The degree of predictability of earthquakes in several regions of China: statistical analysis of historical data

Chunsheng Lu[a,b,*]

[a]*Centre for Advanced Materials Technology (CAMT), School of Aerospace, Mechanical and Mechatronic Engineering,*
The University of Sydney, Sydney NSW 2006, Australia
[b]*State key Laboratory of Nonlinear Mechanics (LNM), Institute of Mechanics, Chinese Academy of Sciences, Beijing 100080, China*

Abstract

The stress release model, a stochastic version of the elastic-rebound theory, is applied to the historical earthquake data from three strong earthquake-prone regions of China, including North China, Southwest China, and the Taiwan seismic regions. The results show that the seismicity along a plate boundary (Taiwan) is more active than in intraplate regions (North and Southwest China). The degree of predictability or regularity of seismic events in these seismic regions, based on both the Akaike information criterion (AIC) and fitted sensitivity parameters, follows the order Taiwan, Southwest China, and North China, which is further identified by numerical simulations.
© 2004 Elsevier Ltd. All rights reserved.

Keywords: Spatio-temporal seismicity; Stress release model; Degree of predictability; Historical earthquake data in China

1. Introduction

In terms of the elastic-rebound theory proposed by Reid (1910), stress in a seismically active region accumulates due to relative movement along faults. When the stress exceeds a certain threshold, for example, the strength of rocks, an earthquake occurs and the accumulated strain energy is released. Direct evidence for this theory was provided by geodetic data measured before and after the 1906 San Francisco earthquake. Since then, the elastic-rebound theory has been the basis of seismic gap model, time- or slip-predictable model, and many long-term earthquake prediction algorithms.

Provided the elastic-rebound theory is universally applicable, a large earthquake should be followed by a period of quiescence during which stress is restored to the level just before the event, whereas the real sequences in earthquakes are more complicated. A strong earthquake can be followed by a period of activation, and sometimes by another earthquake with comparable magnitude. Most plausibly this is consistent with a view that events in the Earth's crust form part of a tightly linked, near-critical process, exhibiting the self-similarity, long-range correlation and power-law distributions, which are characteristic of a physical process in near-critical state (Vere-Jones, 1976; Takayasu and Matsuzaki, 1988; Turcotte, 1992; Kagan, 1994; Main, 1996a). Earthquake prediction is a very difficult and challenging problem (Mogi, 1985; Main, 1996b). Over the last decade, whether earthquake prediction is possible has given rise to much controversy, and pessimistic attitudes seem to be in vogue in seismology. Based on the assumption that earthquakes are a self-organised critical phenomenon (Bak and Tang, 1989), it is declared that earthquakes cannot be predicted (Geller et al., 1997).

Our recent study, on synthetic catalogues generated by physical model, has found that the degree of predictability of earthquakes in a seismic region is closely related to the heterogeneity of tectonic structures (Ben-Zion, 1996; Lu and Vere-Jones, 2001). Of the real data extracted from different kinds of information, the temporal variation of seismicity, as given in earthquake catalogues, may most directly reflect the nature of earthquake-generating stress and tectonic structures (Zhao et al., 1990). However, in statistical analysis of historical earthquake data, we should take considerable care concerning the completeness of historical catalogues used in the modelling. It may, nevertheless, be of interest to apply a similar method to real

* Tel.: + 61-2-9351-2348; fax: +61-2-9351-7060.
   E-mail address: chunsheng.lu@aeromech.usyd.edu.au (C. Lu).

earthquake data. In this paper, such an attempt will be applied to Chinese historical earthquake catalogues by using a simple statistical model. One of the important questions we examine is whether there is a relationship between the degree of predictability and seismic regions.

2. Stress release model

Through a development of the Markov model suggested by Knopoff (1971), the stress release model, a stochastic version of the elastic-rebound theory, was formulated by Vere-Jones (1978), and subsequently applied to the statistical analysis of historical earthquake data from China, Japan, Iran, and New Zealand (Vere-Jones and Deng, 1988; Zheng and Vere-Jones, 1991, 1994; Lu et al., 1999a; Lu and Vere-Jones, 2000, 2001).

In the stress release model, it is assumed that a scalar regional stress level, $X(t)$, increases deterministically between two earthquakes and releases stochastically as a Markov process. The evolution of stress versus time follows the equation

$$X(t) = X(0) + \rho t - S(t), \tag{1}$$

where $X(0)$ is the initial stress level, ρ is the constant loading rate from external tectonic force, and $S(t) = \sum_{t_i < t} S_i$ is the accumulated stress release from earthquakes within the region over the period $(0, t)$, where t_i, S_i are the origin time and stress release associated with the i-th earthquake, respectively (Zheng and Vere-Jones, 1991, 1994).

The value of stress release during an earthquake can be estimated from its magnitude, M, in terms of the empirical formula, $M = 2/3 \log_{10} E + \text{const.}$, where E is the energy released during the earthquake (Gutenberg and Richter, 1954). For simplicity, the stress drop, S, is supposed to be proportional to the square root of the released energy, i.e. $S \propto E^{1/2}$. Then, we have the formula

$$S = 10^{0.75(M - M_0)}, \tag{2}$$

where M_0 is the reference magnitude, and $M_0 = 6.0$ is used in the subsequent analysis.

The probability intensity of an earthquake occurrence is controlled by a risk function, $\Psi(X)$. Generally, the risk increases nonlinearly with the stress level, X. If the solid media had an exact critical strength, $\Psi(X)$ would have to be zero until X reaches the critical strength, and infinite beyond it. By contrast, a finite constant value of $\Psi(X)$ corresponds to a pure random process in which the occurrence of events is independent of the stress level, X. Thus, the simplest choice of $\Psi(X)$ is taken as an exponential function, $\Psi(X) = \exp(\mu + \nu X)$, where μ and ν represent a background constant and the sensitivity to risk, respectively. This is a compromise between the time-predictable and pure random (Poisson) processes.

It is further assumed that the probability distribution of earthquake magnitudes is independent of the stress level,

and as a default, is governed by the standard Gutenberg–Richter law. Here, the key to statistical analysis is that the data in historical earthquake catalogues can be treated as a marked point process in time–stress space with the conditional intensity function

$$\lambda(t) = \exp\{\alpha + \nu[\rho t - S(t)]\}, \tag{3}$$

where $\alpha = \mu + \nu X(0)$, ν, ρ are the parameters to be fitted. Estimates of these parameters are found by maximising the log-likelihood

$$\ln L = \sum_{i=1}^{N} \ln \lambda(t_i) - \int_{T_1}^{T_2} \lambda(t) \mathrm{d}t, \tag{4}$$

where N is the number of events over the observation interval (T_1, T_2) (Harte, 1998).

Comparisons between two models will be based on the AIC, which is defined as

$$\text{AIC} = -2 \ln \hat{L} + 2k, \tag{5}$$

where $\ln \hat{L}$ is the maximum log-likelihood for a given model and k is the number of parameters to be fitted in the model (Akaike, 1977). This represents a rough way of compensating for the effect of adding parameters, and is a useful heuristic measure of the relative effectiveness of different models. For example, in comparing the stress release model with three parameters against the Poisson model with only one ($\nu = \rho = 0$ in Eq. (3)), the more complex model must demonstrate a significantly better fit to justify the additional parameters. In typical cases, model differences which would be significant at around the 5% confidence level correspond to differences in AIC values of around 1.5–2. The best model is that for which AIC has the smallest value.

In order to find a scale-free version of the sensitivity to risk, ν, the Eq. (3) can be rewritten in the form

$$\lambda(t) = \lambda_0 \exp\left\{\eta\left[\frac{t}{t_0} - \frac{S(t)}{S_0}\right]\right\}, \tag{6}$$

where S_0 is the observed mean stress drop, $S_0 = \Sigma S_i / N$, $t_0 = S_0 / \rho$ is the time to build the stress from 0 up to S_0, and $\eta = \nu S_0$ is the scale-free sensitivity to risk.

3. Historical earthquake data

The Chinese historical earthquake catalogue is one of the oldest and most extensive in existence in the world. Here, the data from three strong earthquake-prone regions, including North China, Southwest China, and Taiwan, as illustrated in Fig. 1, are chosen in terms of both the completeness and extensiveness of historical records (Gu, 1983; Zheng and Vere-Jones, 1991; Zhuang and Ma, 1998). We should notice that in all the data sets studied below, aftershocks were deliberately removed from earthquake catalogues using the simple window system of Gardner and Knopoff (1974) since the model framework
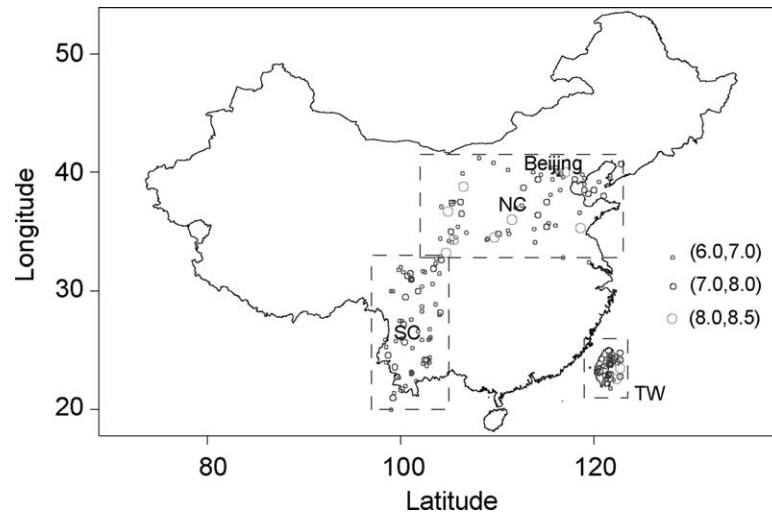
Fig. 1. The epicentre distribution of events in three earthquake-prone regions of China: North China (NC); Southwest China (SC); and Taiwan (TW). Earthquakes with the magnitude $M \geq 6.5$ in Taiwan region are illustrated.

presented above is only suitable to the analysis of mainshocks.

3.1. North China

The seismic region selected here (Fig. 1) represents a substantial part of northern China, and comprises the north-eastern coastal region, excluding the Yellow Sea, but including the Bohai Sea, and the main part of the Ordos Plateau, including in particular its western and southern boundaries. It is essentially bounded by latitudes 32°N and 42°N, longitudes 104°E and 102°N, and covers about the 500-year period from 1480 to 2000, where the catalogue is believed to be complete for events with magnitude $M \geq 6.0$ (see Fig. 1). The number of events is 65.

3.2. Southwest China

The tectonic structure in the Southwest China region (Fig. 1) is much more complicated than the North China region. There are a few main faults in this region, such as the Longmenshan fault, the Mabian fault, the Xianshuihe fault, etc. From north to south, the orientation of the maximum principal stress gradually changes from northeast to north-west (Ma, 1987). The region studied here is essentially bounded by latitudes 20°N and 34°N, longitudes 98°E and 105°E, and covers the 100-year period from 1900 to 2000, since the recording before 1900 is incomplete for events with magnitude $M \geq 6.0$. The number of events is 66.

3.3. Taiwan

Of the three seismic regions shown in Fig. 1, the Taiwan region, which is located on the boundary between the Pacific and Eurasian plates, is the most active. The seismic pattern is quite different from the other two regions. The axis of

the maximum compression principal stress has roughly an E–W direction. The catalogue in the Taiwan region is incomplete and most of larger earthquakes in the ocean to the east were not recorded before 1900. Thus, the catalogue used here covers the 100-year period from 1900 to 2000 for events with a relatively higher threshold $M \geq 6.5$. The number of events is 44.

4. Data fitting and simulation

Two kinds of basic statistical models, the Poisson model and the stress release model, were applied to the analysis of the earthquake catalogues mentioned above. These two models represent two limiting cases: the pure random process and a simplified pseudo-periodic process. In the following we shall examine the results based on both the data fitting and simulations.

4.1. The degree of predictability

As discussed above, the relative effectiveness of different models (e.g. the stress release model and the Poisson model in this paper) in fitting the data can be determined by the difference of their AIC values, $\Delta AIC = AIC_p - AIC_s$, where AIC_s and AIC_p are the AIC values calculated

Table 1
The AIC values calculated using the simple stress release model (SRM) for the three seismic regions of China, where ΔAIC_s represents the AIC value of the SRM model and ΔAIC_p the AIC value of the Poisson model, $\Delta AIC = AIC_p - AIC_s$, and N is the number of events

Region	N	AIC_s	AIC_p	ΔAIC	$\Delta AIC/N$
North China	65	398.46	402.33	3.87	0.06
Southwest China	66	184.93	188.85	3.92	0.06
Taiwan	44	155.10	162.25	7.15	0.16

Table 2
The fitted parameters using the stress release model, as formulated in Eq. (3), to the three seismic regions of China.

Region	N	α	ν	ρ
North China	65	-2.453	0.010	1.169
Southwest China	66	-1.488	0.025	2.936
Taiwan	44	-1.293	0.052	3.060

Table 3
The fitted parameters using the modified version of the stress release model, as formulated in Eq. (6), to the three seismic regions of China

Region	N	t_0	S_0	η
North China	65	0.608	0.711	0.007
Southwest China	66	0.191	0.561	0.014
Taiwan	44	0.316	0.968	0.051

by the stress release model and the Poisson model, respectively. As listed in Table 1, the stress release model fits the data better than the Poisson model in all the three regions, and the differences of AIC values are substantial, i.e. $\Delta AIC \geq 2$. If we take the Poisson model as a reference, the changes of ΔAIC values in the three regions from large to small follow the order Taiwan, Southwest China, and North China. However, such a direct comparison might be misleading, as the different catalogues yield different numbers of events. In order to allow for this effect, we use the indicator, $\Delta AIC/N$, as a measure of the improvement in performance which is approximately independent of sample size (Kagan and Knopoff, 1977; Di Luccio et al., 1997). Apart from a little difference between the North China and Southwest China regions, a similar order was obtained (see Table 1). Here the larger the value of the ratio, $\Delta AIC/N$, the better the seismicity is fitted by the alternative model (i.e. the stress release model) relative to the Poisson model. Thus, the values of $\Delta AIC/N$ give an indication of the degree of periodicity of each catalogue, at least in the sense of the degree of departure from the Poisson model. They also provide a measure of the increase in predictability for each model over the base level provided by the Poisson model (Vere-Jones, 1998).

The fitted parameters in the intensity functions are set out in Tables 2 and 3, and the magnitude versus time, as well as
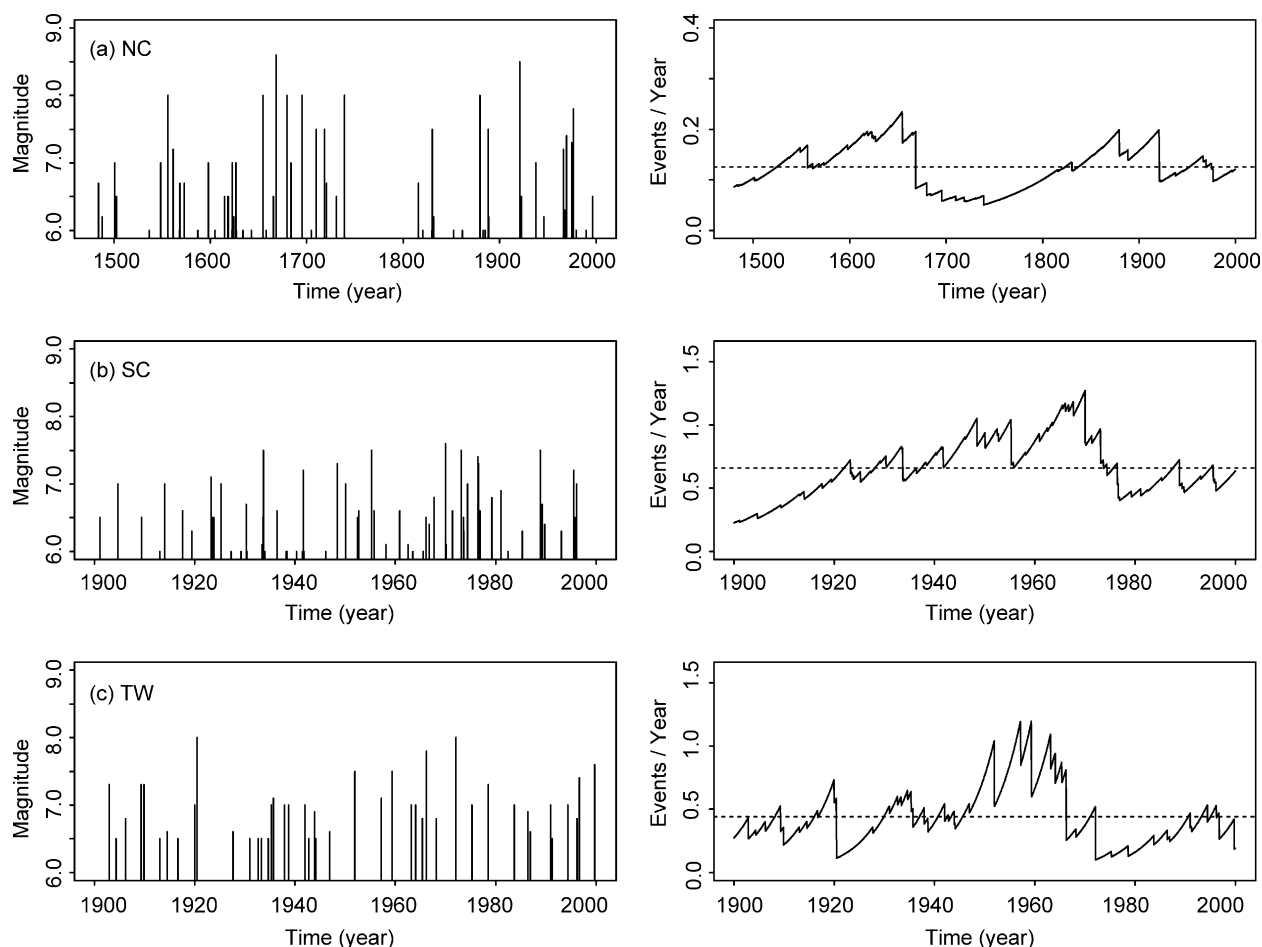

Fig. 2. The magnitude versus time (left column) and the risk function (events/year) versus time (right column) for each region calculated by the stress release model (solid line) and the Poisson model (dotted line): (a) North China (NC); (b) Southwest China (SC); and (c) Taiwan (TW) seismic regions.
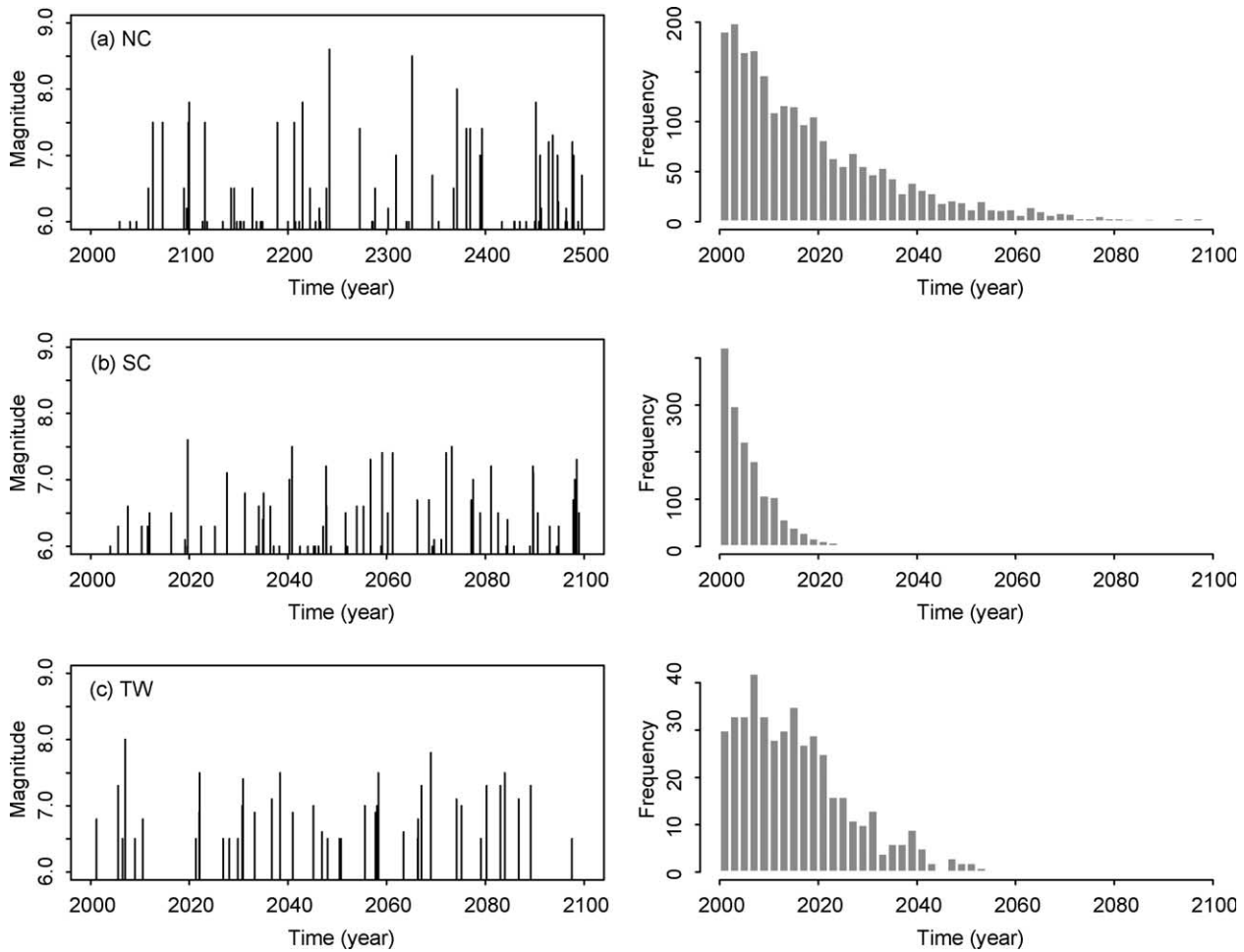
Fig. 3. The simulated events for the three regions, where the left hand column indicates the simulated catalogues, and the right hand column the simulated frequency. Here, the magnitude of next event is randomly chosen from the historical data (see Fig. 2) rather than the Gutenberg–Richter law.

the corresponding risk function (events/year), calculated for each region is shown in Fig. 2. It is obvious that the main differences between these regions lie in the values of ν or η, which represent the sensitivity to risk as formulated in Eqs. (3) and (6). In other words, the results show that the seismicity in the plate boundary region (Taiwan) is more active than those in intraplate regions (North China and Southwest China). The sensitivity to risk in the three regions follows the order Taiwan, Southwest China, and North China, which is consistent with the results in terms of the AIC values.

4.2. Simulations

Using the fitted parameters, the stress release model can be used to simulate the sequence of earthquakes forward in time. An example of the simulations, for each of the three regions, is shown in Fig. 3. The time history forward for the next 100 years was taken, namely over the period from 2000 to 2100. By using a large number of such simulations, we can easily find the probability distribution for the time to

the next earthquake (the right column in Fig. 3). Here we define the width of the 90% equitailed probability interval, from the fifth to the ninety-fifth percentiles, in units of the median interval, as a measure of the relative concentration, denoted by r_c (see Table 4 for details). The simulated results indicate that the degree of predictability, measured roughly by the ratio r_c, follows the same order, Taiwan, Southwest China, and North China, as obtained above.

Table 4
Prediction performance in the three seismic regions of China, where t_m is the mean interval which is defined as the ratio of the total time interval to the number of events with $M \geq M_c$ ($M_c = 6.0$ for North China and Southwest China regions, and $M_c = 6.5$ for Taiwan region), and r_c is the relative concentration, $r_c = (p_{95\%} - p_{5\%})/t_m$

Region	t_m	$p_{5\%}$	$p_{95\%}$	r_c
North China	20.675	1.115	59.495	2.822
Southwest China	5.714	0.336	15.826	2.711
Taiwan	15.895	1.431	36.693	2.350

5. Discussion and conclusion

In this paper, the stress release model, a simple stochastic version of the elastic-rebound theory, is applied to the historical earthquake data from North China, Southwest China, and Taiwan regions. The results show that, seismicity in the plate boundary region (Taiwan) is more active than those in the intraplate regions (North China and Southwest China). The degree of predictability or regularity in these seismic regions, based on both the AIC values and the sensitivity parameters fitted, follows the order Taiwan, Southwest China, and North China, which is in good agreement with that obtained from simulations.

It is worth noting that, in the stress release model, the stress field in a seismic region is treated as a scalar, which seems to be a reasonable approximation, if the real stress field has a dominant principal component. The real stress distribution, however, might be much more complicated. For example, geophysical consideration suggests that the North China region should be treated as containing at least two major components, the coastal region and the Ordos Plateau, roughly separated by the Taihang Mountains. To the east of this region, the principal stress is compressive, oriented in an E–W direction, and corresponds to pressure from the Pacific Plate. To its west, the stress is mainly in the N–S direction and driven by pressure from the Indian Plate and Tibetan Plateau (Ma, 1987; Ma et al., 1990). Detailed analysis can be carried out using a linear combination of the stress release model, or the 'linked stress release model' (see, Zheng and Vere-Jones, 1991, 1994; Lu et al., 1999a; Lu and Vere-Jones, 2000).

We should also notice that the length and time scales in the North China region are substantially different from those in the Southwest China and Taiwan regions. The influence of these factors on the predictability is not easy to determine, since the usable part of historical data is extremely short. An alternative method is to study the synthetic catalogues generated by a geophysical model (Lu et al., 1999b; Lu and Vere-Jones, 2001). Despite the crudity of the model used here in physical terms, it can still provide some useful information for forecasting long-term risk in a seismic region.

References

Akaike, H., 1977. On entropy maximisation principle, In: Krishnaiah, P.R., (Ed.), Applications of Statistics, Amsterdam, North Holland, pp. 27–41.

Bak, P., Tang, C., 1989. Earthquakes as a self-organized critical phenomenon. Journal of Geophysical Research 94, 15635–15637.

Ben-Zion, Y., 1996. Stress, slip, and earthquakes in models of complex single-fault systems incorporating brittle and creep deformations. Journal of Geophysical Research 101, 5677–5706.

Di Luccio, F., Console, R., Imoto, M., Murru, M., 1997. Analysis of short time-space range seismicity patterns in Italy. Annali Di Geofisica 40, 783–798.

Gardner, J.K., Knopoff, L., 1974. Is the sequence of earthquakes in southern California, with aftershocks removed, Poissonian? Bulletin of the Seismological Society of America 64, 1363–1367.

Geller, R., Jackson, D., Kagan, Y., Mulargia, F., 1997. Earthquake cannot be predicted. Science 275, 1616–1617.

Gu, G.X., 1983. Chinese Earthquake Catalogue, Science Press, Beijing, in Chinese.

Gutenberg, B., Richter, C.F., 1954. Seismicity of the Earth and Associated Phenomenon, Princeton University Press, Princeton.

Harte, D., 1998. Documentation for the Statistical Seismology Library, Research Report 98/10, School of Mathematical and Computing Sciences, Victoria University of Wellington, New Zealand.

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

Kagan, Y.Y., Knopoff, L., 1977. Earthquake risk prediction as a stochastic process. Physics of the Earth and Planetary Interiors 14, 97–108.

Knopoff, L., 1971. A stochastic model for the occurrence of main-sequence earthquakes. Reviews of Geophysical and Space Physics 9, 175–188.

Lu, C., Vere-Jones, D., 2000. Application of linked stress release model to historical earthquake data: comparison between two kinds of tectonic seismicity. Pure and Applied Geophysics 157, 2351–2364.

Lu, C., Vere-Jones, D., 2001. Statistical analysis of synthetic earthquake catalogues generated by models with various levels of fault zone disorder. Journal of Geophysical Research 106, 11115–11125.

Lu, C., Harte, D., Bebbington, M., 1999a. A linked stress release model for historical Japanese earthquakes: coupling among major seismic regions. Earth, Planets and Space 51, 907–916.

Lu, C., Vere-Jones, D., Takayasu, H., 1999b. Avalanche behaviour and statistical properties in a microcrack coalescence process. Physical Review Letters 82, 347–350.

Ma, X.Y. (Ed.), 1987. Lithospheric Dynamics Atlas of China, China Cartographic Publisher, Beijing.

Ma, Z., Fu, Z., Zhang, Y., Wang, C., Zhang, G., Liu, D., 1990. Earthquake Prediction: Nine Major Earthquakes in China, Seismological Press, Beijing.

Main, I., 1996a. Statistical physics, seismogenesis, and seismic hazard. Reviews of Geophysics 34, 433–462.

Main, I., 1996b. Long odds on earthquake prediction. Nature 385, 19–20.

Mogi, K., 1985. Earthquake Prediction, Academic Press, Tokyo.

Reid, H.F., 1910. The mechanism of the earthquake, In: Report of the State Earthquake Investigation Commission, Vol. 2. Carnegie Institute of Washington, Washington, pp. 16–28.

Takayasu, H., Matsuzaki, M., 1988. Dynamical phase transition in threshold elements. Physics Letters A 131, 244–247.

Turcotte, D.L., 1992. Fractals and Chaos in Geology and Geophysics, Cambridge University Press, Cambridge.

Vere-Jones, D., 1976. A branching model for crack propagation. Pure and Applied Geophysics 114, 711–726.

Vere-Jones, D., 1978. Earthquake prediction: a statistician's view. Journal of Physical Earth 26, 129–146.

Vere-Jones, D., 1998. Probabilities and information gain for earthquake forecasting. Computational Seismology 30, 249–263.

Vere-Jones, D., Deng, Y.L., 1988. A point process analysis of historical earthquakes from North China. Earthquake Research in China 2, 165–181.

Zhao, Z., Oike, K., Matsumura, K., Ishikawa, Y., 1990. Stress field in the continental part of China derived from temporal variations of seismic activity. Tectonophysics 178, 357–372.

Zheng, X., Vere-Jones, D., 1991. Application of stress release models to historical earthquakes from North China. Pure and Applied Geophysics 135, 559–576.

Zheng, X., Vere-Jones, D., 1994. Further applications of the stochastic stress release model to historical earthquake data. Tectonophysics 229, 101–121.

Zhuang, J., Ma, L., 1998. The stress release model and results from modelling features of some seismic regions in China. Acta Seismologica Sinica 11, 59–70.