doi:10.1007/s11589-011-0796-0

Underground water stress release models*

Yong Li[1] Shenjun Dang[2] Shaochuan Lü[3,♠]

[1] *School of Mathematical Sciences, Beijing Normal University, Beijing* 100875, *China*
[2] *Affiliated School of Tsinghua University, Beijing* 100875, *China*
[3] *School of Mathematical Sciences, Beijing Normal University, Beijing* 100875, *China*

Abstract The accumulation of tectonic stress may cause earthquakes at some epochs. However, in most cases, it leads to crustal deformations. Underground water level is a sensitive indication of the crustal deformations. We incorporate the information of the underground water level into the stress release models (SRM), and obtain the underground water stress release model (USRM). We apply USRM to the earthquakes occurred at Tangshan region. The analysis shows that the underground water stress release model outperforms both Poisson model and stress release model. Monte Carlo simulation shows that the simulated seismicity by USRM is very close to the real seismicity.

Key words: **SRM; underground water data; parameter inference; conditional intensity; AIC**
CLC number: P315.01 **Document code**: A

1 Introduction

The earthquake generation is a complicated dynamical process. Despite that many pre-seismic anomalies or precursors were reported before the occurrence of some large earthquakes, none of them is regarded as a convincing predictor for the occurrence of large earthquakes. Major obstacles of developing trustable prediction algorithm lie in limited observability and lack of information about the motion of subterranean substance. The occurrence of earthquakes is a result of the brittle fracture of the unstable subterranean substance in a critical state. The fracture of subterranean substance can happen anytime when a slight disturbance occurs, making the prediction of the exact occurrence time of an earthquake almost untractable. As a result, probabilistic forecasting based on the stochastic models is crucial for evaluating the risks and uncertainties of forecasting schemes (Aki, 1989).

In 1970s, Vere-Jones (1970, 1976, 1978) proposed some stochastic point process models for the description of the occurrence of earthquakes. These studies are very stimulating and highly regarded in scientific communities, particularly among statisticians and seismologists. One of widely applied stochastic models is the stress release model (SRM), a stochastic version of elastic rebound model, which is viable for the description of seismicity in a relatively large spatial and temporal scale. Further advances in model reformulation and applications have been made by Shi et al. (1998) and Liu et al. (1998).

The controlling role of the stress release model is a process indicating the evolution of regional stress level $X(t)$. The model presumes the loading stress $X(t)$ accumulates linearly over time and release randomly through large quakes, which is written by

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

where $X(0)$ is the initial state of the stress, ρ is the constant loading rate of the stress and $S(t)$ is the accumulative stress release through earthquakes from time 0 up to t. The accumulative release of stress is denoted by

$$S(t) = \sum_{0 \le t_i < t} S_i, \tag{2}$$

where S_i is the stress drop due to the ith earthquake occurring at time t_i . Assuming the stress drop depends only on the magnitude, S_i is usually given by

$$S_i = 10^{\eta(m_i - m_0)}, \tag{3}$$

* **Received** 28 April 2011; **accepted in revised form** 13 July 2011; **published** 10 August 2011.
♠ **Corresponding author. e-mail**: shaochuan.lv@gmail.com

where m_0 is the threshold magnitude. Generally, η varies from 0.75 to 0.9. We take η being 0.75.

The model is specified by the risk function $\psi(x)$ or the conditional intensity function $\lambda(t)$ in terms of the context of point process, which gives the instantaneous occurrence rate of events at an infinitesimal time interval $(t, t+\Delta t)$, i.e.,

$$\lambda(t)\mathrm{dt} = \mathrm{P}\{\text{There are earthquakes occurring in}(\mathrm{t}, \mathrm{t}+\Delta\mathrm{t})\}.$$

For the purpose of inference, simulation and prediction, it is essential to specify the conditional intensity explicitly.

For Poisson model, the occurrence rate of the earthquakes remains a constant, regardless of the stress state $X(t)$. For stress release model, however, the "risk" function is assumed being exponentially increasing with respect to the stress state, i.e.,

$$\psi(x) = A\exp(D + Bx). \tag{4}$$

Then the conditional intensity rate of the SRM is given by

$$\lambda(t) = \psi[X(t)]. \tag{5}$$

From formulae (1–5), we have

$$\lambda(t) = \exp\{a + b[t - cS(t)]\}, \tag{6}$$

where a, b, c are parameters to be estimated by maximum likelihood method. The log-likelihood over time interval [0, T] is given by

$$\log L = \sum_{i=1}^{N(T)} \log\lambda(t_i) - \int_0^T \lambda(t)\mathrm{d}t. \tag{7}$$

SRM assumes the regional stress level is increasing linearly and suddenly drops through large earthquakes. However, the accumulating stress in a region can be released not only by quakes, but also by many other means, such as slow earthquakes, non-seismic slips etc.. Accordingly, these contributing factors should also be considered for investigation of the evolution of regional stress field. Slow earthquakes or static earthquakes, non-seismic slips will generally result in the crust deformation, hence causing changes of the underground water level. Therefore, we incorporate the underground water data into the SRM model, forming the underground water stress release model (USRM).

2 Principles of underground water SRM

The introduction of the underground water level into SRM is not straightforward since the effects of the loading stress on the crust deformation is not a direct indication of the regional stress state, not mentioning other factors such as rainfalls and underground water pumping, which may also cause uprising or declining of the underground water level. Instead of directly introducing the underground water level into SRM, we utilize the absolute value of the instantaneous changing rate of the underground water level as a source of associated information about the crust deformation caused by the regional stress accumulation. Let $v(t)$ denotes the instantaneous changing rate of the underground water level, the overall variation of the underground water level from 0 up to t is given by $\int_0^t |v(s)|\mathrm{d}s$. From equation (1), the regional stress level $X(t)$ is rewritten by

$$X(t) = X(0) + \rho t - S(t) - K\int_0^t |v(s)|\mathrm{d}s, \tag{8}$$

where the last term in the expression is the effect of overall change of underground water level due to the crust deformation. From the above discussion, the conditional intensity of the USRM is given by

$$\begin{aligned}\lambda(t) &= \exp\{X(0) + \rho t - S(t) - K\int_0^t |v(s)|\mathrm{d}s\} \\ &= \exp\{a + b[t - cS(t)] - K\int_0^t |v(s)|\mathrm{d}s\},\end{aligned} \tag{9}$$

where a, b, c and k are parameters to be estimated by maximum likelihood method.

3 Underground water data and analysis

The catalogue used in this analysis includes events occurred between January 1, 1977 and December 31, 2004, within the confine of a circle centered at (118°14′E, 39°41′N). We exclude all the aftershocks out of the catalogue and utilize only the remaining main shocks both in SRM and USRM. Denote the main shocks occurred in this circle with radius R and threshold magnitude M_0 by Cat(R, M_0). It turns out the total number of large earthquakes with magnitude greater than 6 in Richter scale is less than 10. To warrant having sufficient data in the analysis, the magnitude threshold is lowered to at least 5.5. The data used in the analysis are Cat(700, 5.5) and Cat(600, 5.5) with 27 and 23 main shocks in total respectively.

The underground water level records used in this analysis is from No. 2 well in Tangshan, from January 01, 1977 up to December 31, 2004, with 10 227 daily observations. Let ω_i denote the underground water level in ith day and denote $\omega = (\omega_1,\ \omega_2,\ \cdots,\ \omega_{10227})$. For small portion of missing data, we replace them by linear interpolations. The underground water level in the No.2 well of Tangshan is indicated in Figure 1.

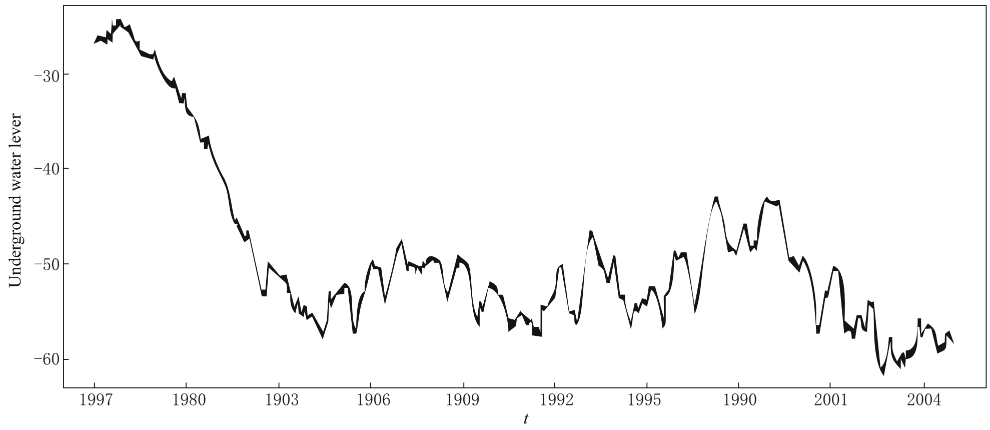

Figure 1 Underground water level (from January 1, 1977 to December 31, 2004).

According to equation (9), the changing rate of the underground water level is a key component of the underground water stress release model. Note that it is the changing rate of the underground water level forming the critical part in USRM rather than the level of the underground water level itself. However, there is no explicit form of $v(t)$ available. We approximate $\int_0^t |v(s)|\mathrm{d}s$ by

$$\int_0^t |v(s)|\mathrm{d}s \approx m\sum\nolimits_{i=1}^{n'} |v_i^{(m)}|, \qquad (10)$$

where n'=min$\{i : i \geq t/m, i \in N\}$ is the number of divided time intervals and $v_i^{(m)}$ denotes the changing rate of the underground water level in the ith time interval, i.e., the tangent of the line obtained by linear regression for m observations in the ith interval.

4 Results and conclusions of underground water stress release model

Let $t_1, t_2, \cdots, t_n$ be the occurrence times of earthquakes over time interval $(0,\ T]$. According to equations (7) and (9), the log-likelihood of the underground water stress release model is given by

$$\log L(a,\ b,\ c,\ k;\ t_1,\ \cdots,\ t_n) = \sum\nolimits_{i=1}^{n}\{a + b[t_i - cS(t_i)] - K\int_0^t |v(s)|\mathrm{d}s\} - \int_0^T \exp\{a + b[x - cS(x)] - K\int_0^x |v(s)|\mathrm{d}s\}\mathrm{d}x. \qquad (11)$$

Given the log-likelihood for the observations, the four unknown parameters are obtained via the method of maximum likelihood estimation.

For obtaining the changing rate of the underground water level, we choose time windows with 10, 20, 30, 60, 90, 120, 150, 180 and 200 observations (days). The optimal window width is determined according to the performance of the models. The following two tables list the estimated parameters, log-likelihoods and AIC values of the three models, i.e. Poisson model, SRM and USRM, for the catalogues Cat(700, 5.5) and Cat(600, 5.5) respectively, where m^* denotes the optimal window width in the fitting.

Table 1 Estimated parameters, likelihoods, AIC values of three models for Cat(700, 5.5)

Model	a	b	c	k	log$L$	AIC
Poisson	−5.937 0	–	–	–	−187.297 6	376.595 3
SRM	−5.799 3	0.000 832	0.017 521	–	−181.770 0	369.540 0
USRM	−19.050 0	0.006 055	0.005 955	0.186 5	−176.294 9	360.589 8

Table 2 Estimated parameters, likelihoods, AIC values of three models for Cat(600,5.5)

Model	a	b	c	k	log$L$	AIC
Poisson	−6.097 3	–	–	–	−163.237 7	328.475 4
SRM	−5.913 8	0.000 885	0.019 766	–	−157.925 2	321.850 4
USRM	−17.751 0	0.005 470	0.006 792	0.168 5	−152.824 1	313.648 2

We compare the performance of models for the two catalogues in terms of AIC. From the tables, it is indicated that AIC values of both SRM and USRM are significantly less than that of Poisson model for each catalogue. Therefore, the SRM and USRM obviously outperform the Poisson model. By comparing SRM with USRM, it is suggested that the AIC value of USRM is also less than that of SRM for both catalogues. In this sense, the USRM gives the best fit to the two catalogues.

Note that the three parameters a, b, c have different interpretations in the three models. In Poisson model, the intensity rate is a constant e^a. For SRM, however, parameters b and c will be positive as it suggests the regional stress level accumulates linearly and release randomly through large earthquakes. The small b is due to the relatively short time interval considered in the model. Similarly, the three parameters b, c, k in USRM are all positive with k greater than 0.03, suggesting that the variation of underground water level is an indication of stress release.

We also show the estimated conditional intensities for three models and show M-T plots for two catalogues in Figures 2 and 3.

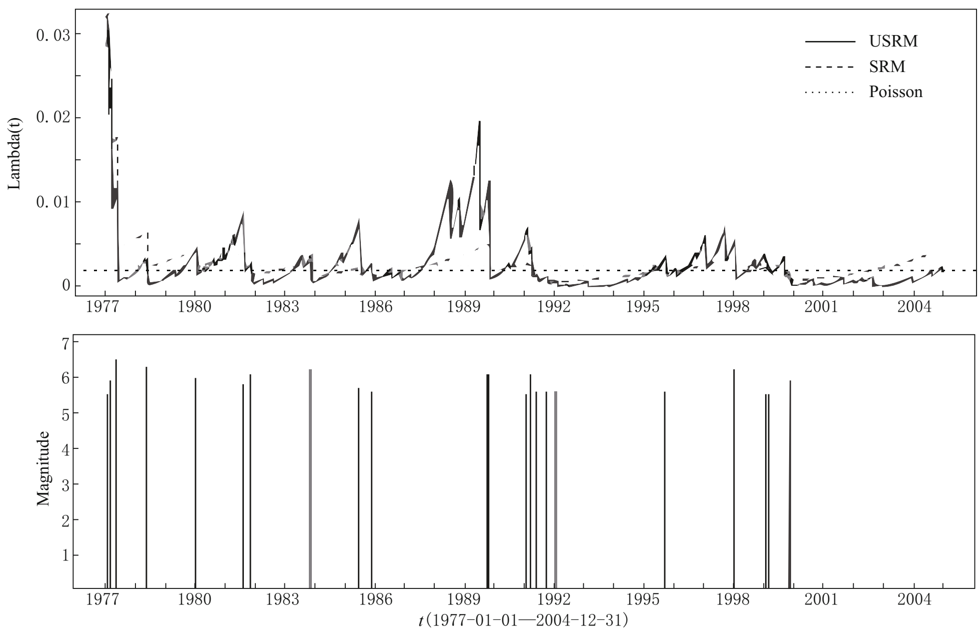

Figure 2 Estimated conditional intensities for three models and M-T plot for Cat(700, 5.5).

From Figures 2 and 3, it is suggested that the intensity curve for Poisson model is a horizontal line, but the intensity functions for SRM and USRM are noncontinuous curves with jumps at the occurrence times of earhquakes. The USRM will give better predictions for the occurrence of large quakes as indicated by the increasing probabilities for the occurrence of large earthquakes at epochs when the estimated intensity rate of USRM is large. We will demonstrate this by Monte Carlo simulations in the following sections.

5 Simulation schemes of USRM

Given a, b, c, k, a series of events can be simulated according to the history of the process (Deng and Liang, 1992). The distribution of the magnitude is determined by G-R law, and written by

$$\log N(M) = a - bM, \tag{12}$$

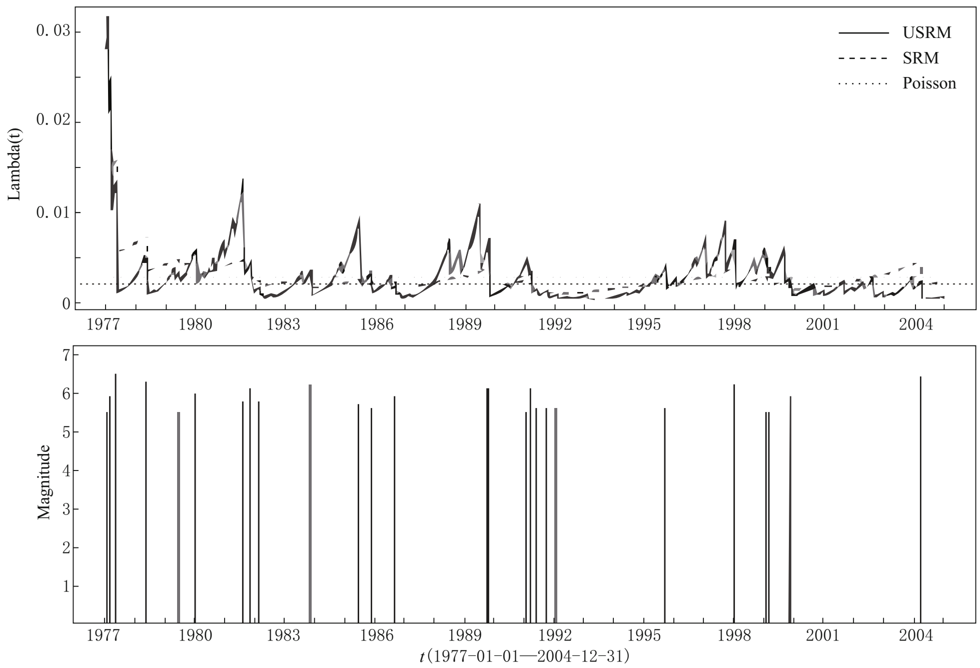

Figure 3 Estimated conditional intensities for three models and M-T plot for Cat (600, 5.5).

where $N(M)$ is the total number of events with magnitude greater than M. Therefore, the probability density function of the magnitude is written by

$$f(m) = \begin{cases} \beta e^{-\beta(m-m_0)}, & m \geq m_0 \\ 0, & \text{others} \end{cases}, \tag{13}$$

where m_0 is the threshold magnitude. The unknown parameter β is estimated by maximum likelihood method. Given N earthquakes with magnitude $m_1, m_2, \cdots, m_N$ the likelihood is

$$L(\beta;\, m_1, \cdots, m_N) = \beta^N e^{-\beta \sum_{i=1}^{N}(m_i - m_0)}. \tag{14}$$

After taking logarithm, the log-likelihood of $m_1, m_2, \cdots, m_N$ is written by

$$\log L(\beta; m_1, \cdots, m_N) = N \log \beta - \beta \sum_{i=1}^{N}(m_i - m_0). \tag{15}$$

The explicit form of MLE for β is therefore written by

$$\hat{\beta} = \frac{N}{\sum_{i=1}^{N}(m_i - m_0)'}. \tag{16}$$

See Aki (1965) for the estimation of β. Given β and the occurrence times of events at $t_1, \cdots, t_n$ in the time interval $[T_1, T_2]$, the magnitude of earthquakes can be simulated as follows

1) evaluate the intensity rate $\lambda(t_i)$ and $\lambda(t_{i-1})$;

2) generate a uniform random number U on the unit interval (0, 1);

3) if $U \leq \lambda(t_i)/\lambda(t_{i-1})$, generate an exponential random number $m' \sim \exp(\beta)$, the magnitude of the ith earthquake is given by $m_i = m_0 + m'$;

4) return to equation (1), replace t_i by t_{i+1} and replace t_{i-1} by t_i, simulate the magnitude of the next earthquake.

We simulate the occurrence of events from January 1, 2004 to December 31, 2004 by using the earthquake catalogue Cat(700, 5.5) and underground water level before 2004. The probability of the occurrence of at least one event with magnitude greater than 5 in 2004 is obtained by Monte Carlo methods. We repeat 10 000 times simulations for the occurrence of events, given the history of the process before 2004. Among the 10 000 simulations, there are 6 721 simulations including at least one event, and for the 6 721 simulations with at least one event, there are 5 268 simulations including exactly just one event and 1 453 simulations including at least two events. From the simulations, it is suggested that the probability of the occurrence of at least an earthquake greater than 5.5 is about 0.672 1 and conditioned on the occurrence of at least an earthquake, the probability of the occurrence of exactly one earthquake is about 0.783 8. In conclusion, the simulation suggests that it is very likely that at least an earthquake with magnitude greater than 5.5 would occur in 2004, but it is unlikely that more earthquakes greater than 5.5 would occur in this period as the seismicity is not very active in 2004. As a matter of fact, there was exactly just one quake greater than 5 occurring in 2004, i.e., the earthquake at Inner Mongolia on March 24.

6 Discussion and conclusions

Stress release model is a statistical model based on a physical model, the elastic rebound model, which is widely used among statisticians and seismologists. We reformulate the original SRM by utilizing the underground water level and obtaining a new class of SRM, i.e. the USRM. The main contributions of this paper is as follows.

1) We suggest the information of underground water level can be incorporated into the conditional intensity function of the SRM , forming the so called USRM and provide the methods for parametric inference and simulation of the model;

2) The USRM is applicable for the catalogue data within the confine of a circle centered at Tangshan with magnitude greater than 5.5. The performance of the model is better than both SRM and Poisson model in terms of some information theoretical criterion such as AIC;

3) In dealing with the underground water level data, we introduce some new conceptions such as the local changing rate and window size of the observations;

However, more detailed analysis should be carried out in future studies.

1) According to the fitting results and the estimated conditional intensity function as shown in Figures 2 and 3, the size of the variations of the stress is notable, which might be caused by heavy rainfall, drought or other factors. Li et al. (2001) discussed the effect of rainfall on the underground water level and suggested a method to rule out the effect of rainfall on the underground water level. The model might be improved by the method.

2) In our analysis, R in $\mathrm{Cat}(R, M_0)$ is relatively large since only one well is used as the source of the information of underground water level. The model might be improved if we combine the data of several wells together when using the USRM.

3) For the stress release model, it is sensible to divide the region into several sub-regions in terms of the tectonic environment. However, if a refined division of the region is carried out, the available records of the underground water level (less than 30 years) and large earthquakes might be in short for a detailed statistical analysis. Zhuang and Ma (1998) suggested that the North China district can be divided into four sub-regions when dealing with the earthquake catalogue from North China region, which might be also feasible when applying USRM.

4) Except for the underground water level data, other factors might also be manifestations of the accumulation of regional stress. Therefore, the model could be further improved by incorporation of other related factors into USRM.

5) The validity of current model in other region should be further investigated. Furthermore, the uncertainty of the parameter estimates of the USRM is not discussed in this paper, but as Wang et al. (2010) discussed, it is not small for some stochastic point process models and can not be ignored. The influence of the uncertainty will be considered in the further study. Moreover, retrospective test (e.g., Kagan et al., 2007, Wang et al., 2011) is useful to evaluate the prediction power of model when we have more data available.

Acknowledgements This research is financial supported by the National Natural Science Foundation of China under the grant No. 10871026. Two referees suggestions are acknowledged.

References

Akaike H (1977). On entropy maximization principle. In: Krishnaian P R ed. *Application of Statistics*. North Holland, Amsterdam, 27–41.

Aki K (1965). Maximum likelihood estimation of b in the formula $\log N = a - bM$ and its confidence limits. *Bull Earthq Res Inst Tokyo Univ* **43**: 237–239.

Aki K (1989). Ideal probabilistic earthquake prediction. *Tectono Physics* **169**: 197–198.

Deng Y and Liang Z (1992). *Stochastic Point Processes and Their Applications*. Science Press, Beijing, (in Chinese).

Kagan Y Y, Jackson D D and Rong Y F (2007). A testable five-year forecast of moderate and large earthquakes in southern California based on smoothed seismicity. *Seism Res Lett* **78**(1): 94–98.

Knopoff L (1971). A stochastic model for occurrence of main-sequence earthquake. *Rev Geophys Space Phys* **9**: 175–188.

Li Y, Yang X and Ma L (2001). A rudimentary study of the associations between underground water level and seisimicity. *Journal of Beijing Normal University* (*Natural Science Version*) **37**: 739–745 (in Chinese with English abstract).

Liu J, Vere-Jones D, Ma L, Shi Y and Zhuang J (1998). The coupling stress release models and its applications.

J Seismol **20**: 232–239 (in Chinese with English abstract).

Shi Y, Liu J, Zhuang J and Ma L (1998). Application of mechanical and statistical models to study of seismicity and prediction of synthetic earthquakes. *Acta Seismologica Sinica* **11**: 421–430.

Vere-Jones D (1970). Stochastic models for earthquake occurrence (with discuss). *J Roy Stat Soc B* **32**: 1–62.

Vere-Jones D (1976). A branching model for crack propagation. *Pure Appl Geophys* **114**: 711–726.

Vere-Jones D (1978). Earthquake prediction: A statistics view. *J Phys Earth* **26**: 129–146.

Wang Q, Schoenberg F and Jackson D D (2010). Standard errors of parameter estimates in the ETAS model. *Bull Seismol Soc Amer* **100**(5A): 1 989–2 001, doi:10.1785/0120100001.

Wang Q, Jackson D D and Kagan Y Y (2011). California earthquake forecasts based on smoothed seismicity: Model choices. *Bull Seismol Soc Amer* **101**: 1 422–1 430.

Zhuang J and Ma L (1998). Stress release models and simulations of earthquakes in several seismic active zones. *J Seismol* **20**: 48–57 (in Chinese with English abstract).