J. Appl. Probab. **38A**, 222–231 (2001)
Printed in England

AN EVALUATION OF CHINESE ANNUAL EARTHQUAKE PREDICTIONS, 1990–1998

YAOLIN SHI,[1] *University of Science and Technology of China and Academica Sinica*
JIE LIU[2] AND
GUOMIN ZHANG,[2] *China Seismological Bureau*

Abstract

The annual earthquake predictions of the China Seismological Bureau (CSB) are evaluated by means of an R score (an R score is approximately 0 for completely random guesses, and approximately 1 for completely successful predictions). The average R score of the annual predictions in China in the period 1990–1998 is about 0.184, significantly larger than 0.0. However, background seismicity is higher in seismically active regions. If a 'random guess' prediction is chosen to be proportional to the background seismicity, the expected R score is 0.123, and the nine-year mean R score of 0.184 as observed is only marginally higher than this background value. Monte Carlo tests indicate that the probability of attaining an R score of *actual prediction by background seismicity based on random guess* is about $\frac{1}{3}$. It is concluded that earthquake prediction in China is still in a very preliminary stage, barely above a pure chance level.

Keywords: Earthquake prediction; prediction evaluation; Chinese earthquake

AMS 2000 Subject Classification: Primary 86A15

1. Introduction

Forecasting earthquakes is at the frontier of research in seismology. David Vere-Jones (1970) was amongst the first to apply statistical methods in research on predicting earthquakes. The China Seismological Bureau (CSB), formerly the State Seismological Bureau (SSB), is the only governmental institution in the world that is dedicated to monitoring precursors of earthquakes and earthquake prediction. The CSB supervises 8 research institutes and 30 provincial or municipal seismological bureaux, with about 800 observational stations in total covering most seismically active and populated areas of China. The provincial or municipal bureaux are responsible for local predictions. They also report their information to the Center for Analysis and Prediction (CAP) of the CSB. The CAP is in charge of analyzing earthquake risks and makes predictions on a nationwide basis. These predictions are not available to the public or media: they are reported to the government, and only the government is authorized by law to

Received August 2000.
[1] Postal address: Graduate School, University of Science and Technology, Beijing 100039, China. Email: `shiyl@sun.ihep.ac.cn`
[2] Postal address: Center for Analysis and Prediction, China Seismological Bureau, Beijing 100036, China.

make decisions on mitigation actions. Details of Chinese earthquake predictions can be found in Mei *et al.* (1993) and SSB (1996).

Chinese earthquake predictions are usually made in progressive steps of long-term, medium-term, and short-term (or imminent) predictions. Typical medium-term predictions of one year are made at a 'national consultative meeting on seismic tendency' (Wu, 1997). These predictions are reported to the State Council. The State Council then issues formal government documents to ministries of the State Council and provincial and municipal governments.

The annual predictions are made by consensus of experts from the CAP and provincial seismological bureaux, based on a series of preparatory meetings at provincial level. Various data of seismicity parameters, deformation, apparent resistivity, ground water, ground stress, gravity, magnetic field, etc. are used in these predictions (e.g. Mei *et al.*, 1993; SSB, 1996). How these CSB predictions should be evaluated is a question of great interest to seismologists all over the world: earthquake predictions and their evaluation are very topical questions that are strongly debated (e.g. Stark, 1996; Geller *et al.*, 1997). The Chinese annual predictions have largely remained unknown because at the time they are made they are confidential. An objective evaluation of the CSB predictions is certainly worthwhile.

In this paper, we evaluate the CSB predictions in the 1990s only, because since 1990 the form of these CSB predictions has been more uniform. Our evaluation is based on the original prediction maps from the official Chinese State Council documents (CAP, SSB 1990–1998) so as to guarantee that all our data are reliable and that the predictions have indeed been made in advance of the predicted year. Our tests are somewhat different from testing a single prediction algorithm, such as M8, which can be applied to any region for which a catalogue is available. We can only make an evaluation of different strategies for making predictions and compare these with CSB predictions.

All the CSB annual predictions were made in advance of the predicted year by consensus of experts. Stark (1996) pointed out that there are many potential pitfalls in assigning statistical significance to 'successful' earthquake predictions. He suggested (1996, 1997) that it is preferable to treat the observed seismicity as given and compare the predictions with randomly generated predictions of various types. The benefit is that the stochastic component of the null hypothesis is in the prediction, not the Earth. This procedure was also taken by Kagan (1996) in comparing predictions based on foreshocks to the VAN predictions. In this paper we evaluate the actual prediction of each year by an R score, comparing the observed R score with two different prediction algorithms to judge its statistical significance.

2. *R* score

Many systems of scores have been proposed for the evaluation of earthquake predictions: here we also use an R score method. It is in fact difficult to use a single index to cover all important aspects in evaluating predictions, and an R score may not be the best score scheme. We choose an R score partly because it is the CSB 'official' evaluation method; this is convenient when comparing various prediction methods in China.

The total region is divided into N cells. A *positive prediction* (for a given cell) means that an earthquake of magnitude in a specified range (the 'predicted magnitude range', which in our case means 5.0 or larger) is predicted to occur in the cell within a

TABLE 1: Earthquake prediction counts.

Activity	Predictions Positive	Negative	Total activity
Earthquake	n_1^1	n_1^0	N_1
No event	n_0^1	n_0^0	N_0
Total predictions	N^1	N^0	N

time window of one year. A *negative prediction* is a prediction that no such earthquake will occur in the given cell within the one year time span. The R score is based on the following counts, which are entries in Table 1:

N^1, N^0 the number of positive (negative) predictions;
N_1, N_0 the number of cells where earthquakes do (do not) occur;
n_1^1, n_0^1 the number of positive predictions with (without) earthquake occurrences;
n_1^0, n_0^0 the number of negative predictions with (without) earthquake occurrences.

It follows that $N^1 = n_1^1 + n_0^1$, $N^0 = n_1^0 + n_0^0$, $N_1 = n_1^1 + n_1^0$, $N_0 = n_0^1 + n_0^0$ and $N = N^1 + N^0 = N_1 + N_0$. Further, define the *failure rate* $a = n_1^0/N_1$, the *false alarm rate* $b = n_0^1/N_0$, the *success rate of positive prediction* $c = n_1^1/N_1$, and the *success rate of negative prediction* $d = n_0^0/N_0$. Finally the score R is defined as $R = c - b$ (then we must also have $R = d - a = 1 - a - b = c + d - 1$). We can interpret R as

$$R = \frac{\text{predicted earthquakes}}{\text{total earthquakes}} - \frac{\text{false positive predictions}}{\text{number of aseismic areas}} = \frac{n_1^1}{N_1} - \frac{n_0^1}{N_0}.$$

When all positive and negative predictions are correct ($a = b = 0$, $c = d = 1$), $R = 1$. When all predictions are wrong ($a = b = 1$, $c = d = 0$), $R = -1$. A random prediction leads to $R \approx 0$ (see the appendix), while a meaningful prediction must have $R > 0$. This R score provides a simple but objective way of evaluating the overall performance of predictions.

3. *R* scores of Chinese annual predictions

We apply the R score scheme described above to test the CSB annual predictions for China since 1990.

Figure 1 shows the distribution of earthquakes with magnitude $M \geq 5.0$ in China from 1900 to 1989. The map is divided into cells of size $0.5° \times 0.5°$ (in terms of latitude and longitude), making 3743 cells in all. For the purposes of our analysis we excluded areas (cells) which cannot be monitored.

Figure 2 shows the cell frequencies of occurrence of earthquakes with $M \geq 5.0$, calculated from spatially smoothed grid data from 1900 to 1989. The value of each cell is averaged from the cell and its eight surrounding cells.

Annual predicted areas and earthquake occurrences from 1990 to 1998 were drawn, as shown for example in Figure 3 for the year 1990. The circled areas are actual predictions made as a result of the CSB annual meeting. For convenience of statistics, we call a cell a *positive predicted area* if it covers any part of the circled area, i.e. the circumscribed polygon of the prediction circle defines exactly the predicted area (grey areas in Figure 3). Any earthquake with $M_s \geq 5$ occurring inside the polygon is

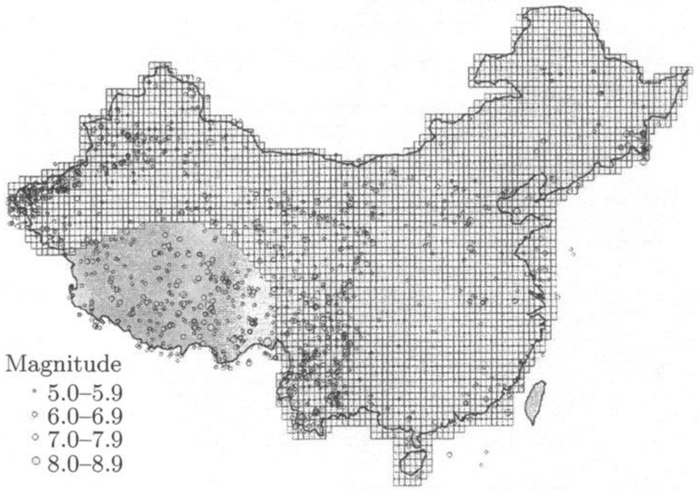

FIGURE 1: Epicenter distribution in earthquake prediction monitored area of China (from 1900 to 1989). In total 3743 cells used for statistics in Table 2. Shaded areas are not included because of lack of precursor data.

TABLE 2: Statistics on Chinese Earthquake Predictions, 1990–1998.

	n_1^1	N_1	c	n_0^1	N_0	b	R	p_r
1990	2	12	0.167	197	3731	0.053	0.114	0.131
1991	5	19	0.263	343	3724	0.092	0.171	0.026
1992	3	10	0.300	336	3733	0.090	0.210	0.055
1993	3	14	0.214	285	3729	0.076	0.138	0.087
1994	1	10	0.100	205	3733	0.055	0.045	0.432
1995	5	18	0.278	300	3725	0.081	0.197	0.012
1996	4	11	0.364	406	3732	0.109	0.255	0.025
1997	4	11	0.364	339	3732	0.099	0.265	0.014
1998	3	8	0.375	306	3735	0.082	0.293	0.023
Average	3.33	12.55	0.265	301.9	3730.4	0.081	0.184	0.044

considered a success, and outside the polygon is considered a failure. Only main shocks are counted (i.e. aftershocks are ignored in the statistics). Aftershocks are deleted by using an algorithm of Keilis-Borok and Knopoff (1980).

4. Comparison with prediction strategies

4.1. Completely random prediction

Although for a large number of samples, random prediction would score 0, it is quite possible for small samples that random prediction may yield a positive R score. The probability of making predictions equal to or better than the actual ones by random

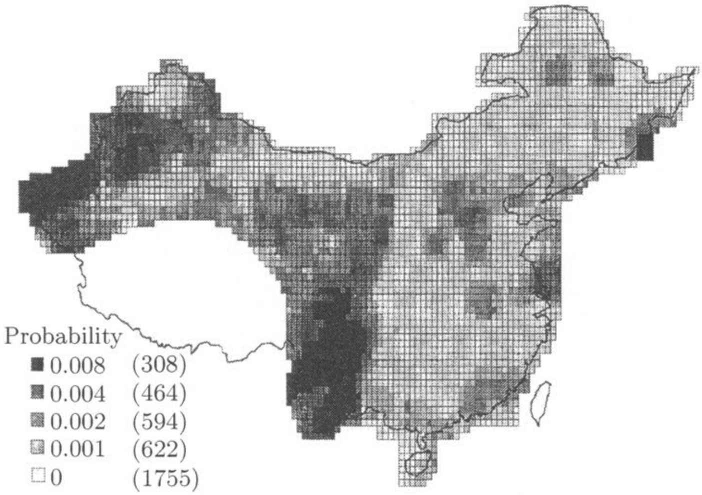

FIGURE 2: Spatially smoothed cellular annual frequency (to be regarded as an approximation to the background probability) of earthquake occurrence calculated from the average number of each cell and its eight surrounding cells.

guessing, i.e. picking out N_1 cells randomly from the N cells and containing at least n_1^1 earthquake cells, is calculated as

$$p_r = \sum_{i=n_1^1}^{N_1} \binom{N_1}{i} \binom{N_0}{N^1 - i} \Big/ \binom{N}{N_1}.$$

From 1990 to 1998, an annual average of 13 earthquakes occurred in the monitored area of 3743 cells. In an average CSB prediction, which covers 305 cells, there are 3 earthquakes correctly predicted, 10 earthquakes missed and 302 cells of false alarms. The average R score is 0.184, and the geometric average of p_r is 0.044. The R scores for the CSB predictions are significantly superior to those of random prediction.

4.2. Prediction based on observed long-term frequencies

Observation over long periods has shown that the relative frequencies of earthquakes differ in different regions. Two conclusions follow immediately: first, that completely random prediction cannot be expected to be satisfactory, and second, that it is reasonable to attempt to 'predict' earthquakes at differential rates for different regions. As a start we therefore use the long-term observed annual frequencies p_i in cell i as a basis for 'background probability prediction'.

One possible strategy is to choose cell i for prediction by tossing a biased coin with probability *proportional* to p_i, kp_i say, where k is the ratio of the number of positively predicted cells to the annual average number of earthquakes. Then (see the appendix) the expectation R_2 of this score is

$$R_2 \equiv \mathrm{E}(R) = k\sigma_p^2\Big(\frac{1}{\bar{p}} + \frac{1}{1-\bar{p}}\Big) = \frac{k\sigma_p^2}{\bar{p}(1-\bar{p})},$$

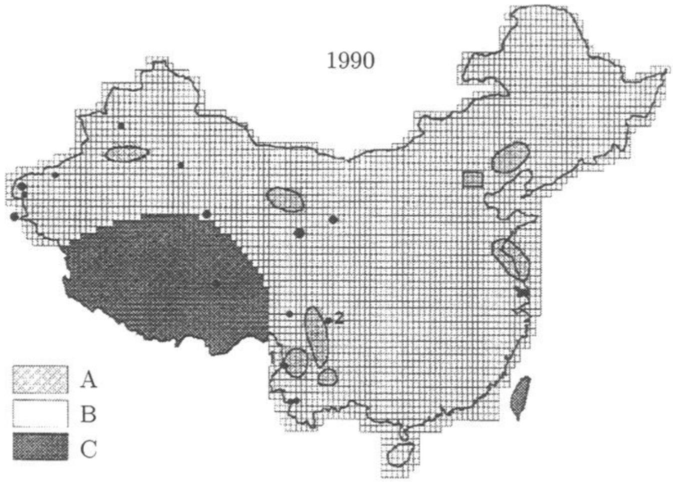

FIGURE 3: Predictions and earthquakes ($M \geq 5$) of 1990. Circled areas are the actual CSB predicted risk regions of the year. The grey area A shows cells regarded as having an earthquake predicted in the statistics; the white area B shows cells with no earthquakes predicted; the dark area C is areas not included in statistics because of lack of precursor data.

TABLE 3: Statistics on R score measures, 1990–1998.

	R	R_2	R_{2A}	p_A
1990	0.114	0.097	0.116	0.654
1991	0.171	0.170	0.110	0.382
1992	0.210	0.165	0.099	0.273
1993	0.138	0.139	0.154	0.643
1994	0.045	0.100	0.096	0.830
1995	0.197	0.149	0.096	0.190
1996	0.255	0.200	0.122	0.145
1997	0.265	0.182	0.143	0.245
1998	0.293	0.151	0.171	0.318
Average	0.184	0.150	0.123	0.350

where $\bar{p} = \sum p_i/N$, the average p_i over the N cells. The results from 1990 to 1998 are summarized in Table 3 together with the actual scores R. For most years except 1994, R_2 is smaller than or equal to the R score computed from the actual predictions. The nine-year average R_2 is 0.150, also smaller than the nine-year average 0.184 of the observed R scores. We conclude that CSB predictions are marginally better than these background probability predictions.

The expectation R_2 is based on two assumptions: that the earthquakes occurring in the 1990s followed the 1900–1989 background probabilities, and that these probabilities are known exactly. Clearly, neither of these assumptions is true, so we also conducted some Monte Carlo tests. For each year, we randomly chose the same number of cells as for the CSB prediction, such that the chance of each cell being chosen was proportional to its background probability, and then calculated the R score. We carried out 5000 such tests for each year. The results (see also the next paragraph), yielded the yearly average R score shown in Table 3 as R_{2A}. The nine-year average R_{2A} is 0.123, smaller than the theoretical expectation R_2 of 0.150, probably because both the long-term frequencies are only estimates of the background probabilities, and earthquakes in the 1990s may not exactly follow these probabilities. The nine-year average R_{2A} of 0.123 is also smaller than the actual R score 0.184, indicating that CSB predictions are marginally better than guesses based on these backbround probabilities.

We made other comparisons of the CSB and these Monte Carlo 'predictions':

(1) Of the annual number of actual earthquakes (nine-year average 12.55), the CSB gave a nine-year average of 3.33 positive predictions compared with 2.56 from the Monte Carlo predictions.

(2) For each year we constructed histograms (see Figure 4), based on 5000 tests, of the number of earthquakes with positive predictions.

(3) For each year we calculated the proportion p_A of the 5000 tests with R scores larger than the R score realized from the CSB predictions (see Table 3).

From the viewpoint of CSB-based predictions, the worst and best cases were 1994 and 1996, when 83% and 15% of the 5000 tests had background probability prediction R scores larger than the CSB-based R scores. The average p_A is 0.350, meaning that the chance of the CSB prediction being better (in terms of the R score) than the background probability prediction is about $\frac{2}{3}$.

5. Discussion and conclusion

Earthquake prediction is a highly topical subject. We have suggested two prediction strategies and compared their R scores with the R score of the data for CSB-based predictions. The completely random strategy ignores the spatial variability of seismicity, and is markedly inferior to the CSB predictions.

The second strategy considers the spatial distribution of seismicity over the period 1900–1989, even though it is always possible for a large earthquake to occur where none has been seen before. CSB prediction is only marginally better than background-based random prediction. In comparison with background probability prediction for high-risk areas only, the R score of CSB prediction is lower, but can be increased significantly by combination with the high-risk area prediction. Improved CSB prediction could be superior to high-risk area only prediction because it uses more than probability-based information. However on its own, this cannot yet be shown to give significant improvement: CSB prediction is still empirical and in a preliminary stage of development. Serious advances are needed to make predictions that can be approved by most scientists.

Appendix. *R* score expectations for different prediction strategies

Our discussion and all quantities relate to a given time unit, typically, one year. Our concern is with making predictions for N cells.

Suppose that (in the given year) the probability of earthquake occurrence in cell i is p_i which is obtained from long-term data and may vary from one cell to another. Define variables

$$U_i = \begin{cases} 1 & \text{if an earthquake occurs in cell } i, \\ 0 & \text{if no earthquake occurs in cell } i, \end{cases}$$

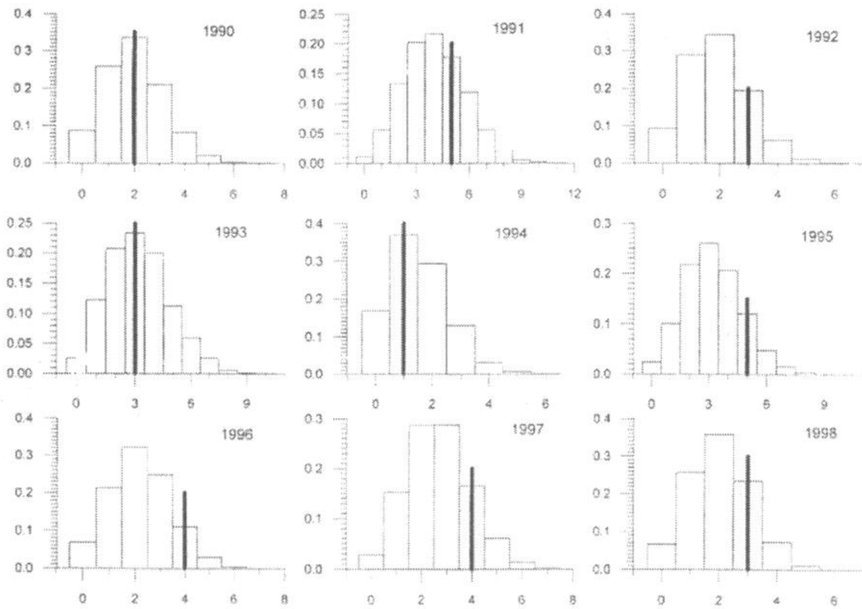

FIGURE 4: Histograms for 1990 to 1998, based on 5000 tests, of the number of correct positive predictions using the proportional background probability prediction algorithm (see text). The heavy line for each year shows the number of actual earthquakes with positive predictions based on the CSB method.

and

$$V_i = \begin{cases} 1 & \text{if an earthquake is predicted to occur in cell } i, \\ 0 & \text{if no earthquake is predicted for cell } i. \end{cases}$$

Then R is defined by

$$\begin{aligned} R &= 1 - a - b \\ &= 1 - \left(\frac{\sum_i (1 - V_i) U_i}{\sum_j U_j} + \frac{\sum_i V_i (1 - U_i)}{\sum_j (1 - U_j)} \right) \\ &= \frac{\sum_i V_i U_i}{\sum_j U_j} - \frac{\sum_i V_i (1 - U_i)}{\sum_j (1 - U_j)} . \end{aligned}$$

We discuss $\mathrm{E}(R)$ for three different prediction strategies.

Case 1: *Predictions are made completely at random.* Each cell is selected to carry predictions as Bernoulli trials with a constant probability p independent of any other factors. Then the expectation of R over the V_i is zero, because

$$\mathrm{E}(R) = p \frac{\sum_i U_i}{\sum_j U_j} - p \frac{\sum_i (1 - U_i)}{\sum_j (1 - U_j)} = p - p = 0.$$

Note that this is independent both of any chosen value of p, and of the values of all p_i.

Case 2. *Predictions based on long-term rates.* If the value of p_i for each cell is known from long-term records, then we may use a rule that a cell is given a positive prediction with probability proportional to p_i, kp_i say. In practice, k would be chosen to be the ratio of the number of cells predicted to the number of expected earthquakes. We have

$$\begin{aligned} R_2 \equiv \mathrm{E}(R) &= \mathrm{E}\Big[\sum_i V_i\Big(\frac{U_i}{\sum_j U_j} - \frac{1-U_i}{N-\sum_j U_j}\Big)\Big] \\ &= \mathrm{E}\Big[\frac{1}{(\sum_j U_j)(N-\sum_j U_j)}\,\mathrm{E}\Big(\sum_i V_i(NU_i - \textstyle\sum_j U_j)\;\Big|\;\{U_i\}\Big)\Big] \\ &= \mathrm{E}\Big(\frac{N\sum_i kp_iU_i - (\sum_j U_j)\sum_i kp_i}{(\sum_j U_j)(N-\sum_j U_j)}\Big). \end{aligned}$$

Write

$$\sigma_p^2 = \frac{\sum(p_i-\bar{p})^2}{N} = \int_0^1 (p-\bar{p})^2\,\mathrm{d}F(p), \qquad \bar{p} = \frac{\sum p_i}{N},$$

where $F(\cdot)$ describes the distribution of values p_i over the N cells. Since $\mathrm{E}\big(U_i \mid \sum_j U_j\big) = \big(\sum_j U_j\big)p_i/(N\bar{p})$, we can first condition on $\sum_j U_j$ and then finally obtain

$$R_2 = k\,\mathrm{E}\Big(\frac{\sum_i p_i^2/\bar{p} - N\bar{p}}{N-\sum_j U_j}\Big) = k\,\mathrm{E}\Big(\frac{\sigma_p^2}{\bar{p}(1-\sum_j U_j/N)}\Big) \approx \frac{k\sigma_p^2}{\bar{p}(1-\bar{p})}.$$

Case 3. *Predictions only for higher p_i.* Suppose now that we predict events only for cells with higher probabilities p_i, i.e. we choose cell i if and only if its background probability $p_i > p_t$ for some threshold probability p_t. In this case,

$$\mathrm{E}\Big[\sum_{i:\,p_i>p_t}\frac{U_iV_i}{N}\Big] = \int_{p_t}^1 p\,\mathrm{d}F(p),$$

and then

$$\mathrm{E}(R) = \frac{1}{\bar{p}(1-\bar{p})}\int_{p_t}^1 (p-\bar{p})\,\mathrm{d}F(p).$$

Acknowledgment

We thank David Vere-Jones for reading the manuscript and contributing to the appendix. We also thank Ma Li, C.-Y. Wang, Yan Kagan and Keilis-Borok for valuable suggestions. This research is partially supported by the National Key Project of Science and Technology Development Program (G98040706) and the National Natural Science Foundation of China (NSFC-49774236).

References

CAP, SSB (Center for Analysis and Prediction of the State Seismological Bureau) (1990–1998). *Annual Summary Report* on National Earthquake Trends, 1990, ..., 1998.

Geller, R. G., Jackson, D. D., Kagan, Y. Y. and Mulargia, F. (1997). Earthquakes cannot be predicted. *Science* **275**, 1616–1617.

KAGAN, Y. Y. (1996). VAN earthquake predictions—an attempt as statistical evolution. *Geophys. Res. Lett.* **23**, 1315–1318.

KEILIS-BOROK, V. I. AND KNOPOFF, L. (1980). Burst of aftershocks of strong earthquakes. *Nature* **283**, 259–263.

MEI SHIRONG, FENG DEYI, ZHANG GUOMIN, ZHU YUEQING, GAO XU AND ZHANG ZHAOCHENG (1993). *Introduction to Earthquake Prediction in China.* Seismological Press, Beijing.

SSB (STATE SEISMOLOGICAL BUREAU) (1996). *The Selected Papers of Earthquake Prediction in China.* Seismological Press, Beijing.

STARK, P. B. (1996). A few considerations for ascribing statistical significance to earthquake predictions. *Geophy. Res. Lett.* **23**, 1399–1402.

STARK, P. B., (1997). Earthquake prediction: the null hypothesis. *Geophys. J. Internat.* **131**, 495–498.

VERE-JONES, D. (1970). Stochastic models for earthquake occurrence (with discussion). *J. Roy. Statist. Soc. Ser. B* **32**, 1–62.

WU, F. T. (1997). The annual earthquake prediction conference in China (National consultative meeting on seismic tendency). *Pure Appl. Geophys.* **149**, 249–264.