

# Using earthquake intensities to forecast earthquake occurrence times

J. R. Holliday<sup>1,2</sup>, J. B. Rundle<sup>1,2</sup>, K. F. Tiampo<sup>3</sup>, and D. L. Turcotte<sup>4</sup>

<sup>1</sup>Computational Science and Engineering Center, University of California, Davis, USA

<sup>2</sup>Department of Physics, University of California, Davis, USA

<sup>3</sup>Department of Earth Sciences, University of Western Ontario, Canada

<sup>4</sup>Department of Geology, University of California, Davis, USA

Received: 19 July 2006 – Revised: 4 October 2006 – Accepted: 10 October 2006 – Published: 31 October 2006

**Abstract.** It is well known that earthquakes do not occur randomly in space and time. Foreshocks, aftershocks, precursory activation, and quiescence are just some of the patterns recognized by seismologists. Using the Pattern Informatics technique along with relative intensity analysis, we create a scoring method based on time dependent relative operating characteristic diagrams and show that the occurrences of large earthquakes in California correlate with time intervals where fluctuations in small earthquakes are suppressed relative to the long term average. We estimate a probability of less than 1% that this coincidence is due to random clustering. Furthermore, we show that the methods used to obtain these results may be applicable to other parts of the world.

## 1 Introduction

While there is yet no proven method for the reliable short time prediction of earthquakes (minutes to months), it is currently possible to make probabilistic hazard assessments for earthquake risk. The pattern informatics (PI) method for earthquake forecasting identifies geographical regions that have systematic and large fluctuations in seismic activity of the smallest events and quantifies their temporal variations (Rundle et al., 2002; Tiampo et al., 2002b,a; Rundle et al., 2003). The output of this method analysis is a map of areas in a seismogenic region where earthquakes are forecast to occur in a future time span, generally five to ten years.

Recent advances in the PI method show considerable improvement, especially when combined with data from a relative intensity (RI) analyses (Holliday et al., 2005, 2006). RI maps are an alternative approach for earthquake forecasting that specify the locations of the highest seismic activity of the smallest magnitude earthquakes. These forecasts

are based on the hypothesis that future large earthquakes will occur where most smaller earthquakes have occurred in the recent past. Using relative (or receiver) operating characteristic (ROC) analyses (Jolliffe and Stephenson, 2003) to optimize our forecasts, we are able to better specify where future earthquakes are most likely to occur. The question remains, however, of when they are most likely to occur.

We attempt to address this question by analyzing the performance of PI maps against RI maps through time using ROC diagrams. Our hypothesis is that as a region evolves towards a major earthquake in response to persistent loading or stress increase there will be a precursory and systematic change in the separation of the PI and RI curves in an ROC analysis. Since these two measures are sensitive to different effects, we expect the time-dependent differences in the area between the two curves to be sensitive to upcoming events.

## 2 Rationale

There are numerous possible explanations for our hypothesis and subsequent results. From a statistical point of view, it is widely accepted that the observed earthquake scaling laws indicate the existence of phenomena closely associated with proximity of the system to a critical point (Burridge and Knopoff, 1967; Rundle and Jackson, 1977; Carlson et al., 1994; Helmstetter and Sornette, 2002; Main and Al-Kindy, 2002; Chen et al., 1991; Turcotte, 1997; Sornette, 2000; Fisher et al., 1997; Rundle et al., 1996; Klein et al., 1997). More specifically, it has been proposed that earthquake dynamics are associated either with a second order critical point (Carlson et al., 1994; Helmstetter and Sornette, 2002; Main and Al-Kindy, 2002; Chen et al., 1991; Turcotte, 1997; Sornette, 2000; Fisher et al., 1997) or a mean field spinodal (Rundle et al., 1996; Klein et al., 1997) that can be understood as a line of critical points. Mean field theories have been proposed (Sornette, 2000; Fisher et al., 1997; Rundle

*Correspondence to:* J. R. Holliday  
(holliday@cse.ucdavis.edu)

et al., 1996; Klein et al., 1997) to explain the phenomenology associated with scaling and nucleation processes of earthquakes. If mean field equations do describe earthquakes, the dynamics must be operating outside the critical region, and fluctuations are correspondingly reduced. This reduction would effectively enhance the forecasting performance of the RI map, which emphasizes sites with high historical rates of activity, and would degrade the forecasting performance of the PI map, which emphasizes sites with large fluctuations from historical rates of activity.

From a geophysical point of view, there are a growing number of cases reported where the occurrence of a large earthquake is preceded by a regional increase in seismic energy release (Jaumé and Sykes, 1999). This phenomenon is known as accelerating moment release (AMR) and is due primarily to an increase in the number of intermediate-size events that occur within a characteristic distance of the main shock and that scale with magnitude. AMR is characterized by a decrease in the rate of regional seismicity, causing growing fluctuations from historic rates which enhances the forecasting performance of the PI map, followed by a rapid rebound back to historic levels, causing a decrease in fluctuations from historic rates which enhances the forecasting performance of the RI map. Sammis and Bowman (2006) have proposed a number of physical models to explain AMR. These include:

- An analogy with critical phase transitions where the correlation length of the stress field rapidly increases.
- An erosion of a stress shadow from some previous, large event.
- A slow, silent earthquake propagating upward on a ductile extension loading the seismogenic crust above.

We note that the existence of such a seismicity pattern appears to require a certain regional fault system structure and density. Simulation models using a hierarchical distribution of fault sizes match this pattern well, but other types of fault distributions may also support AMR (Jaumé and Sykes, 1999). Conversely, some real-world fault distributions may not support AMR as a predictive tool.

### 3 Procedure

For this new analysis, we start with a regional catalog of earthquake epicenters and create a time series by course-graining it in regular time intervals (typically daily to monthly). In order to ensure catalog completeness (Gutenberg-Richter scaling) we only look at events above some threshold magnitude  $m_c$ . We then tile the region with a spatially course-grained, regular mesh of  $N$  boxes. The number of earthquakes in a box  $i$ , located at  $x_i$ , at a time  $t$  is denoted by  $n(x_i; t)$ .

We next construct an RI map  $I(x_i; t_0, t_2)$  by computing the number of earthquakes  $n(x_i; t_0, t_2)$  in each coarse-grained box  $i$  from some base time  $t_0$  until a later time  $t_2$  that will be allowed to vary:

$$n(x_i; t_0, t_2) = \sum_{t=t_0}^{t_2} n(x_i; t). \quad (1)$$

We then regard  $n(x_i; t_0, t_2)$  as a non-normalized probability for the location of future events  $m \geq m_T$  for times  $t > t_2$ , where  $m_T$  is the forecast threshold. We normalize this probability by requiring the integral over the region be equal to unity:

$$I(x_i; t_0, t_2) = \frac{n(x_i; t_0, t_2)}{\sum_{i=1}^N n(x_i; t_0, t_2)}. \quad (2)$$

As can be seen,  $I(x_i; t_0, t_2)$  is a measure of the historic seismic rate as a function of location. Previous work (Rundle et al., 2002; Tiampo et al., 2002a; Holliday et al., 2005) has indicated that this normalized probability is by itself a good predictor of locations for future large events with  $m_T \approx 5$ .

We next construct a PI map  $\Delta I(x_i; t_0, t_1, t_2)$  by building upon the RI map and computing the average change in earthquake intensity over a time interval  $\Delta t = t_2 - t_1$ . Consistent with previous work (Rundle et al., 2002; Tiampo et al., 2002a), we typically choose  $\Delta t = 13$  years, although we note that shorter time intervals may be appropriate depending on the quality of the input catalog. We compute  $n(x_i; t_b, t_1)$  and  $n(x_i; t_b, t_2)$  for the two times  $t_1$  and  $t_2$  ( $t_2 > t_1$ ) beginning at some base time  $t_b$ , where  $t_1 > t_b > t_0$ , and calculate the change in numbers of events  $\Delta n(x_i; t_b, t_1, t_2)$  from  $t_1$  to  $t_2$ :

$$\Delta n(x_i; t_b, t_1, t_2) = n(x_i; t_b, t_2) - n(x_i; t_b, t_1). \quad (3)$$

We then normalize  $\Delta n(x_i; t_b, t_1, t_2)$  to have zero mean and unit variance:

$$\Delta n'(x_i; t_b, t_1, t_2) = \frac{\Delta n(x_i; t_b, t_1, t_2) - \mu}{\sigma}, \quad (4)$$

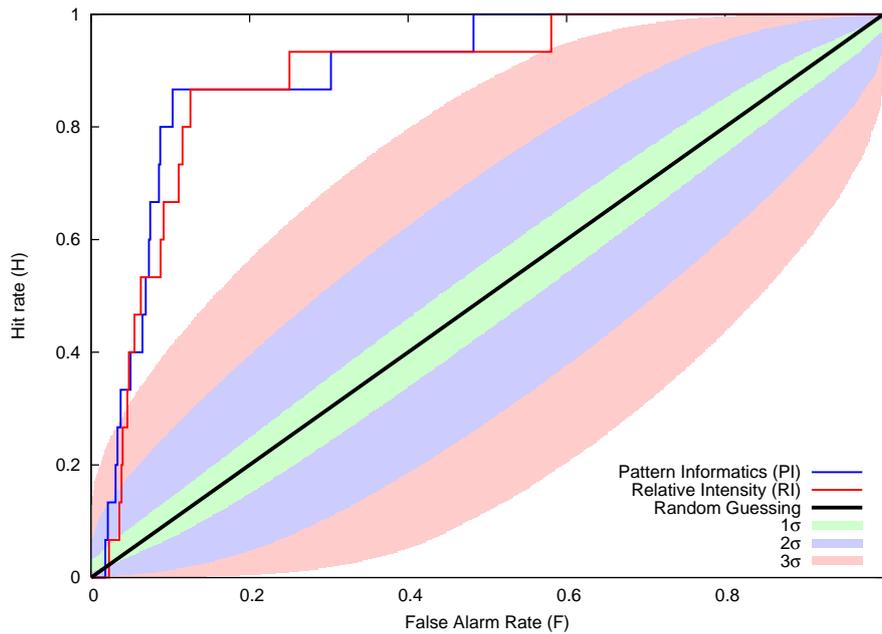
where  $\mu$  and  $\sigma$  are respectively the mean and variance of  $\Delta n(x_i; t_b, t_1, t_2)$ . Finally we average  $\Delta n'(x_i; t_b, t_1, t_2)$  over all base times  $t_b$  in the range  $t_0 < t_b < t_1$ :

$$\Delta \bar{n}'(x_i; t_0, t_1, t_2) = \frac{\sum_{t_b=t_0}^{t_1-1} \Delta n'(x_i; t_b, t_1, t_2)}{t_1 - t_0 - 1}. \quad (5)$$

This step helps to reduce the relative importance of random fluctuations (noise) in the cataloged seismic activity. The normalized probability is again constructed by requiring the integral over the region be equal to unity:

$$\Delta I(x_i; t_0, t_1, t_2) = \frac{\Delta \bar{n}'(x_i; t_0, t_1, t_2)}{\sum_{i=1}^N \Delta \bar{n}'(x_i; t_0, t_1, t_2)}. \quad (6)$$

Previous work (Rundle et al., 2002; Tiampo et al., 2002a; Holliday et al., 2005) has indicated that this normalized probability when squared is by itself also a good predictor of locations for future large events with  $m_T \approx 5$ . Given its construction, this probability can be viewed as a probability based upon the squared change in earthquake intensity.



**Fig. 1.** Example relative operating characteristic (ROC) diagram. Shown is a plot of hit rates,  $H$ , as a function of false alarm rates,  $F$ , for a sample PI forecast (blue), sample RI forecast (red), and random guessing (black). Confidence intervals for the one-, two- and three-sigma levels are shown as well (Holliday et al., 2005; Zechar and Jordan, 2005).

We next convert the normalized probabilities into binary forecasts by using a decision threshold  $D$ . Locations where the normalized probability is greater than  $D$  constitute locations where future large events are hypothesized to preferentially occur. Binary forecasts are a well-known and utilized method for constructing forecasts of future event locations and have been widely used in tornado and severe storm forecasting (Jolliffe and Stephenson, 2003; Holliday et al., 2005). Comparison of the forecast utility of the two binary forecast maps can be used to establish which map is a better predictor of the locations of future large earthquakes during a future evaluation period  $t > t_2$ . The better map is the one that scores more highly on the comparison test for the particular value of  $D$  that is used (Holliday et al., 2005; Jolliffe and Stephenson, 2003).

For a given value of  $D$ , we set  $B_{RI}(x_i)=1$  where  $I(x_i; t_0, t_2) > D$  and  $B_{RI}(x_i)=0$  otherwise. Similarly, we set  $B_{PI}(x_i)=1$  where  $\Delta I(x_i; t_0, t_1, t_2) > D$  and  $B_{PI}(x_i)=0$  otherwise. The set of boxes where  $B_{RI}(x_i)=1$  or  $B_{PI}(x_i)=1$  then constitute locations where future events  $m \geq m_T$  are hypothesized to be likely to occur under the chosen forecast. The locations where  $B_{RI}(x_i)=0$  or  $B_{PI}(x_i)=0$  are sites where future events  $m \geq m_T$  are hypothesized to be unlikely to occur. In past work (Rundle et al., 2002; Tiampo et al., 2002a; Holliday et al., 2005) we have taken  $m_T=5$ , but we relax that restriction in this analysis and allow  $m_T$  to vary.

ROC diagrams are next constructed for the binary RI and PI maps. The ROC diagram is a parametric plot of the hit rate,  $H(D)$ , as a function of the false alarm rate,  $F(D)$ , as  $D$  is varied from 0 to 1 (Holliday et al., 2005; Jolliffe and

Stephenson, 2003). A perfect forecast of occurrence would consist of two line segments: the first connecting the points  $\{F, H\}=\{0, 0\}$  to  $\{F, H\}=\{0, 1\}$ , and the second connecting  $\{F, H\}=\{0, 1\}$  to  $\{F, H\}=\{1, 1\}$ . A curve of this type can be described as forecasting all future earthquakes ( $H=1$ ) with no false alarms ( $F=0$ ). The line  $H=F$  occupies a special status and corresponds to a completely random forecast (Holliday et al., 2005; Jolliffe and Stephenson, 2003) where the false alarm rate is the same as the hit rate and no information is produced by the forecast. Alternatively, we can say that the marginal utility (Chung, 1994) of an additional hotspot,  $dH/dF$ , equals unity for a random forecast. A sample ROC diagram is presented in Fig. 1.

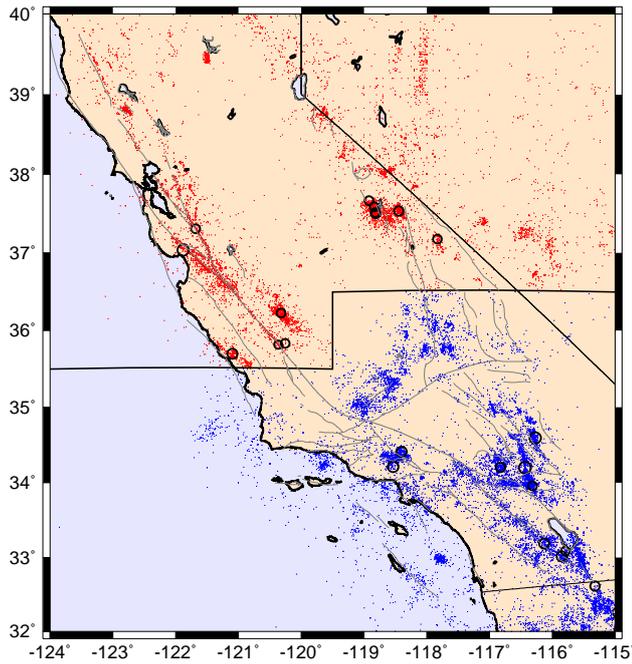
For a given time-dependent forecast  $H(F; t)$ , we are motivated to consider the time-dependent Pierce Skill Score  $H(F; t) - F$  (Jolliffe and Stephenson, 2003), which measures the increase in performance of  $H(F; t)$  relative to the random forecast  $H=F$ . We define the Pierce function  $A(t)$  as the area between  $H(F; t)$  and the random forecast:

$$A(t) = \int_0^{F_{\max}} \{H(F; t) - F\} dF, \tag{7}$$

where the upper limit on the range of integration,  $F_{\max}$ , is a parameter whose value is set by the requirement that the marginal utility (Chung, 1994) of the forecast of occurrence  $H(F; t)$  exceeds that of the random forecast  $H=F$ :

$$\frac{d}{dF} \{H(F; t) - F\} > 0. \tag{8}$$

Since  $H(F; t)$  curves are monotonically increasing,  $F_{\max}$  is determined as the value of  $F$  for which  $dH(F; t)/dF=1$ .



**Fig. 2.** Map of earthquake ( $m \geq 3.0$ ) epicenters in California from 1932 to the present. Circles are events with  $m \geq 6.0$  since 1960. Red epicenters define the area used to analyze seismicity in northern California; blue epicenters define the area used for southern California.

Finally, we hypothesize that as a region evolves towards a major earthquake in response to persistent loading or stress increase there will be a precursory and systematic change in the separation of  $A(t)$  for the RI maps and  $A(t)$  for the PI maps. Since these two measures are sensitive to different effects, we expect the time-dependent differences in the area between the two curves to be sensitive to upcoming events. Thus we plot the Pierce difference function  $\Delta A(t)$ , where

$$\Delta A(t) = A_{\text{RI}}(t) - A_{\text{PI}}(t), \quad (9)$$

as a function of time over a range of choices for  $m_T$ .

In particular, we first consider the Gutenberg-Richter frequency-magnitude relation  $f = 10^a \cdot 10^{-bm}$ , where  $f$  is the number of events per unit time with magnitude larger than  $m$  and  $a$  and  $b$  are constants.  $a$  specifies the level of activity in the region, and  $b \cong 1$ . To construct ROC curves, we consider  $t$  to be the current time at each time step and test the RI map and PI map by forecasting locations of earthquakes having  $m \geq m_T$  during  $t_2$  to  $t$ , where  $m_T$  is some forecast threshold magnitude. Note that  $f^{-1}$  specifies a time scale for events larger than  $m$ : 1 event with  $m \geq 6.0$  is associated on average with 10  $m \geq 5.0$  events, 100  $m \geq 4.0$  events, etc. Without prior knowledge of the optimal value for  $m_T$ , we average the results for a scale-invariant distribution of 1000  $m_T \geq 3.0$  events, 794  $m_T \geq 3.1$  events, 631  $m_T \geq 3.2$  events, ..., 10  $m_T \geq 5.0$  events, etc. At some point, which catalog and region dependent, we terminate the sequence due to

increasingly poor statistics. To control the number of earthquakes with  $m \geq m_T$  in the “snapshot window” ( $t_2 \rightarrow t$ ), we determine the value of  $t_2$  that most closely produces the desired number of events. We then average  $\Delta A(t)$  over all choices of  $m_T$ , yielding  $\Delta \mathcal{A}(t)$ , and identify periods where  $\Delta \mathcal{A}(t) > 0$  as times where future large events are hypothesized to be likely to occur. These are time periods where fluctuations, as measured by PI, are decreased relative to historic rates, as measured by RI.

A central idea is that the length of the snapshot window is not fixed in time; it is instead fixed by earthquake number at each threshold magnitude  $m_T$ . Nature appears to measure “earthquake time” in numbers of events, rather than in years (Varotsos et al., 2006). This time scale is evidently based on stress accumulation and release, that is, earthquake numbers, rather than in months or years (Klein et al., 1997). We call this procedure of forecasting earthquake occurrence times ensemble intensity differencing, or EID.

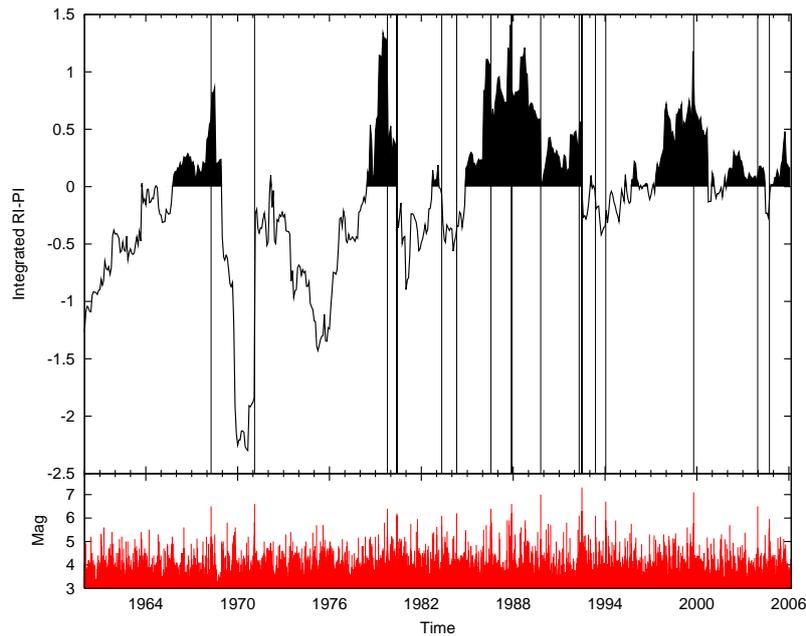
### 3.1 Itemized procedure

In cookbook format, the EID method as applied in this paper is composed of the following steps:

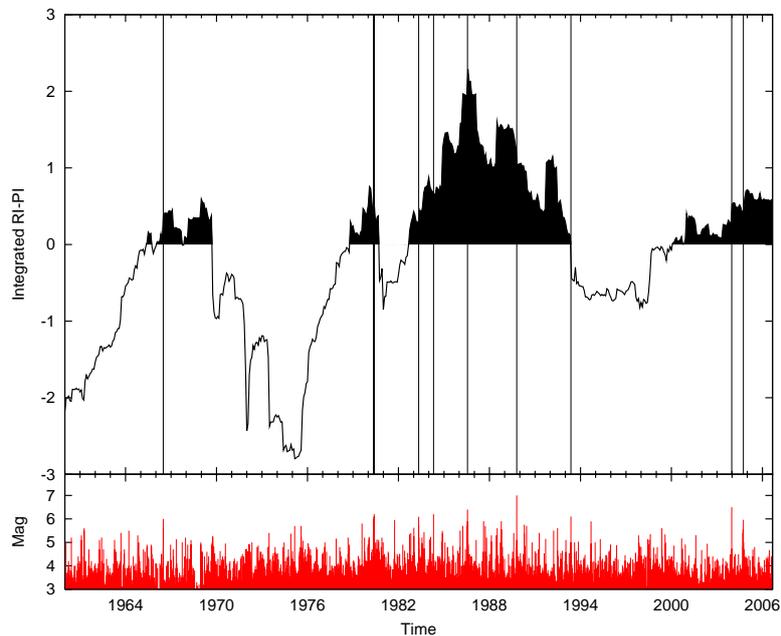
1. The seismically active region is binned into boxes of some characteristic size, and all events having  $m \geq m_c$  are used. These boxes are labeled  $x_i$ .
2. The seismicity obtained from the regional catalog for each day in each box is considered to be uniformly spread over that box. The resulting intensities for each box forms a time series.
3. Three time parameters are determined:  $t_0$ ,  $t_1$ , and  $t_2$ .  $t_0$  is chosen to be the base time. For California, we typically take  $t_0 = 1$  January 1932.  $t_2$  is chosen such that the number of  $m > m_T$  events during the time period  $t_2 \rightarrow t$  is equal to some value specified by the regional  $b$ -value.  $t_1$  is chosen such that  $t_2 - t_1 = \Delta t$ .
4. RI,  $I(x_i; t_0, t_2)$ , and PI,  $\Delta I(x_i; t_0, t_1, t_2)$ , maps are created for the region.
5. These maps are converted to binary forecasts, and ROC diagrams are constructed for the snapshot window  $t_2 \rightarrow t$ .  $\Delta A(t)$  is calculated by integrating  $A_{\text{RI}}(t) - A_{\text{PI}}(t)$  over  $F \in [0, F_{\text{max}}]$ .
6. Finally,  $\Delta A(t)$  is averaged (or for simplicity summed) for a range of snapshot windows yielding  $\Delta \mathcal{A}(t)$ . If  $\Delta \mathcal{A}(t) > 0$ , a warning is issued that future large earthquakes are likely to occur. If  $\Delta \mathcal{A}(t) < 0$ , no such warning is issued.

## 4 Application to California

To apply this method to California, we use the ANSS catalog of earthquakes between latitude 32° N and 40° N, and lon-



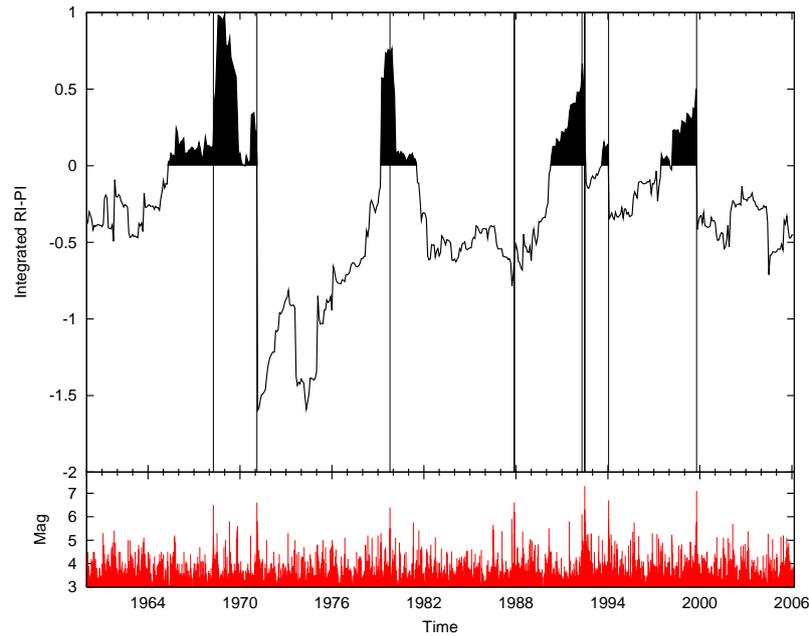
**Fig. 3.** Value of the Pierce difference function  $\Delta\mathcal{A}(t)$  (top) and magnitude (bottom) as a function of time for events occurring on the entire map area of Fig. 2 (red and blue epicenters). Vertical black lines represent times of major earthquakes having  $m \geq 6.0$ . Pierce differences are computed for a scale-invariant distribution of magnitude thresholds in the snapshot window from  $m_T=3.0$  to  $m_T=5.0$ . Area integration is performed for  $F \in [0.00, 0.90]$ .



**Fig. 4.** Value of the Pierce difference function  $\Delta\mathcal{A}(t)$  (top) and magnitude (bottom) as a function of time for events occurring on the northern map area of Fig. 2 (red epicenters). Vertical black lines represent times of major earthquakes having  $m \geq 6.0$ . Pierce differences are computed for a scale-invariant distribution of magnitude thresholds in the snapshot window from  $m_T=3.0$  to  $m_T=5.0$ . Area integration is performed for  $F \in [0.00, 0.90]$ .

gitudes  $125^\circ$  W and  $115^\circ$  W. Only events above a threshold magnitude  $m_c=3$  are used to ensure catalog completeness. Fig. 2 shows the event locations with “northern” epicenters shown in red and “southern” epicenters shown in blue. The

coarse-grained mesh we use for this region consists of boxes having a side length of  $0.1^\circ$ , about 11 km at these latitudes, approximately the rupture length of an  $m \sim 6$  earthquake.



**Fig. 5.** Value of the Pierce difference function  $\Delta\mathcal{A}(t)$  (top) and magnitude (bottom) as a function of time for events occurring on the southern map area of Fig. 2 (blue epicenters). Vertical black lines represent times of major earthquakes having  $m \geq 6.0$ . Pierce differences are computed for a scale-invariant distribution of magnitude thresholds in the snapshot window from  $m_T = 3.0$  to  $m_T = 5.0$ . Area integration is performed for  $F \in [0.00, 0.90]$ .

Results for analyzing this entire region are shown in Fig. 3. At the top of this plot is the Pierce difference function  $\Delta\mathcal{A}(t)$  as a function of time from 1 January 1960 to 31 December 2005. At the bottom is the earthquake magnitude as a function of time over the same period. The vertical lines are the times of all  $m \geq 6.0$  events in the region. Information for these events is given in Table 1. While many of the events fall within (black) time intervals where  $\Delta\mathcal{A}(t) > 0$ , the results appear noisy and some events are missed. We can improve this result by carefully treating northern and southern California separately.

Results for analyzing Northern and Southern California separately are shown in Figs. 4 and 5. The choice of where to divide the total region was made by considering the fault structure and local seismicity profile near  $36^\circ$  N latitude. From Fig. 2 we see that there are twelve  $m \geq 6$  events in northern California and ten such events in southern California. These major events are concentrated into distinct episodes corresponding to distinct main shocks. In the northern California plot, all major episodes fall during (black) time intervals where  $\Delta\mathcal{A}(t) > 0$  or they terminate such a time interval. In the southern California plot, seven of the eight major episodes fall during (black) time intervals where  $\Delta\mathcal{A}(t) > 0$  or they terminate such a time interval. If a binomial probability distribution is assumed, the chance that random clustering of these major earthquake episodes could produce this temporal concordance can be computed. For Fig. 4, where black time intervals constitute 36.8% of the total, we compute a 0.19% chance that the concordance is due to random

clustering. For Fig. 5, the respective numbers are 19% of the total time interval, and 0.0058% chance due to random clustering. Our results support the hypothesis that major earthquake episodes preferentially occur during time intervals when  $\Delta\mathcal{A}(t) > 0$ . Furthermore, we note that currently in northern California  $\Delta\mathcal{A}(t) > 0$ .

In an attempt to identify precisely where the  $\Delta\mathcal{A}(t) > 0$  signal is originating, we further subdivided the region. As we continue to shrink the test region, however, we increase our sensitivity to random (as opposed to systematic) fluctuations and also to uncertainties in the event locations. In addition, we find that the snapshot window often grows too large ( $t_2$  approaching  $t_0$ ) as there are fewer events contained in the subregion. These effects each cause the forecast to degrade significantly and make it difficult to isolate individual source locations for the  $\Delta\mathcal{A}(t) > 0$  signal. With these limitations, we are only able to say that the signal is most likely coming from the north-western part of California between latitudes  $37^\circ$  N and  $39^\circ$  N and west of the valley.

## 5 Application to Sumatra

To apply this method to Sumatra, we used the ANSS catalog of earthquakes between latitude  $12^\circ$  S and  $5^\circ$  N, and longitudes  $95^\circ$  E and  $115^\circ$  E. Only events above a threshold magnitude  $m_c = 5$  were used to ensure catalog completeness. Figure 6 shows the event locations. The coarse-grained mesh we

**Table 1.** Dates, locations, and magnitudes of earthquakes with  $m \geq 6.0$  since 1960 in California. Latitudes and Longitudes are colored to correspond with Fig. 2.

Date	Latitude	Longitude	Mag
9 April 1968	33.1900° N	116.129° W	6.5
9 February 1971	34.4112° N	118.401° W	6.6
15 October 1979	32.6137° N	115.318° W	6.4
25 May 1980	37.5905° N	118.833° W	6.1
25 May 1980	37.6673° N	118.918° W	6.0
25 May 1980	37.5185° N	118.820° W	6.1
27 May 1980	37.5002° N	118.808° W	6.2
2 May 1983	36.2277° N	120.318° W	6.0
24 April 1984	37.3097° N	121.679° W	6.2
21 July 1986	37.5387° N	118.443° W	6.4
24 November 1987	33.0900° N	115.792° W	6.2
24 November 1987	33.0150° N	115.852° W	6.6
18 October 1989	37.0362° N	121.880° W	7.0
23 April 1992	33.9600° N	116.317° W	6.1
28 June 1992	34.2000° N	116.437° W	7.3
28 June 1992	34.2030° N	116.827° W	6.3
17 May 1993	37.1763° N	117.832° W	6.1
17 January 1994	34.2130° N	118.537° W	6.7
16 October 1999	34.5940° N	116.271° W	7.1
22 December 2003	35.7002° N	121.097° W	6.5
28 September 2004	35.8182° N	120.366° W	6.0

**Table 2.** Dates, locations, and magnitudes of earthquakes with  $m \geq 8.0$  since 1980 in Sumatra.

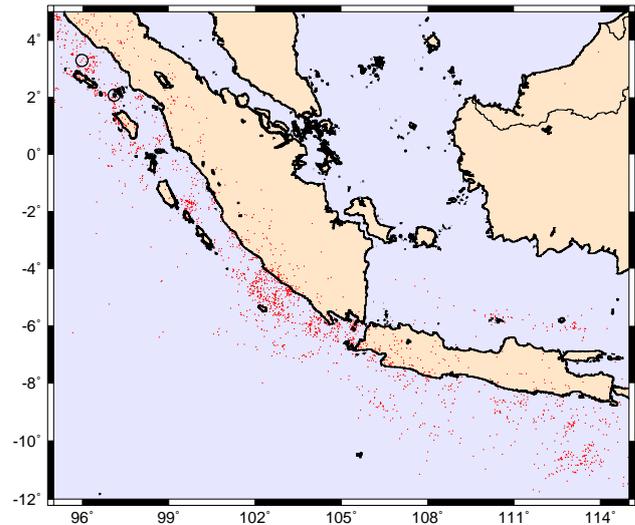
Date	Latitude	Longitude	Mag
26 December 2004	3.295° N	95.982° E	9.0
28 March 2005	2.085° N	97.108° E	8.6

used for this region consists of boxes having a side length of  $1^\circ$ .

Results for analyzing this entire region are shown in Fig. 7. At the top of this plot is the time dependent Pierce difference function  $\Delta\mathcal{A}(t)$  as a function of time from 1 January 1980 to 31 May 2006. At the bottom is the earthquake magnitude as a function of time over the same period. The vertical lines mark the times of the two  $m \geq 8$  events in the region. It can be seen that for the majority of the time period, the time dependent Pierce difference function was negative. It only became positive approximately one year before the tsunami-causing  $m=9$  event on 26 December 2004. We note that currently in Sumatra,  $\Delta\mathcal{A}(t) > 0$ .

## 6 Conclusions

While initial results from our EID analysis are promising, we are aware that more refinements and testing are needed. In particular, we currently lack a standardized method for de-



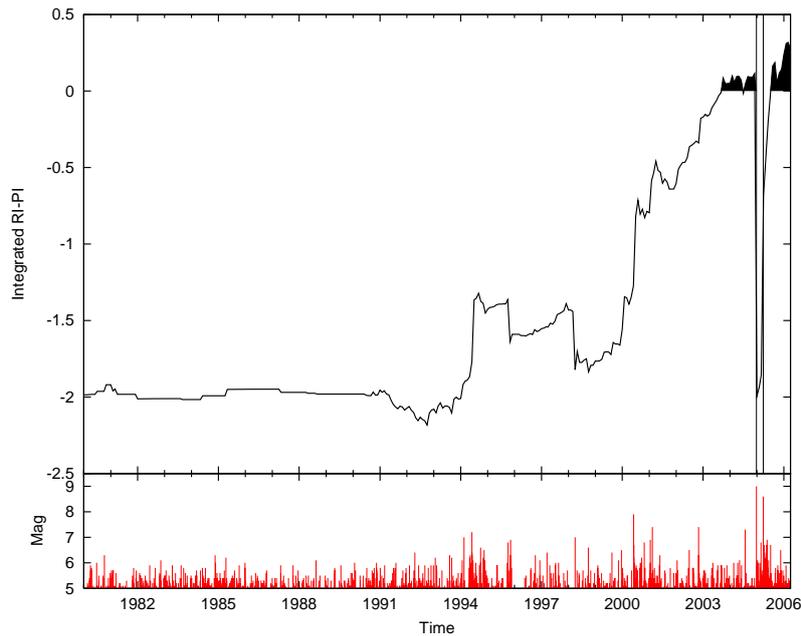
**Fig. 6.** Map of earthquake ( $m \geq 5.0$ ) epicenters in Sumatra from 1950 to the present. Circles are the two events with  $m \geq 8.0$  since 1980.

termining the bounds of the seismogenic region to be tested. In the California application above, the choice of where to divide northern California from southern California was made by considering primarily the general fault structures, the clusters of past event locations, and the ease to which the division could be incorporated into the analysis. Clearly this inserts an amount of subjective prejudice into the analysis. A universal procedure needs to be developed and utilized. Also, we need to better understand the dependence on  $m_c$  and  $m_T$  as different regions (with different fault topologies and different Gutenberg-Richter scaling parameters) and different catalogs are studied.

An interesting product of this analysis is the notion that relevant time intervals are based on stress accumulation and release and are therefore measured in numbers smaller earthquakes at threshold magnitudes  $m_T$ . One way to view this is as a Fourier series, only in terms of numbers of events. The fundamental mode of, say, a  $m=6$  event is one earthquake during the time computed from  $\tau=1/f$ , where  $f=10^a \cdot 10^{-bm}$ . During this time  $\tau$  on average, assuming  $b \cong 1$ , there are two events having  $m > m_T=5.7$  (first harmonic), three events having  $m > m_T=5.5$  (second harmonic), and so forth. Because of clustering and statistics, however, a time window that contains 1000  $m > 3$  events is not exactly the same as a window that contains 100  $m > 4$  events; the window length is scale dependent. Our process of summing over an ensemble of  $m_T$  values averages out these statistics.

### 6.1 Summary

To summarize our results, we compare the performance of two probability measures that define the locations of future earthquake occurrence: the spatially coarse-grained relative



**Fig. 7.** Value of the Pierce difference function  $\Delta\mathcal{A}(t)$  (top) and magnitude (bottom) as a function of time for events occurring on the map area of Fig. 6. Vertical black lines represent times of major earthquakes having  $m \geq 8.0$ . Pierce differences are computed for a scale-invariant distribution of magnitude thresholds in the snapshot window from  $m_T = 5.0$  to  $m_T = 8.0$ . Area integration is performed for  $F \in [0.00, 1.00]$ .

intensity map and the pattern informatics map. By examining the time-dependent differences between an ensemble collection of these two measures, we find that future large earthquakes tend to correlate with time intervals where fluctuations in small earthquakes are suppressed relative to the long term average. In particular, we find that since 1960, major earthquakes in California with magnitudes  $m \geq 6$  tend to preferentially occur during intervals of time when  $\Delta\mathcal{A}(t) > 0$ . We also find that the procedure used to obtain this result may be valid over other parts of the world. Work is currently underway to apply these methods to Japan and Taiwan using the respective local catalogs. Finally, we note that currently  $\Delta\mathcal{A}(t) > 0$  in both northern California and in Sumatra.

*Acknowledgements.* This work has been supported by NASA Headquarters under the Earth System Science Fellowship Grant NGT5 (JRH), by a grant from the US Department of Energy, Office of Basic Energy Sciences to the University of California, Davis DE-FG03-95ER14499 (JRH and JBR), by an HSERC Discovery grant (KFT), and through additional funding from the National Science Foundation (DLT) and the National Aeronautics and Space Administration under grants through the Jet Propulsion Laboratory to the University of California, Davis.

Edited by: G. Zoeller

Reviewed by: M. Eneva

## References

- Burridge, R. and Knopoff, L.: Model and theoretical seismicity, *Bull. Seism. Soc. Am.*, 57, 341–371, 1967.
- Carlson, J. M., Langer, J. S., and Shaw, B. E.: Dynamics of earthquake faults, *Rev. Mod. Phys.*, 66, 657–670, 1994.
- Chen, K., Bak, P., and Obukhov, S. P.: Self-organized criticality in a crack-propagation model of earthquakes, *Phys. Rev. A*, 43, 625–630, 1991.
- Chung, J. W.: *Utility and Production Functions*, Blackwell, Oxford, 1994.
- Fisher, D. S., Dahmen, K., Ramanathan, S., and Ben-Zion, Y.: Statistics of Earthquakes in Simple Models of Heterogeneous Faults, *Phys. Rev. Lett.*, 78, 4885–4888, 1997.
- Helmstetter, A. and Sornette, D.: Subcritical and supercritical regimes in epidemic models of earthquake aftershocks, *J. Geophys. Res.*, 107, 2237, doi:10.1029/2001JB001580, 2002.
- Holliday, J. R., Nanjo, K. Z., Tiampo, K. F., Rundle, J. B., and Turcotte, D. L.: Earthquake forecasting and its verification, *Nonlin. Processes Geophys.*, 12, 965–977, 2005, <http://www.nonlin-processes-geophys.net/12/965/2005/>.
- Holliday, J. R., Rundle, J. B., Tiampo, K. F., Klein, W., and Donnellan, A.: Systematic procedural and sensitivity analysis of the pattern informatics method for forecasting large ( $M \geq 5$ ) earthquake events in southern California, *Pure Appl. Geophys.*, in press, 2006.
- Jaumé, S. C. and Sykes, L. R.: Evolving towards a critical point: A review of accelerating seismic moment/energy release prior to large and great earthquakes, *Pure Appl. Geophys.*, 155, 279–306, 1999.
- Jolliffe, I. T. and Stephenson, D. B.: *Forecast Verification*, John Wiley, Chichester, 2003.

- Klein, W., Rundle, J. B., and Ferguson, C. D.: Scaling and Nucleation in Models of Earthquake Faults, *Phys. Rev. Lett.*, 78, 3793–3796, 1997.
- Main, I. G. and Al-Kindy, F. H.: Entropy, energy, and proximity to criticality in global earthquake populations, *Geophys. Res. Lett.*, 29, 4078, doi:10.1029/2001GL014078, 2002.
- Rundle, J. B. and Jackson, D. D.: Numerical simulation of earthquake sequences, *Bull. Seism. Soc. Am.*, 67, 1363–1377, 1977.
- Rundle, J. B., Klein, W., and Gross, S. J.: Dynamics of a Traveling Density Wave Model for Earthquakes, *Phys. Rev. Lett.*, 76, 4285–4288, 1996.
- Rundle, J. B., Tiampo, K. F., Klein, W., and Martins, J. S. S.: Self-organization in leaky threshold systems: The influence of near-mean field dynamics and its implications for earthquakes, neurobiology, and forecasting, *Proc. Natl. Acad. Sci. USA*, 99, 2514–2521, Suppl. 1, 2002.
- Rundle, J. B., Turcotte, D. L., Shcherbakov, R., Klein, W., and Sammis, C.: Statistical physics approach to understanding the multi-scale dynamics of earthquake fault systems, *Rev. Geophys.*, 41, 1019, doi:10.1029/2003RG000135, 2003.
- Sammis, C. G. and Bowman, D. D.: Competing Models for Accelerating Moment Release Before Large Earthquakes, 5th Annual ACES International Workshop, 2006.
- Sornette, D.: *Critical Phenomena in the Natural Sciences*, Springer, Berlin, 2000.
- Tiampo, K. F., Rundle, J. B., McGinnis, S., Gross, S. J., and Klein, W.: Eigenpatterns in southern California seismicity, *J. Geophys. Res.*, 107, 2354, doi:10.1029/2001JB000562, 2002a.
- Tiampo, K. F., Rundle, J. B., McGinnis, S., and Klein, W.: Pattern dynamics and forecast methods in seismically active regions, *Pure App. Geophys.*, 159, 2429–2467, 2002b.
- Turcotte, D. L.: *Fractals & Chaos in Geology & Geophysics*, Cambridge University Press, Cambridge, 2nd edn., 1997.
- Varotsos, P. A., Sarlis, N. V., Skordas, E. S., Tanaka, H. K., and Lazaridou, M. S.: Attempt to distinguish long-range temporal correlations from the statistics of the increments by natural time analysis, *Phys. Rev. E*, 74, 0211 231–02112 312, 2006.
- Zechar, J. D. and Jordan, T. H.: Evaluation techniques for alarm-based forecasts, *EOS Trans. AGU*, Fall Meeting, abstr. no.: S41D-04, 2005.