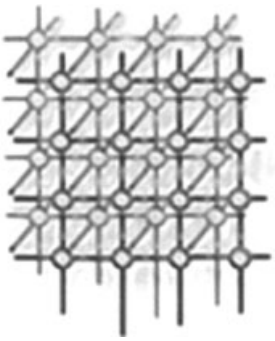
CONCURRENCY AND COMPUTATION: PRACTICE AND EXPERIENCE
Concurrency Computat.: Pract. Exper. 2010; **22**:1559–1568
Published online 17 November 2009 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/cpe.1515



Study on the forecast effects of PI method to the north and southwest China

Yongxian Zhang[1,2], Xiaotao Zhang[3], Xiangchu Yin[2,3] and Yongjia Wu[1,*,†]

[1]*China Earthquake Networks Center, Beijing 100045, China*
[2]*LNM, Institute of Mechanics, Chinese Academy of Sciences, Beijing 100080, China*
[3]*Institute of Earthquake Science, China Earthquake Administration, Beijing 100036, China*

SUMMARY

In this paper, the codes of Pattern Informatics (PI) method put forward by Rundle *et al.* have been worked out according to their algorithm published, and the retrospective forecast of PI method to North China (28.0°–42.0° N, 108.0°–125.0° E) and to Southwest China (22.0°–28.3° N, 98.0°–106.0° E) has been tested. The results show that the hit rates in different regions show a great difference. In Southwest China, 32 earthquakes with M_L5.0 or larger have occurred during the predicted time period 2000–2007, and 26 out of the 32 earthquakes occurred in or near the hot spots. In North China, the total number of M_L5.0 or larger was 12 during the predicted time period 2000–2007, and only 3 out of the 12 earthquakes occurred in or near the hot spots. From our results, we hold that if the PI method could be applied to all kinds of regions, the parameters associated with time points and time windows should be chosen carefully to obtain the higher hit rate. We also found that the aftershocks in a strong earthquake sequence affect the PI results obviously. Copyright © 2009 John Wiley & Sons, Ltd.

Received 25 October 2008; Accepted 24 July 2009

KEY WORDS: PI method; retrospective forecast test; north China; southwest China

1. INTRODUCTION

Pattern informatics (PI) method is one of the new approaches to earthquake forecasting in 10 year scale, which is recently introduced by Rundle *et al.* [1–4] and defined in mathematical terms and

*Correspondence to: Yongjia Wu, China Earthquake Networks Center, Beijing 100045, China.
†E-mail: wuyongj@mails.gucas.ac.cn

Contract/grant sponsor: Earthquake Joint Funds of China Earthquake Administration
Contract/grant sponsor: NSFC; contract/grant numbers: 10232050, 10572140

provided a rational explanation for each step of the process by Tiampo *et al.* [5,6]. This method is proved to be much better than a simple measure of past seismicity. Holliday *et al.* performed a systematic analysis of the procedure and found optimal choices for the southern California region by varying the ordering of the steps and the parameter values [7–9].

The PI method is based on the statistical mechanics of complex systems and can quantify temporal variations in seismicity. Although it could not give short-term predictions of future earthquakes, it does reduce the areas of earthquake risk relative to those given by long-term hazard assessments. The result is a map of areas in a seismogenic region (hot spots) where earthquakes are likely to occur during a specified period in the future. A forecast map of locations (hot spots) of future $M>5$ earthquakes for California in the period of 2000–2010 was published in 2002 [3] (http://quakesim.jpl.nasa.gov/scorecard.html). Subsequently, 19 of the 20 California earthquakes with magnitudes $M>5$ occurred in or immediately adjacent to the resulting hot spots till February 2008, while the areas of the hot spots only cover 4% of the map area (*Performance Analysis of Earthquake Forecasts*, presentation of Rundles *et al.* on *the 6th ACES International Workshop Cairns*, Australia, 11–16 May 2008). Nanjo *et al.* modified the PI method for use with the Japanese catalogs and successfully forecast the 23 October 2004 $M=6.8$ Niigata earthquake [10]. Chen *et al.* modified the PI method for use with Chinese Taiwan catalogs [11] and found the Chi Chi Ms7.6 earthquake located in the hot spot area.

How about the effects of PI method when applied to China mainland? To test the validity of this method for continental earthquakes, Jiang and Wu [12] studied the PI map with codes from Chen in Sichuan-Yunnan region of China and made retrospective forecast test for earthquakes with Ms $\geq$ 5.5. Their results show that the PI forecast outperforms not only random forecast but also the simple number counting approach based on the clustering hypothesis of earthquakes. They also found that if the 'forecast time window' was shortened to 3 years, the forecast capability of the PI model decreased significantly, albeit outperforming random forecast.

The objective of this paper is to apply the PI method to different regions in China mainland to see if there exists significant discrepancy in the forecasting ability in different regions. We choose north China and southwest China as the retrospective forecasting regions in this paper.

2. ABOUT THE CODES AND ALGORITHM OF PI METHOD

Following the detailed steps of the PI method in the literature [7], we write codes in Fortran language, which can obtain hot spots map of the PI method in the region of interest. The detailed utilization of the PI method for earthquake forecasting is as follows [7]:

(1) The region of interest is divided into N_B square boxes with linear dimension Δx. Boxes are identified by a subscript i and are centered at x_i. For each box, there is a time series $N_i(t)$, which is the number of earthquakes per unit time at time t larger than the lower cut-off magnitude M_c. The time series in box i is defined between a base time t_b and the present time t.
(2) All earthquakes in the region of interest with magnitudes greater than a lower cutoff magnitude M_c are included. The lower cutoff magnitude M_c is specified in order to ensure completeness of the data through time, from an initial time t_0 to a final time t_2.

DOI: 10.1002/cpe

(3) Three time intervals are considered:
   (a) A reference time interval from t_b to t_1.
   (b) A second time interval from t_b to t_2, $t_2 > t_1$. The change interval over which seismic activity changes are determined is then $t_2 - t_1$. The time t_b is chosen to lie between t_0 and t_1. The objective is to quantify anomalous seismic activity in the change interval t_1 to t_2 relative to the reference interval t_b to t_1.
   (c) The forecast time interval t_2 to t_3, for which the forecast is valid. We take the change and forecast intervals to have the same length.

(4) The seismic intensity in box i, $I_i(t_b, t)$, between two times $t_b < t$, can then be defined as the average number of earthquakes with magnitudes greater than M_c that occur in the box per unit time during the specified time interval t_b to t. Therefore, using discrete notation, we can write:

$$I_i(t_b, t) = \frac{1}{t - t_b} \sum_{t' = t_b}^{t} N_i(t')$$

where the sum is performed over increments of the time series, say days.

(5) In order to compare the intensities from different time intervals, we require that they have the same statistical properties. We therefore normalize the seismic intensities by subtracting the mean seismic activity of all boxes and dividing by the standard deviation of the seismic activity in all boxes. The statistically normalized seismic intensity of box i during the time interval t_b to t is then defined by

$$\hat{I}_i(t_b, t) = \frac{I_i(t_b, t) - \langle I_i(t_b, t) \rangle}{\sigma(t_b, t)}$$

where $\langle I_i(t_b, t) \rangle$ is the mean intensity averaged over all the boxes and $\sigma(t_b, t)$ is the standard deviation of intensity over all the boxes.

(6) Our measure of anomalous seismicity in box i is the difference between the two normalized seismic intensities:

$$\Delta I_i(t_b, t_1, t_2) = \hat{I}_i(t_b, t_2) - \hat{I}_i(t_b, t_1)$$

(7) To reduce the relative importance of random fluctuations (noise) in seismic activity, we compute the average change in intensity, $\Delta I_i(t_0, t_1, t_2)$ over all possible pairs of normalized intensity maps having the same change interval:

$$\overline{\Delta I_i(t_0, t_1, t_2)} = \frac{1}{t_1 - t_0} \sum_{t_b = t_0}^{t_1} \Delta I_i(t_b, t_1, t_2)$$

where the sum is performed over increments of the time series, which here are days.

(8) We define the probability of a future earthquake in box i, $P_i(t_0, t_1, t_2)$, as the square of the average intensity change:

$$P_i(t_0, t_1, t_2) = \overline{\Delta I_i(t_b, t_1, t_2)}^2$$

DOI: 10.1002/cpe

(9) To identify anomalous regions, we wish to compute the change in the probability $P_i(t_0, t_1, t_2)$, relative to the background so that we subtract the mean probability over all boxes. We denote this change in the probability by

$$\Delta P_i(t_0, t_1, t_2) = P_i(t_0, t_1, t_2) - \langle P_i(t_0, t_1, t_2)\rangle$$

where $\langle P_i(t_0, t_1, t_2)\rangle$ is the background probability for a large earthquake.

Hot spots are defined to be the regions where $\Delta P_i(t_0, t_1, t_2)$ is positive. In these regions, $P_i(t_0, t_1, t_2)$ is larger than the average value for all boxes (the background level). Note that since the intensities are squared in defining probabilities the hot spots may be due to either the increases of seismic activity during the change time interval (activation) or due to the decreases (quiescence). We hypothesize that earthquakes with magnitudes larger than $M_c + 2$ will occur preferentially in hot spots during the forecast time interval t_2 to t_3.

3. CHECK OF OUR CODES

In order to make sure that if our codes work correctly, we calculated the hot spot map of the California region with the same parameters chosen in the literature [7].

First, we had downloaded earthquake catalogs since 1932 from the web page of 'http://www.data.scec.org/ftp/catalogs/SCSN/'. Here, we have named it as the 'Southern California catalog'. Meanwhile, we had downloaded earthquake catalogs since 1932 from the web page of 'http://www.ncedc.org/ncedc/catalog-search.html'. Here, we have named it as the 'Northern California catalog'. The

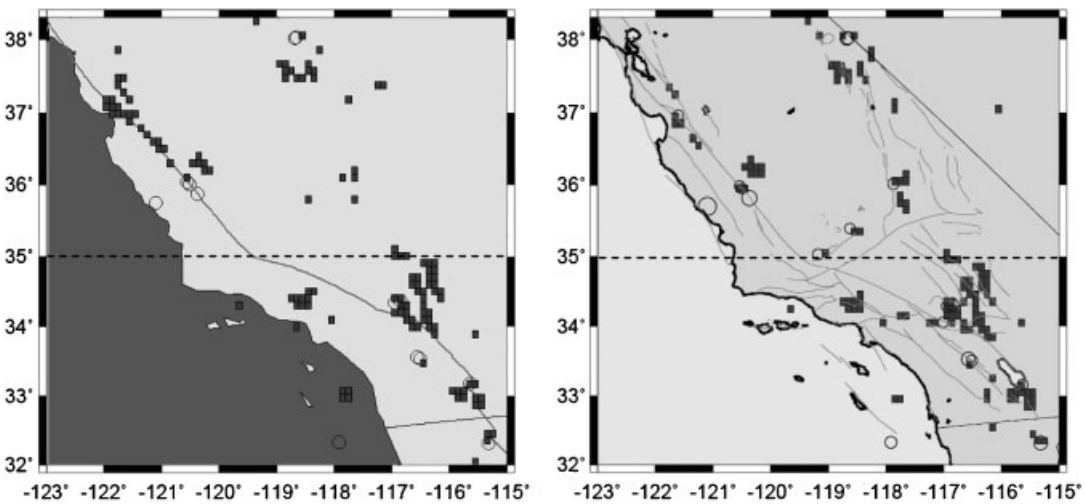


Figure 1. Comparison of our results (left) with those by Holliday *et al.* [7] (right) ($t_0 = 1932$, $t_1 = 1990$, $t_2 = 2000$, $t_3 = 2010$).

DOI: 10.1002/cpe

catalogue used in our research work is a combined one with earthquakes in the area of (32–35°N, 115–123°W) from the 'Southern California catalog' and earthquakes in the area of (35–38.3°N, 115–123°W) from the 'Northern California catalog'.

Second, the parameters in the PI computation are chosen as follows: The initial time was 1932/01/01, the change interval was from 1989/12/31 to 1999/12/31, and the forecast interval was from 1999/12/31 to 2010/01/01. The region of interest (32–38.3°N, 115–123°W) was divided into 5040 boxes. The lower magnitude cutoff was taken to be 3.0.

The forecast map obtained by our program is shown in Figure 1 (left). The results are almost in consistence with those of Holiday *et al.* [7] (Figure 1, right), and a little difference between the two maps may be caused by different earthquake catalogs.

4. DATA AND COMPUTING PARAMETERS CHOSEN

North China (28.0°–42.0° N, 108.0°–125.0° E) and Southwest China (22.0°–28.3° N, 98.0° –106.0° E) are rich in large earthquakes (Figure 2).

The amazing Tangshan $M7.8$ earthquake (39.41°N, 118.0° E) which killed more than 240 000 lives in 1978 occurred in North China, and the tragic Wenchuan $M8.0$ earthquake (31.0° N, 103.4° E) which killed more than 70 000 lives in 2008 occurred in Southwest China.

We divide the seismogenic region to be studied into a grid of square boxes with the size of $0.1° \times 0.1°$, this size is related to the magnitude of $M5$ earthquakes to be forecasted; hence, if PI anomaly occurs in a region with a chain of square boxes, a larger earthquake with $M>5.0$ could be predicted. For Southwest China, the total number of boxes is 5040. For North China, the total number of boxes is 23 800.

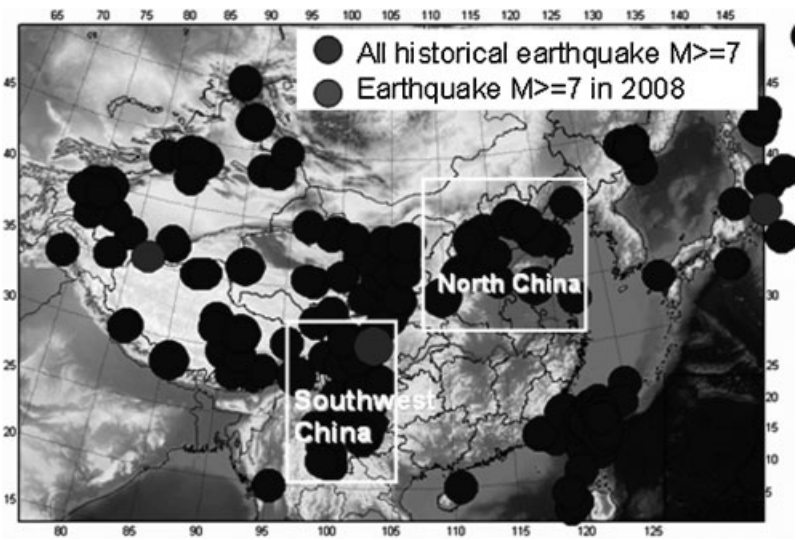


Figure 2. Abridged general view of the locations of the two regions studied in this paper.

DOI: 10.1002/cpe

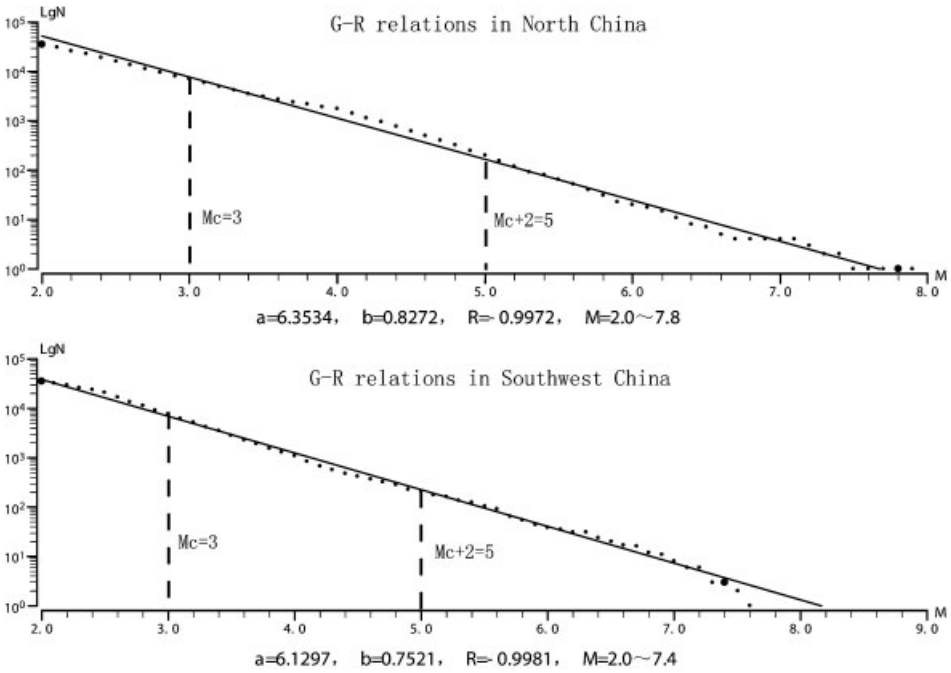


Figure 3. Completeness of catalogue test by G-R relation in north China and southwest China.

The earthquake catalogue is from 1970. According to the seismic monitoring ability in North China and Southwest China, the lower cutoff magnitude M_c could be chosen as 3.0 to ensure completeness of the data from an initial time $t_0 = 1970$ to a final time $t_2 = 1999$ (Figure 3).

We chose the beginning time $t_b = 1973$, and the reference time interval from t_b to $t_1 = 1991$. The forecast time interval is t_2 to t_3, here $t_3 = 2007$.

5. RESULTS

Generally, hot spots are defined as the regions where $\Delta P_i(t_0, t_1, t_2)$ is positive. Our result shows that there are more hot spots when the threshold of $\Delta P_i(t_0, t_1, t_2)$ is lower. In order to raise the hitting rate and reduce the missing rate, we have to make the decision threshold of possibility gain $\Delta P_i(t_0, t_1, t_2)$. After trying different thresholds of $\Delta P_i(t_0, t_1, t_2)$, we found a good fitness of $\Delta P_i(t_0, t_1, t_2)$, under which the hit rate is relatively higher and the miss rate is relatively lower.

Following the literature [7], we define that: during t_2 to t_3, if an earthquake occurs in a hotspot box or within the Moore neighborhood of the box, this is a success (the eight boxes surrounding the hotspot box are defined as 'the Moore neighborhood' [13]); If no earthquake occurs in a non-hotspot box, this is also a success; If no earthquake occurs either in a hotspot box or within the Moore neighborhood of the hotspot box, this is a false alarm; If an earthquake occurs in a box,

DOI: 10.1002/cpe

which is neither the hotspot box nor the Moore neighborhood of the hotspot box, this is a failure to forecast.

According to the above definitions, values a(Forecast = yes, Observed = yes), b(Forecast = yes, Observed = no), c(Forecast = no, Observed = yes), and d(Forecast = no, Observed = no) are obtained for the hot spot map. The fraction of colored boxes, also called the probability of forecast of occurrence, is $r = (a + b)/N$, where the total number of boxes is $N = a + b + c + d$. The hit rate is $H = a/(a + c)$ and is the fraction of large earthquakes that occur on a hot spot. The false alarm rate is $F = b/(b + d)$ and is the fraction of non-observed earthquakes that are incorrectly forecast.

5.1. Hot spot map of southwest China

When we take the threshold possibility as $\log_{10}(\Delta P_i(t_0, t_1, t_2)/\langle P_i(t_0, t_1, t_2)\rangle) = -3.2$, the best good fitness hot spot map could be obtained as shown in Figure 4.

The values of a, b, c, d are listed in Table I. From this table we can see that there are 32 earthquakes of $M \geq 5.0$ that occurred during the forecasting period January 2000–2007. Twenty-six out of the 32 earthquakes occurred in or near the hot spots; hence, the hitting rate $H = a/(a+c) = 26/32 = 0.813$, and the false alarm rate $F = b/(b + d) = \frac{140}{5008} = 0.028$.

Comparing with hot spot map of California region in the literature [7], the hitting rate for California region is $H = \frac{23}{32} = 0.719$, and the false alarm rate $F = \frac{104}{5008} = 0.021$. It seems that the forecasting effect of the PI method in Southwest China is better than that in the California region.

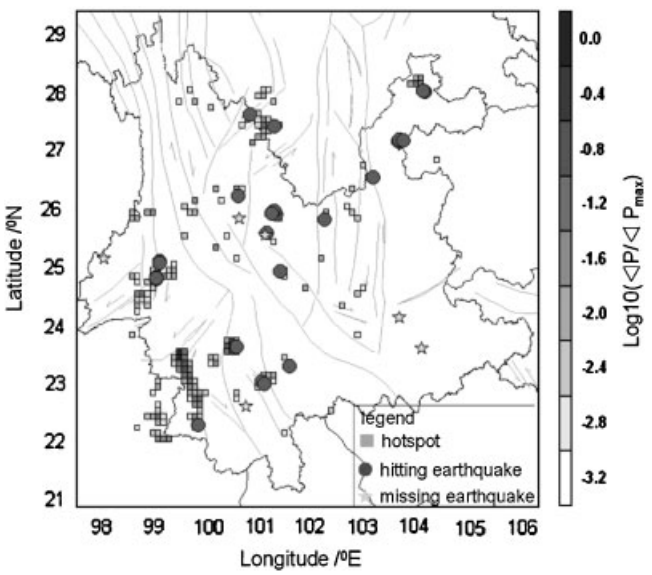


Figure 4. Hot spot map in southwest China under the threshold possibility $\log_{10}(\Delta P_i(t_0, t_1, t_2)/\langle P_i(t_0, t_1, t_2)\rangle) = -3.2$.

DOI: 10.1002/cpe

Table I. Numbers of boxes counted for check of PI method effects in southwest China.

	Observed	
Forecast	Yes	No
Yes	a (26)	b (140)
No	c (6)	d (4868)

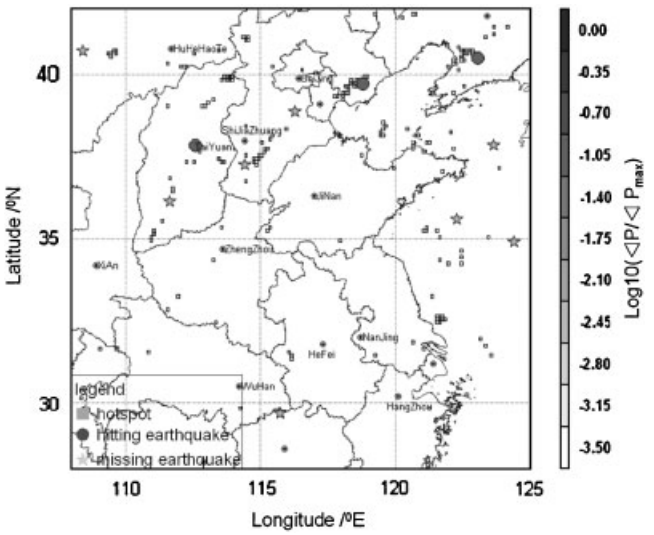


Figure 5. Hot spot map in north China under the threshold possibility $\log_{10}(\Delta P_i(t_0, t_1, t_2)/\langle P_i(t_0, t_1, t_2)\rangle) = -3.5$.

5.2. Hot spot map of north China

When we take the threshold possibility as $\log_{10}(\Delta P_i(t_0, t_1, t_2)/\langle P_i(t_0, t_1, t_2)\rangle) = -3.5$, the best good fitness hot spot map could be obtained as shown in Figure 5.

The values of a, b, c, d are listed in Table II. From this table we can see that there are 12 earthquakes of $M \geq 5.0$ occurred during the forecasting period January 2000–2007. Three out of the 12 earthquakes occurred in or near the hot spots; hence, the hitting rate $H = a/(a+c) = \frac{3}{12} = 0.25$, and the false alarm rate $F = b/(b+d) = \frac{183}{23\,794} = 0.008$.

Comparing with the hot spot map of the California region in the literature [7] and the above hot spot map of southwest China, the hitting rate for north China is much lower, although the false rate is much lower, too. However, no matter in southwest China, California, or north China, the forecasting effect of the PI method outperforms a random map greatly. According to the literature [7], the forecasting effect of the PI method outperforms RI (relative intensity) method generally.

DOI: 10.1002/cpe

Table II. Check of forecasting effects of PI method in southwest China.

	Observed	
Forecast	Yes	No
Yes	a (3)	b (183)
No	c (9)	d (23 611)

6. CONCLUSIONS AND DISCUSSIONS

The hit rates in different regions show a great difference. In Southwest China, 32 earthquakes with M_L5.0 or larger have occurred during the time period 2000–2007, and 26 out of the 32 earthquakes occurred in or near the hot spots. In North China, the total number of M_L5.0 or larger is 12 during the time period 2000–2007, and only 3 out of the 12 earthquakes occurred in or near the hot spots. It seems obviously that the PI method give better forecast in Southwest China than in north China.

The PI method is really an optimal method for earthquake forecast in a time scale of 10 years, however, when this method is applied to different regions, the effects show a great difference. For this method, the factors that affect the final results include the time parameters t_0, t_1, t_2, and t_b, and space parameters of scale of the studied region, scale of boxes, and the thresholds of lower cutoff of magnitude, lower cutoff of possibility. In order to obtain the higher hit rate, the suitable study region, suitable time points t_1, t_2 and t_3, suitable thresholds of lower cutoff of magnitude, suitable threshold of $\log 10(\Delta P/\Delta P \max)$, etc. should be chosen after numerous tries.

In our study, we also found that the after shocks in a strong earthquake sequence affect the hot spot map seriously, the effects of aftershock sequence on the PI method is to be studied next.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support from the Earthquake Joint Funds of China earthquake Administration and NSFC under grant No. 10232050, No. 10572140. Prof. J. B. Rundle gave us some comments on PI methods during the 6th ACES International workshop Cairns, Australia 11–16 May 2008. We also show our thanks to CENC(China Earthquake Networks Center) for earthquake data.

REFERENCES

1. Rundle JB, Klein W, Gross SJ, Tiampo KF. Dynamics of seismicity patterns in systems of earthquake faults. *GeoComplexity and the Physics of Earthquakes* (*Geophysical Monograph Series*, vol. 120), Rundle JB, Turcotte DL, Klein WW (eds.). AGU: Washington, DC, 2000; 127–146.
2. Rundle JB, Klein W, Tiampo KF, Gross SJ. Linear pattern dynamics in nonlinear threshold systems. *Physical Review E* 2000; **61**:2418–2432.
3. Rundle JB, Tiampo KF, Klein W, Martins JSS. Self-organization in leaky threshold systems: The influence of near-mean field dynamics and its implications for earthquakes, neuro-biology, and forecasting. *Proceedings of the National Academy of Sciences*, *U.S.A.* 2002; **99**(Suppl. 1):2514–2521.
4. Rundle JB, Turcotte DL, Shcherbakov R, Klein W, Sammis C. Statistical physics approach to understanding the multiscale dynamics of earthquake fault systems. *Reviews of Geophysics* 2003; **41**:1019–1038.
5. Tiampo KF, Rundle JB, McGinnis S, Gross SJ, Klein W. Eigenpatterns in southern California seismicity. *Journal of Geophysical Research* 2002; **107**:2354.

DOI: 10.1002/cpe

6. Tiampo KF, Rundle JB, McGinnis S, Klein W. Pattern dynamics and forecast methods in seismically active regions. *Pure and Applied Geophysics* 2002; **159**:2429–2467.
7. Holliday JR, Nanjo KZ, Tiampo KF, Rundle JB, Turcotte DL. Earthquake forecasting and its verification. *Nonlinear Processes in Geophysics* 2005; **12**:965–977.
8. Holliday JR, Rundle JB, Tiampo KF, Klein W, Donnellan A. Systematic procedural and sensitivity analysis of the pattern informatics method for forecasting large ($M \geq 5$) earthquake events in southern California. *Pure and Applied Geophysics* 2006; **163**:2433–2454.
9. Holliday JR, Rundle JB, Tiampo KF, Klein W, Donnellan A. Modification of the pattern informatics method for forecasting large earthquake events using complex eigenfactors. *Tectonophysics* 2006; **413**:87–91.
10. Nanjo KZ, Rundle JB, Holliday JR, Turcotte DL. Pattern informatics and its application for optimal forecasting of larger earthquakes in Japan. *Pure and Applied Geophysics* 2006; **163**:2417–2432.
11. Chen CC, Rundle JB, Holliday JR, Nanjo KZ, Turcotte DL, Li SC, Tiampo KF. The 1999 Chi-Chi, Taiwan, earthquake as a typical example of seismic activation and quiescence. *Geophysical Research Letters* 2005; **32**:L22315.
12. Jiang CS, Wu ZL. Retrospective forecasting test of a statistical physics model for earthquakes in Sichuan–Yunnan region. *Science in China Series D*: *Earth Sciences* 2008; **51**(10):1401–1410. DOI: 10.1007/s11430-008-0112-6.
13. Moore EF. Machine models of self reproduction. *Proceedings of the Fourteenth Symposium on Applied Mathematics*, American Mathematical Society, 1962; 17–33.