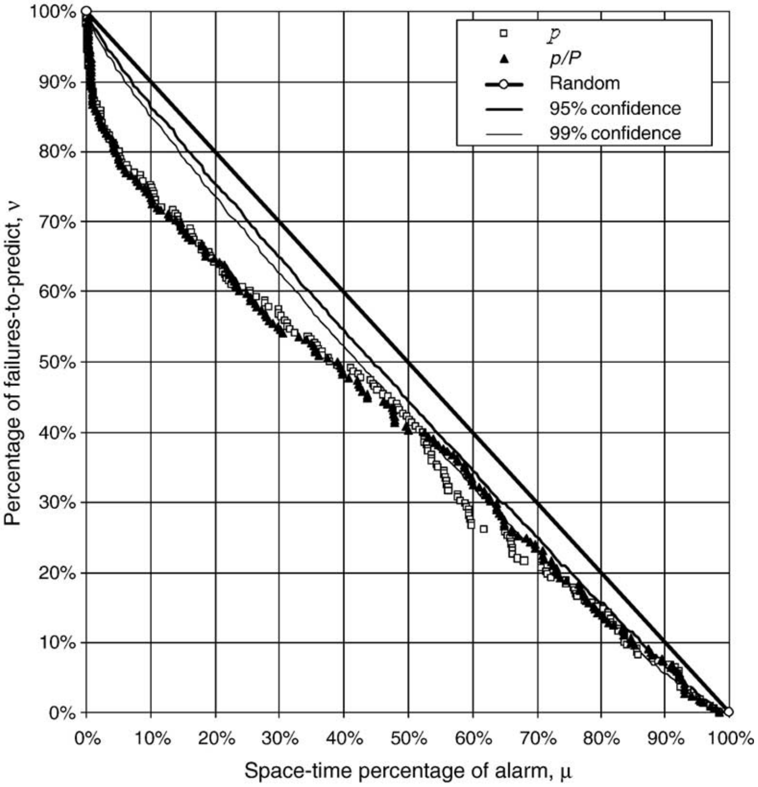

Fig. 1. The percentage of the failures-to-predict v versus the percentage of the alerted space–time volume μ: $\{\mu_p(E), v_p(E)\}$ and $\{\mu_{p/P}(E), v_{p/P}(E)\}$ generated by the 218 earthquakes with magnitude MwHRV ≥ 5.8 and depth ≥ 70 km.

cluded from the set of target earthquakes, the prediction results do not outscore at all the random guessing (Fig. 2).

6. Discussion and conclusion

The reader may notice that the updated statistics of the real-time forecast experiment lead to a slightly different, although expected, conclusions than those reported earlier (Kossobokov, 2004a). It is rather natural, since the theory predicts the confidence reaching any level $(1-\alpha)$ arbitrary close to 1 for any algorithm (or hypothesis) that has even an arbitrary small advantage over the null-hypothesis. In general, it may take a long time. In our case, the interval of μ where the forecast induced predictions outscore random guessing have grown from (0, 20%) after one year to much larger range after another year and a half. As it was pointed out previously and confirmed with newly accumulated data, the advantage essentially related to evident clustering of earthquakes disappears, as soon as the catalog of target earthquakes gets rid of aftershocks.

It should be noted that the choice of specific definition of aftershocks (Keilis-Borok et al., 1980; as well as the choice of specific sample catalog for estimation of μ) was not made for the purposes of this particular study and essentially is not critical for the conclusion: other definitions (e.g., Gardner and Knopoff, 1974) produce (μ, ν)-diagrams very similar to those in Fig. 2. Moreover, in each of these definitions, any earthquake that falls into the space–time aftershock box and has magnitude larger than that of the main shock is treated as another separate main shock. Fig. 2 permits making a suggestion that an aftershock definition could be reformulated into a prediction algorithm (i.e., switching alarms in the currently opened aftershock boxes), which effectiveness compares to the one arising from "the short-term probability p" and "probability ratio p/P" forecasts.

Fig. 3 demonstrates the outcome of a somewhat "absurd post-diction" of the earthquakes in the period from April 10, 1992 to September 13, 1994 predicted by using the probability maps computed ten years later. This one example could be viewed as an illustration that indicates (1) how a trial resulting from a more adequate null-hypothesis with clustered events would be rejected by the independent random guessing null-hypothesis, as well as (2) what random deviations from

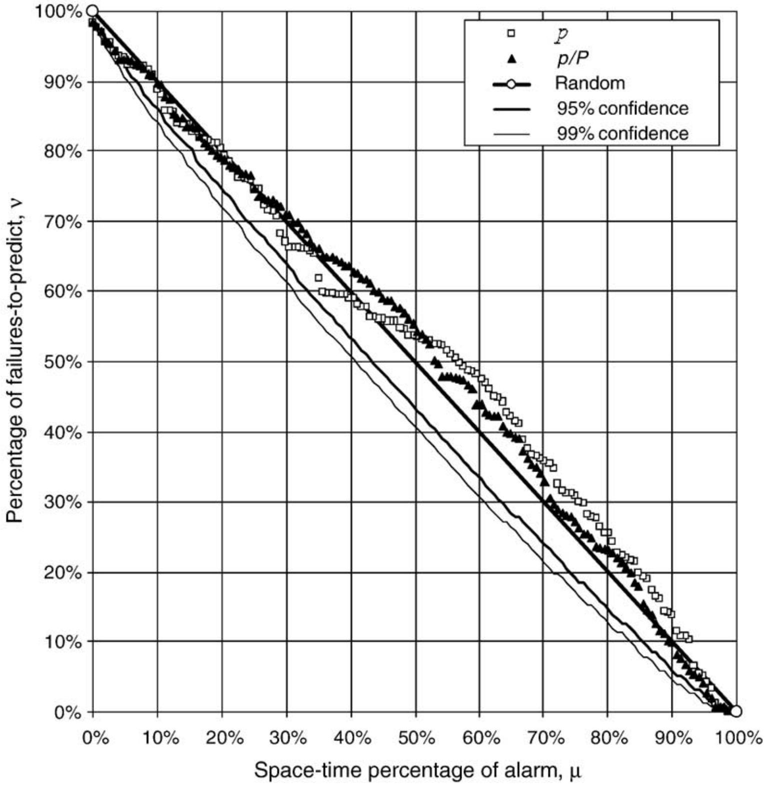

Fig. 2. The percentage of the failures-to-predict ν versus the percentage of the alerted space–time volume μ: $\{\mu_p(E), \nu_p(E)\}$ and $\{\mu_{p/P}(E), \nu_{p/P}(E)\}$ generated by the 151 main shocks with magnitude MwHRV ≥ 5.8 and depth ≥ 70 km.

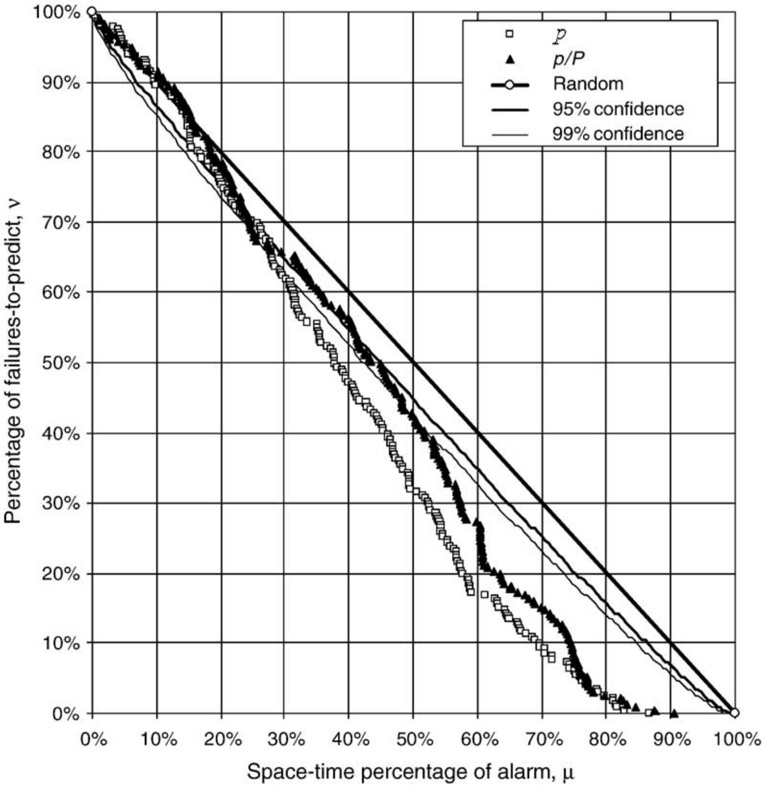

Fig. 3. The percentage of the failures-to-predict μ versus the percentage of the alerted space–time volume μ: $\{\mu_p(E), \nu_p(E)\}$ and $\{\mu_{p/P}(E), \nu_{p/P}(E)\}$ generated by "prediction" of the 231 earthquakes with magnitude MwHRV ≥ 5.8 and depth ≥ 70 km in April 10, 1992–September 13, 1994 using the *p* and *p/P* maps computed for April 10, 2002–September 13, 2004.

the diagonal one should expect in case of naturally clustered earthquakes (in fact, the deviation is larger than in Figs. 1 and 2).

Finally, it appears rather evident that the efficiency of the forecast induced predictions is poor: the sum of errors $\mu+\nu$ even with aftershocks included into statistic has the minimal values of 83.3% for the "probability *p*" and of 82.7% for the "probability ratio *p*/*P*", which is inferior to a very simple prediction strategy (Molchan, 2003) that guarantees the value about 70%. Specifically, the simple strategy suggests "given a large event, wait three-quarters of the recurrence time and then call an alert to be canceled only by the next large event". For further comparison, in 1992–2003 the results of the ongoing global test of the M8 algorithm predictions aimed at magnitude 8.0 or larger earthquakes sum up to $\mu+\nu=0.302+0.222=52.4\%$ (Kossobokov, 2004b), which distance to the diagonal of random guessing is three times larger than that of the probabilistic forecast induced predictions. Of course, the loss function $\gamma(\nu,\mu)=\mu+\nu$ is not the only one of practical interest (e.g., see Molchan, 2003), therefore, it deserves mentioning that if we consider the two combinations of the M8 predictions with the forecast induced ones, then their lower envelopes $\Gamma_{\mathrm{M8}\&p}$ and $\Gamma_{\mathrm{M8}\&p/P}$ would rest on the *p* or *p*/*P* generated points for the space–time volumes of the immediate fore- or aftershocks only (i.e., $\mu<1\%$).

Many of the existing "precursors" and "candidates to precursors" can be reformulated into algorithms, preferably, by their discoverers, so that to provide an opportunity for their objective verification in the prediction experiments. Any result of such an experiment, either a success or a failure, clarifies our knowledge about seismic processes by supporting or rejecting the underlying conceptions. In this sense the forecast experiment, which David D. Jackson and Yan Y. Kagan carry on in real-time, can serve as an example to follow.

Acknowledgements

The author thanks David D. Jackson and Yan Y. Kagan for an existing opportunity of independent evaluation of their method. The remarks by Yan Y. Kagan and the anonymous reviewer were very helpful for improving the original manuscript. The research has been partly supported by the Program No. 13 “Change in Environment and Climate: Natural Catastrophes” of the Presidium of the Russian Academy of Sciences and EC Sixth Framework Programme (Priority FP6-2003-NEST-PATH “Tackling Complexity in Science”, Project E2-C2 “Extreme Events: Causes and Consequences”).

References

Allen, C.R., Edwards, W., Hall, W.J., Knopoff, L., Raleigh, C.B., Savit, C.H., Toksoz, M.N., Turner, R.H., 1976. Predicting earthquakes: A scientific and technical evaluation—with implications for society. Panel on Earthquake Prediction of the Committee on Seismology, Assembly of Mathematical and Physical Sciences, National Research Council, U.S. National Academy of Sciences, Washington, DC.

Cyranoski, D., 2004. A seismic shift in thinking. Nature 431, 1032–1034.

Gardner, J., Knopoff, L., 1974. Is the sequence of earthquakes in S. California with aftershocks removed Poissonian? Bull. Seismol. Soc. Am. 64 (5), 1363–1367.

Geller, R.J., 1997. Earthquake prediction: a critical review. Geophys. J. Int. 131, 425–450.

Jackson, D.D., Kagan, Y.Y., 1999. Testable earthquake forecasts for 1999. Seismol. Res. Lett. 70, 393–403.

Kagan, Y.Y., Jackson, D.D., 2000. Probabilistic forecasting of earthquakes. Geophys. J. Int. 143, 438–453.

Keilis-Borok, V.I., 1982. A worldwide test of three long-term premonitory seismicity patterns: a review. Tectonophysics 85, 47–60.

Keilis-Borok, V.I., 1964. Seismology and logics. Research in Geophysics vol. 2. M.I.T. Press, pp. 61–79.

Keilis-Borok, V.I., Malinovskaya, L.N., 1964. One regularity in the occurrence of strong earthquakes. J. Geophys. Res. 69, 3019–3024.

Keilis-Borok, V.I., Soloviev, A.A. (Eds.), 2003. Nonlinear Dynamics of the Lithosphere and Earthquake Prediction. Springer, Heidelberg. 337 pp.

Keilis-Borok, V.I., Knopoff, L., Rotwain, I.M., 1980. Bursts of aftershocks, long-term precursors of strong earthquakes. Nature 283, 259–263.

Kossobokov, V.G., 2004a. “The West Pacific short-term forecast of earthquakes with magnitude MwHRV ≥ 5.8” by D. Jackson and Y. Kagan: An independent evaluation of efficiency. In: The analysis of geodynamic and seismic processes. Moscow: GEOS, 201–211 (Computational Seismology Anniversary Issue 35; in Russian).

Kossobokov, V.G., 2004b. Earthquake prediction: basics, achievements, perspectives. Acta Geod. Geoph. Hung. 39 (2–3), 205–221.

Kossobokov, V., Shebalin, P., 2003. 4. Earthquake prediction. In: Keilis-Borok, V.I., Soloviev, A.A. (Eds.), Nonlinear Dynamics of the Lithosphere and Earthquake Prediction. Springer, Heidelberg, pp. 141–207.

Kossobokov, V.G., Keilis-Borok, V.I., Romashkova, L.L., Healy, J.H., 1999. Testing earthquake prediction algorithms: Statistically significant real-time prediction of the largest earthquakes in the Circum-Pacific, 1992–1997. Phys. Earth Planet. Inter. 111 (3–4), 187–196.

Kossobokov, V.G., Keilis-Borok, V.I., Turcotte, D.L., Malamud, B.D., 2000. Implications of a statistical physics approach for earthquake hazard assessment and forecasting. Pure Appl. Geophys. 157, 2323–2349.

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

Molchan, G.M., 2003. 5. Earthquake prediction strategies: a theoretical analysis. In: Keilis-Borok, V.I., Soloviev, A.A. (Eds.), Nonlinear Dynamics of the Lithosphere and Earthquake Prediction. Springer, Heidelberg, pp. 208–237.

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

Nature Debates, 1999. http://www.nature.com/nature/debates/earthquake/equake_frameset.html.

Wyss, M. (Ed.), 1991. Evaluation of Proposed Earthquake Precursors. AGU, Washington, D.C.

Wyss, M., 1997a. Cannot earthquakes be predicted? Science 278, 487–488.

Wyss, M., 1997b. Second round of evaluation of proposed earthquake precursors. Pure Appl. Geophys. 149, 3–16.