Study on the Earthquake Potential Regions in North and Northeast China by Pattern Informatics Method

Xiaotao Zhang
Institute of Earthquake Science
CEA
Beijing, China

Yongxian Zhang
Department of Earthquake Prediction
CENC
Beijing, China

Xiangchu Yin
LNM, Institute of Mechanics
CAS
Beijing, China

Abstract—**The Pattern Informatics (PI) method is a new approach to forecast earthquake based on statistical physics. In general, the PI method works for long-term earthquake forecast, but we can obtain the higher spatial resolution of abnormal regions by it. In this paper, we applied the PI method to study seismic risk assessment in north and northeast China. The research suggests that there might be earthquakes with $M \geq 7$ in the juncture of Shanxi, Henan and Shanxi provinces, the juncture of Hebei, Shandong and Henan provinces, and southeast coast of Jiangsu province from Jan. 2009 to Jan. 2050.**

Keywords-pattern informatics method; earthquake; forecasting; relative operating characteristic (ROC) diagram

I. INTRODUCTION

The Pattern Informatics (PI) method was first put forward by Rundle in 2000, and it was a new approach to study of seismicity [1], [2]. The fundamental principles of PI method are given as follows:

The region of interest is divided into many square boxes. For each box, there is a time series, which is the instantaneous number of earthquakes per unit time. Then, the abnormal change of seismic active intensity is extracted from each box as well as the changes of seismic active intensity were normalized. After that, the probability of occurrences of significant earthquakes in each box is computed and then subtracts their corresponding background probability. At last, we extract these boxes (namely, "seismic hotspots" or "hotspots"), which have high probability of earthquake occurrence. In general, the PI method works for long-term earthquake forecast (about 10 years or more), and it can give the higher spatial resolution of abnormal regions (for example, the spatial resolution of abnormal regions is $110km \times 110km$, which is obtained by PI method for earthquake with $M \geq 7$).

In this paper, we introduced the step process of PI method, and illustrated the test standard and evaluation method (ROC diagram) of PI method forecasting effect. Then, we applied the PI method to study seismic risk assessment in north and northeast China. In the study, the research approaches are given as follows:

Firstly, we choose initial values to some parameters, such as the range of study region, the initial time (t_0) of seismic data, the cutoff of the magnitude (M_c), the forecast time interval ($t_2 \sim t_3$) and so on. Secondly, we determine the values of other parameters, such as the scale of box (Δx), the threshold (w) and so on, by multiple retrospective forecasting. Thirdly, we single out the optimum parameter values for the region of interest. At last, we study earthquake risk assessment in the region of interest in future based on the optimum parameters.

II. THE PATTERN INFORMATICS METHOD

Since the PI method is a new approach to study of seismicity, the step process of it, the test standard and evaluation method (ROC diagram) of its forecasting effect are elucidated in details.

A. The Step Process of Pattern Informatics Method

The detailed procedures of PI method are described as follows [3]–[5]:

1) The time series $N_i(t)$: The geographic area of interest is partitioned into N_b square boxes centered on a point x_i and with an edge length Δx . For each box, a time series $N_i(t)$ is defined by counting how many earthquakes with magnitude greater than M_c . The time series in box i is defined between a base time t_b and the present time t .

2) Three time intervals are defined: (Fig. 1)

- The reference time interval from t_b to t_1 ($t_0 < t_b < t_1$).

- The change time interval from t_1 to t_2 .

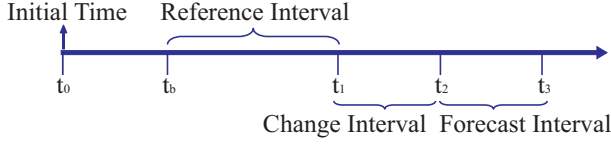- The forecast time interval from t_2 to t_3 , in which the forecast is valid.

Figure 1. Schematic diagram of time intervals

The time t_0 is the initial time, at which the calculation begins, data before t_0 would not be used in PI calculation. The time t_b is shifted between t_0 and t_1, and the time step is Δt. We take the change and forecast interval to have the same length.

3) The seismic intensity function $I_i(t_b,t)$: Between t_b and t ($t_b < t$), the seismic intensity function is defined as the average number of earthquakes with magnitudes greater than M_c that occur in box x_i per unit time. That is,

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

4) Normalizing the seismic intensity function: Seismic intensity is normalized in time by subtracting the spatial mean for all boxes ($< I_i(t_b,t) >$) and dividing by the spatial standard deviation ($\sigma(t_b,t)$)

$$ \hat{I}_i(t_b,t) = \frac{I_i(t_b,t) - < I_i(t_b,t) >}{\sigma(t_b,t)} \qquad (2) $$

5) The anomalous seismicity: In box x_i, the anomalous seismicity is discribed by the differential of normalized mean value between time period of t_b to t_2 and t_b to t_1.

$$ \Delta I_i(t_b,t_1,t_2) = \hat{I}_i(t_b,t_2) - \hat{I}_i(t_b,t_1) \qquad (3) $$

6) Calculating the probability of a future earthquake: The probability of a strong earthquake occurring within a box x_i is defined as this:

$$ P_i(t_0,t_1,t_2) = \overline{\Delta I_i(t_b,t_1,t_2)}^2 \qquad (4) $$

Where,

$$ \overline{\Delta I_i(t_b,t_1,t_2)} = \frac{1}{t_b - t_1} \sum_{t_b=t_0}^{t_1} \Delta I_i(t_b,t_1,t_2) \qquad (5) $$

7) The probability function: Finally, to identify anomalous regions, the new probability function $\Delta P_i(t_0,t_1,t_2)$ is defined by difference between $P_i(t_0,t_1,t_2)$ and its spatial mean value $< P_i(t_0,t_1,t_2) >$.

$$ \Delta P_i(t_0,t_1,t_2) = P_i(t_0,t_1,t_2) - < P_i(t_0,t_1,t_2) > \qquad (6) $$

The "hotspots" are defined to be the boxes (or the regions) where $\Delta P_i(t_0,t_1,t_2)$ is positive. We hypothesize that earthquakes
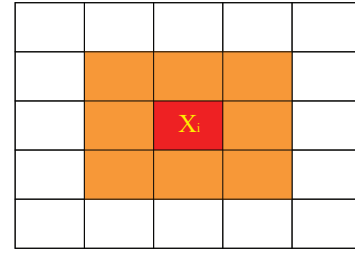

Figure 2. The hotspot box x_i and its Moore neighborhood

with magnitudes larger than $M_c + 2$ will occur in hotspots during the forecast time interval t_2 to t_3.

In this paper, we also introduce the parameter of threshold (w), whose value is determined by $\log_{10}(\Delta P/\Delta P_{max})$, where ΔP is the probability of a future strong earthquake in box x_i, and ΔP_{max} is the maximum of all probabilities. The number of hotspots is determined by the value of threshold (w). And, according to actual need or the retrospective result, we determined the value of w.

B. The Test Standard of Pattern Informatics Method Forecasting Effect

We utilize the rules from the previous work [2]–[4] as the test standard of PI forecasting effect.

- If earthquake occurs in a hotspot box or within the neighborhood of the box, this is a success. The eight boxes surrounding the hotspot box x_i are defined as "the Moore neighborhood" [6] (Fig. 2).

- If no earthquake occurs in a non-hotspot box, this is also a success.

- If no earthquake occurs in a hotspot box or within the Moore neighborhood of the hotspot box, this is a false alarm.

- If earthquake occurs in a box, which is not hotspot box or the Moore neighborhood of the hotspot box, this is a failure to forecast.

C. The Evaluation Method of Pattern Informatics Method Forecasting Effect

At present, there are many methods to evaluate PI method forecasting effect. Such as: a) R-value evaluation method [7], b) Q-value evaluation method [8], c) Bootstrap technique [9], d) Molchan map [10], e) ROC diagram[4] and so on. The use of a) $\sim$ d) for the forecast verification has been proposed by many seismologists. The ROC diagram has been used in the field of atmospheric sciences, and it was firstly introduced to seismology by Holliday in 2005. Since then, the researchers in PI method adopt ROC diagram for the forecast verification of earthquake. In keeping with previous work, we will also consider the ROC diagram in our work.

Here, an example is given to show the application of ROC diagram to evaluate PI method forecasting effect.
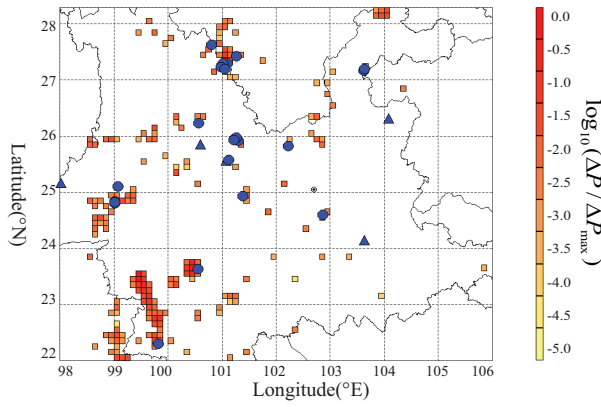
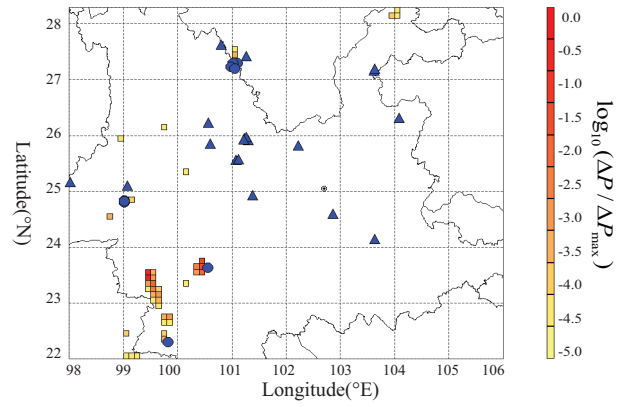Figure 3. The retrospective forecast map at w=-5.0

Figure 4. The retrospective forecast map at w=-2.0

TABLE I. CONTINGENCY TABLE OF RETROSPECTIVE FORECAST AT w=-5.0

Items		Observed		Total
		Yes	No	
Forecast	Yes	24	179	203
	No	5	4832	4837
Total		29	5011	5040

TABLE II. CONTINGENCY TABLE OF RETROSPECTIVE FORECAST AT w=-2.0

Items		Observed		Total
		Yes	No	
Forecast	Yes	10	33	43
	No	19	4978	4997
Total		29	5011	5040

Firstly, retrospective forecast by PI method was conducted for the earthquakes ($M \geq 5$) in Yunnan region. The parameters used in the PI method are chosen as follows: the region studied is located at $22° \sim 28.3°$N and $98° \sim 106°$E. The times are $t_0 = 1/1/1973$, $t_1 = 1/1/1991$, $t_2 = 1/1/1998$, $t_3 = 1/1/2005$, and, this is meant to be a retrospective forecast for the period $1/1/1998 \sim 1/1/2005$. The time step is $\Delta t = 10$ days. The cutoff of the magnitude is $M_c = 3.0$. The scale of box is $\Delta x = 0.1°$, hence, the Yunnan region is divided into 5040 boxes.

Secondly, By PI calculation, the probability (ΔP) of earthquake occurrence in every box is calculated. Based on the rules, we consider the boxes as hotspot, whose probability of a future strong earthquake are positive (that is $\Delta P > 0$). Here, the threshold is $w = -5.0$. At this time, the number of hotspots is 203. The retrospective forecast map obtained by PI method is shown in Fig. 3.

In Fig. 3, the small boxes are the retrospective forecast hotspots for the occurrence of earthquakes with $M \geq 5$ from 1/1/1998 to 1/1/2005, and there were 29 earthquakes with $M \geq 5$ in Yunnan during that period. From Fig. 3 and the contingency table (Tab.1), we can see that 24 earthquakes (solid circles in Fig. 3) were located in the hotspot or its Moore neighborhood. 5 earthquakes (solid triangles in Fig. 3) were located outside the hotspot or its Moore neighborhood. The hit rate is $H = 24/29 = 0.828$. The false alarm rate is $F = 179/5011 = 0.036$.

Then, we change the value of threshold (w), and the number of hotspot is determined by the value of threshold (w). Here, the threshold is $w = -2.0$. At this time, the number of hotspots is 43. The retrospective forecast map is shown in Fig. 4.

From Fig. 4 and the contingency table (Tab.2), we can see that 10 earthquakes (solid circles in Fig. 4) with $M \geq 5$ were located in the hotspot or its Moore neighborhood. 19 earthquakes (solid triangles in Fig. 4) with $M \geq 5$ were located outside the hotspot or its Moore neighborhood. The hit rate is $H = 10/29 = 0.344$. The false alarm rate is $F = 33/5011 = 0.007$.

By gradually changing the threshold value, we can get different retrospective forecast results (hit rate H and false alarm rate F). Then, we draw a ROC diagram obtained by the different results (Fig. 5). In Fig. 5, the solid line denotes the results with the PI forecast. The dashed line denotes the results with the random guess. The filled box correspond to the threshold $w = -2.0$, and the filled triangle correspond to the threshold $w = -5.0$. From Fig. 5, we can see that the PI forecast outperforms random guess (in which the hit rate = the false rate).

III. APPLICATION OF PATTERN INFORMATICS METHOD

In 2008, there were 5 strong earthquakes with magnitude $M \geq 6.9$ occurred one after another in west and southwest China (Tab. 3). However, since 10/1/2003, there are no strong earthquakes with magnitude $M \geq 6$ in north and northeast China. In this paper, we applied the PI method to study strong seismic risk assessment in north and northeast China.

A. Data Used for the Analysis

The region of north and northeast China is located at $30° \sim 54°$N, $108° \sim 135°$E. Historical earthquake catalogue used in this study is provided by the China Earthquake Networks Center. Fig. 6 illustrates the schematic diagram of
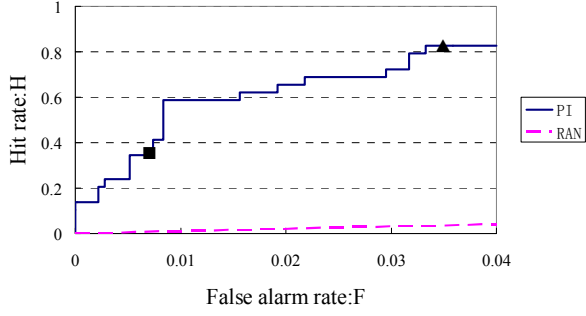
Figure 5. Results of retrospective earthquake forecasting

TABLE III. FIVE STRONG EARTHQUAKES IN WEST AND SOUTHWEST CHINA

Beijing Local Time	Epicenter	Location	Magnitude (M_s)
2008/01/09	(32.5°N,85.2°E)	Gaize	6.9
2008/03/21	(35.6°N,81.6°E)	Yutian	7.3
2008/05/12	(30.95°N,103.4°E)	Wencuan	8.0
2008/08/25	(31°N,83.6°E)	Zhongba	6.8
2008/10/05	(39.5°N,73.9°E)	Wuqia	6.8

the Gutenberg–Richter relation of earthquake. From Fig. 6, we can see that $M_c = 5$ is reasonable in the following study.

B. Application of Pattern Informatics Method to North and Northeast China

Firstly, we choose initial values to some parameters: The initial time was $t_0 = 1/1/1500$. The time step is $\Delta t = 30$ days. The forecast interval is 50 years. The cutoff of the magnitude is $M_c = 5.0$. The "target magnitude" is $M_f \geq 7.0$. Then, we choose different values of the scales of boxes (such as, $\Delta x = 0.5°$, $1.0°$, $1.5°$) for retrospective forecasting. The research results are shown in Fig. 7.

In Fig. 7, the line of RAN corresponds to the results obtained by the random guess. The line of 1.0° is obtained by the approach as follows:

The values of parameters, which are $\Delta t = 30$ days, $t_3 - t_2 = 50$ years, $M_c = 5.0$, $M_f \geq 7.0$, $\Delta x = 1.0°$, are constant. Then , we calculate the hit rate and the false alarm rate by PI method, respectively, when time parameters are: a) $t_0 = 1/1/1500$, $t_1 = 1/1/1750$, $t_2 = 1/1/1800$, $t_3 = 1/1/1850$; b) $t_0 = 1/1/1550$, $t_1 = 1/1/1800$, $t_2 = 1/1/1850$, $t_3 = 1/1/1900$; c) $t_0 = 1/1/1600$, $t_1 = 1/1/1850$, $t_2 = 1/1/1900$, $t_3 = 1/1/1950$; d) $t_0 = 1/1/1650$, $t_1 = 1/1/1900$, $t_2 = 1/1/1950$, $t_3 = 1/1/2000$. At last, we compute the average of hit rates and the average of false alarm rates. Then, we use the average of hit rate and false alarm rate to draw the line of 1.0° in Fig. 7. In this way, the retrospective forecasting results are representative,
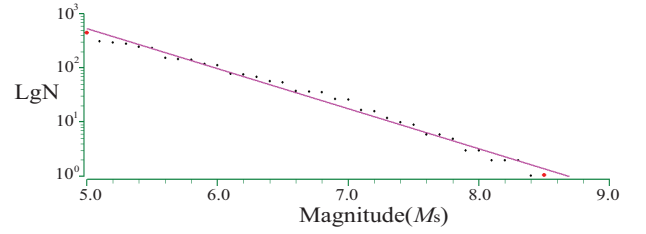

Figure 6. The Gutenberg–Richter relation of earthquakes in north and northeast China
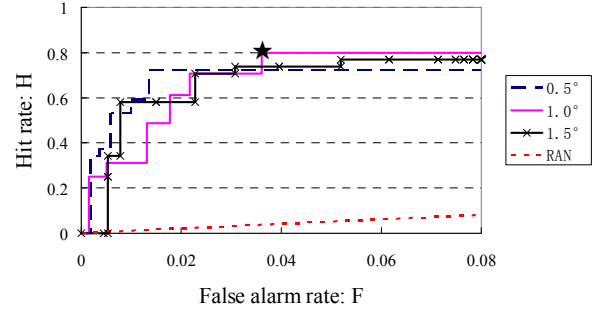

Figure 7. The retrospective forecasting results based on ROC diagram

to avoid random. Besides, the lines of 0.5° and 1.5° are similar as the line of 1.0°.

From Fig. 7, we can see that the "filled star" correspond to hit rate $H = 0.8$, which is higher than others. In this case, the threshold is $w = -1.80$ and the scale of box is $\Delta x = 1.0°$.

Based on the above discussion, we apply the PI method to north and northeast China for future strong earthquakes. In this study, the parameters are chosen as follows:

The times are $t_0 = 1/1/1700$, $t_1 = 1/1/1950$, $t_2 = 1/1/2000$, $t_3 = 1/1/2050$. The time step is $\Delta t = 30$ days. The cutoff of the magnitude is $M_c = 5.0$. The scale of box is $\Delta x = 1.0°$. The threshold is $w = -1.80$. The forecast map is shown in Fig. 8. In Fig. 8, the small boxes are forecast hotspots. The unfilled triangles are earthquakes with $M \geq 7$ from 1/1/1950 to 12/31/1999. The filled circles are earthquakes with $M \geq 7$ from 1/1/2000 to 1/1/2009. The results show that there might be earthquakes with $M \geq 7$ during the period 1/1/2009 to 1/1/2050 in the region circled by unfilled rectangles.

IV. SUMMARY

Although the PI method is in its infancy, the forecasting results obtained by this method are prospective. For example, Rundle has applied the method to southern California for forecasting earthquakes with $M \geq 5$ during the period 2000 to 2010. And, 25 of 27 earthquakes with $M \geq 5$ occurred in the hotspot box or the Moore neighborhood of the hotspot box after the forecast map was published [11]. So, the method has gradually caught attention of seismologists.

In this paper, we applied the PI method to study seismic risk assessment in north and northeast China. The research
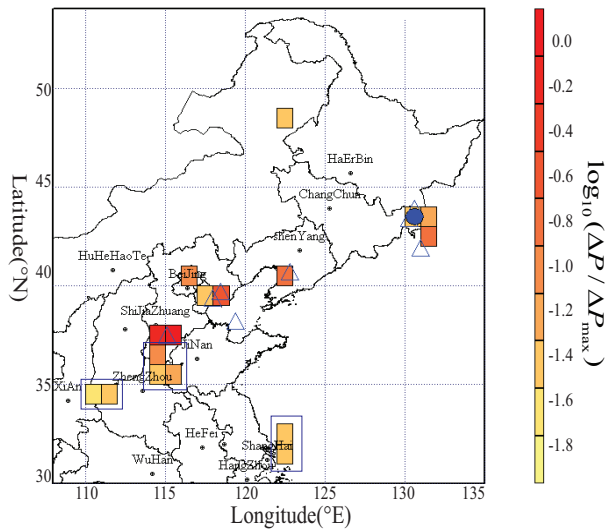
Figure 8.   The forecast map for north and northeast China

suggests that there might be earthquakes ($M \geq 7$) in the juncture of Shanxi, Henan and Shanxi provinces, the juncture of Hebei, Shandong and Henan provinces and southeast coast of Jiangsu province from Jan. 2009 to Jan. 2050.

In this study, some questions are left behind, such as how to set suitable parameters to obtain the best results. Is declustering (deleting the aftershocks) required? And, we will look forward to reader's good suggestions.

REFERENCES

[1]   J.B. Rundle, W. Klein, S.J. Gross, and K.F. Tiampo, Dynamics of seismicity patterns in systems of earthquake faults, Geo-complexity and physics of earthquakes, AGU, Washington, D.C., 2000, pp. 127-146.

[2]   J.B. Rundle, W. Klein, K. Tiampo, and S. Gross, Linear pattern dynamics in nonlinear threshold systems, Phys. Rev. E., 2000, pp. 2418-2432.

[3]   K.F. Tiampo, J.B. Rundle, and W. Klein, Premonitory seismicity changes prior to the Parkfield and Coalinga earthquakes in southern California, Tectonophysics, 2006, pp. 77-86.

[4]   J.R. Holliday, K.Z. Nanjo, K.F. Tiampo, J.B. Rundle, and D.L. Turcotte, earthquake forecasting and its verification, Nonlinear Processes in Geophysics, 2005, pp. 965-977.

[5]   C.C. Chen, J.B. Rundle, J.R. Holliday, K.Z. Nanjo, D.L. Turcotte, S.C. Li, K.F. Tiampo, From Tornadoes to Earthquakes: Forecast verification for Binary events applied to the 1999 Chi-Chi, Taiwan, Earthquake.（In submit）.

[6]   E.F. Moore, Machine models of self reproduction, in proceedings of the fourteenth symposius on applied mathematics, American Mathematical Society, 1962, pp. 17-33.

[7]   S.X. Xu, The evaluation of earthquake prediction ability, In: the album of earthquake prediction methods, 1989, pp. 586-589.

[8]   W.P. Qin, on the problem of grading for one dimensional seismologic predictions, ACTA seismologica Sinca, 1991, pp. 234-242.

[9]   Y.Y. Kagan and D. Jackson, Long-term probabilistic forecasting of earthquakes, J. Geophys. Res. 1994, pp. 13685-13700.

[10] G.M. Molchan, earthquake prediction as a decision-making problem. Pure Appl. Geophys. 1997, pp. 233-247.

[11] http://quakesim.jpl.nasa.gov/scorecard.html