Forecasting the Locations of Future Large Earthquakes: An Analysis and Verification

ROBERT SHCHERBAKOV,[1] DONALD L. TURCOTTE,[2] JOHN B. RUNDLE,[3] KRISTY F. TIAMPO,[1] and JAMES R. HOLLIDAY[3]

Abstract—The objective of this paper is to quantify the use of past seismicity to forecast the locations of future large earthquakes and introduce optimization methods for the model parameters. To achieve this the binary forecast approach is used where the surface of the Earth is divided into $l° \times l°$ cells. The cumulative Benioff strain of $m \geq m_c$ earthquakes that occurred during the training period, ΔT_{tr}, is used to retrospectively forecast the locations of large target earthquakes with magnitudes $\geq m_T$ during the forecast period, ΔT_{for}. The success of a forecast is measured in terms of hit rates (fraction of earthquakes forecast) and false alarm rates (fraction of alarms that do not forecast earthquakes). This binary forecast approach is quantified using a receiver operating characteristic diagram and an error diagram. An optimal forecast can be obtained by taking the maximum value of Pierce's skill score.

Key words: Earthquake forecasting, ROC diagram, error diagram, forecast verification.

1. Introduction

Forecasts of the locations of future major earthquakes play an important role in earthquake preparedness and determining earthquake insurance costs. Many such forecasts have been carried out, one example is the National Seismic Hazard Map for the United States (FRANKEL *et al.*, 1996). This is a probabilistic estimate of the ground-shaking risk. A major input into this assessment is the observed rate of occurrence of past earthquakes. KOSSOBOKOV *et al.* (2000) utilized the rates of occurrence of $m \geq 4.0$ earthquakes globally to forecast the expected rates of occurrence of larger earthquakes. Recently, the Regional Earthquake Likelihood Models (RELM) forecasts were published (FIELD, 2007). This represents the first-generation of forecasts based on a variety of approaches and methods. It also introduces a testing methodology to verify and compare these forecasts (SCHORLEMMER *et al.*, 2007).

In this paper we have presented a method to construct a binary forecast of where future large earthquakes are likely to occur globally. We divide the Earth's surface into $l° \times l°$ cells in order to construct the forecast and evaluate its success. The input for our forecast is a relative intensity (RI) map of $m \geq m_c$ earthquakes that occur in the cells during a specified training period, ΔT_{tr}. As a measure of RI we have utilized the cumulative Benioff strain in each cell during that period. The cumulative Benioff strain has been computed using the data from the CMT (Harvard) catalog (http://www.globalcmt.org). This catalog is complete from 1/1/1976 to 1/1/2008 to magnitude $m_c = 5.5$. The RI map is then converted to a binary forecast by introducing a threshold cumulative Benioff strain. Cells with Benioff strains greater than this threshold constitute *alarm cells* where future earthquakes are forecast to occur. A high threshold reduces the forecast areas but leads to more events that are not predicted. A low threshold reduces the failures to predict but increases the false alarms. We have introduced a standard optimization procedure for binary forecasts in order to select the optimal threshold.

In order to test our method we have performed a retrospective analysis. To define training and forecasting time periods we subdivide the catalog using a base year T_b, a forecasting year T_f, and an end year T_e.

---

[1] Department of Earth Sciences, University of Western Ontario, London, ON N6A 5B7, Canada.
E-mail: rshcherb@uwo.ca; ktiampo@seis.es.uwo.ca
[2] Department of Geology, University of California, Davis, CA 95616, USA. E-mail: turcotte@geology.ucdavis.edu
[3] Department of Physics, University of California, Davis, CA 95616, USA. E-mail: rundle@physics.ucdavis.edu

We have utilized seismicity during the training period from $T_b = 1/1/1976$ to $T_f = 1/1/2004$ ($\Delta T_{tr} \equiv T_f - T_b = 28$ years) to make a forecast for the locations of future target earthquakes of a specified magnitude, $m_T \geq 7.2$. The locations of earthquakes during the forecasting period $T_f = 1/1/2004$ to $T_e = 1/1/2008$ ($\Delta T_{for} \equiv T_e - T_f = 4$ years) are then used to test the validity of the method. The success of the binary forecast has been evaluated using techniques developed in the atmospheric sciences (MASON, 2003). The evaluation is in terms of a hit rate (fraction of events that are successfully forecast) versus a false alarm rate (the fraction of forecast cells in which earthquakes do not occur). Both hit rate and false alarm rate are functions of the tuning parameter (in our case the threshold Benioff strain). Their parametric plot produces a receiver operating characteristic (ROC) diagram. This approach has been used by HOLLIDAY et al. (2005) to test the success of forecasting algorithms in California. Similar analysis was also applied to the verification of the intermediate-term earthquake prediction algorithm M 8 by MINSTER AND WILLIAMS (1999).

We have also quantified our approach using the variation of the error diagram introduced by MOLCHAN (1991), MOLCHAN AND KEILIS-BOROK (2008). The comparison of the error diagram approach and the maximum likelihood method for evaluation of earthquake forecasting methods was given by KAGAN (2007). The connection between the entropy score and log-likelihood estimates with applications to several seismicity models is given in HARTE AND VERE-JONES (2005). The analysis of the error diagram approach applied to California seismicity was performed by ZECHAR AND JORDAN (2008). KOSSOBOKOV AND SHEBALIN (2003) and ZALIAPIN et al. (2003) have applied the error diagram to test a number of earthquake predictions algorithms.

2. Relative Intensity of Global Seismicity

The proposed analysis is based on the RI of past seismicity where large target earthquakes with magnitudes greater than $m_T \geq 7.2$ are forecasted. The moment magnitude m of earthquakes has been calculated from the scalar moment according to the formula $m = \frac{2}{3}\log_{10}(M_0) - 10.7$, where M_0 is a scalar moment reported in the CMT catalog. The grid size plays an important role in this analysis. In this study the Earth's surface has been subdivided into $2° \times 2°$ cells

As a measure of the RI the cumulative Benioff strain has been computed in each cell for a specified training time period ΔT_{tr}. The cumulative Benioff strain B up to time t is defined as

$$B_{xy}(t) = \sum_{i=1}^{N_{xy}(t)} \sqrt{E_{xy}^{(i)}}, \qquad (1)$$

where $E_{xy}^{(i)}$ is the seismic energy release in the ith earthquake and $N_{xy}(t)$ is the number of earthquakes considered up to time t, and (xy) is the coordinates of a given cell. The seismic energy E is related to the seismic moment M_0 as $E \sim M_0$.

The normalized RI in a cell is given by B_{xy}/B_{max} where B_{xy} is the cumulative Benioff strain in the active cell under consideration and B_{max} is the Benioff strain for the most active cell. The cumulative Benioff strain is also normalized by the actual area of the corresponding cell measured in square kilometers.

In order to construct a binary map a decision threshold of RI is applied. All cells with intensities greater than the specified threshold are alarm cells and constitute the binary forecast. To illustrate our approach we consider a specific example. The analysis is based on the seismicity during a training period from $T_b = 1/1/1976$ to $T_f = 1/1/2004$ ($\Delta T_{tr} = 28$ years). The threshold for the normalized Benioff strain $B/B_{max} = 0.0417$ is used. The reason for selecting this particular threshold will be discussed in the next section. The alarm cells in which earthquakes are forecast to occur are illustrated in Fig. 1. During the entire time period ($T_b = 1/1/1976$ to $T_e = 1/1/2008$) there were 1,503 active cells (cells with at least one earthquake with magnitudes $m \geq 5.5$) out of total of 16,200 cells. Note that the area of the active cells depends on the chosen cell size and their geographic location. Also shown in Fig. 1 are the 34 earthquakes with $m_T \geq 7.2$ that occurred during the forecast period $T_f = 1/1/2004$ to $T_e = 1/1/2008$ ($\Delta T_{for} = 4$ years). The next step is to quantify the success of our binary forecast of the locations of these earthquakes.
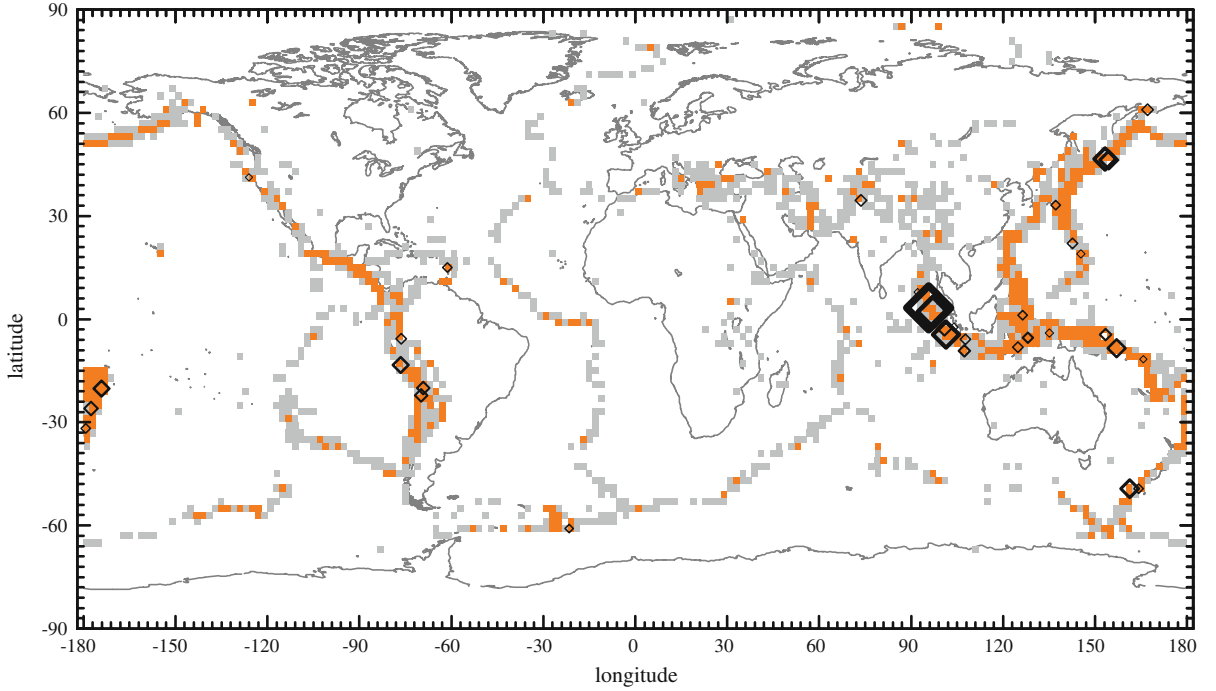
Figure 1
Spatial distribution of $2° \times 2°$ alarm cells. Earthquakes with magnitudes greater than $m \geq 5.5$ are considered during the time period 1/1/1976 until 1/1/2004 ($\Delta T_{\mathrm{tr}} = 28$ years). Also shown as *squares* are the locations of earthquakes with $m_T \geq 7.2$ that occurred during the forecast period $T_f = 1/1/2004$ to $T_e = 1/1/2008$ ($\Delta T_{\mathrm{for}} = 4$ years). The threshold value corresponds to the maximum Pierce's skill score given in Fig. 3a

3. Forecast Verification

The success of a forecast is based on maximizing the fraction of the earthquakes that occur in alarm cells and minimizing the fraction of the alarm cells in which no earthquake occurs. This analysis is carried out by utilizing a *contingency table* approach. Each cell in which an earthquake is forecast to occur constitutes an alarm cell. Earthquakes occur in some of these alarm cells but also in cells where alarms have not been issued. As a result of this test each cell falls in one of four categories: a, alarm and earthquake (successful forecast); b, alarm but not earthquake (false alarm); c, no alarm but an earthquake (failure to predict); d, no alarm and no earthquake (successful forecast of non-occurrence). In addition a is the number of target earthquakes that occurred in alarm cells, b is the number of alarm cells with no earthquakes in them, c is the number of target earthquakes that occurred outside alarm cells, and d is

the number of cells with no target earthquakes and which are not alarm cells.

As a specific example we consider the forecast given in Fig. 1. We retrospectively test this forecast in terms of the locations of $m_T \geq 7.2$ target earthquakes that occurred during the forecast period $T_f = 1/1/2004$ to $T_e = 1/1/2008$. The total number of $2° \times 2°$ cells in which a $m \geq 5.5$ occurred during the entire time period (active cells) was 1,503. During the forecast period $a + c = 34$ earthquakes with $m_T \geq 7.2$ occurred. Of these $a = 23$ occurred in alarm cells (were forecast) and $c = 11$ occurred in the other cells (were not forecast).

The success of a forecast can be quantified using a hit rate, H, and a false alarm rate, F:

$$H = \frac{a}{a+c}, \qquad F = \frac{b}{b+d}. \qquad (2)$$

The hit rate H is the fraction of earthquakes that are successfully forecast. For this example we have

$H = 23/34 = 0.676$. The false alarm rate F is the fraction of cells with no earthquake for which an alarm was issued. The hit rate H and the false alarm rate F define conditional probabilities $H = P(A|E)$ and $F = P(A|\text{not } E)$ respectively, where A denotes an event of having an alarm in given cells and E denotes an event of having a target earthquake in a given cell. For this example we have $F = 371/1469 = 0.253$. The result of a forecast is given in a contingency table (Table 1).

The above example has given the values of H and F for the specific value of the tuning parameter $B/B_{\max} = 0.0417$. The plot of H and F for a range of $B/B_{\max}$ values is known as a ROC diagram. An ROC diagram for our forecast is shown in Fig. 2a. Results are given for large earthquakes with magnitudes $m_T \geq 7.2$. The diagonal line from (0.0) to (1.1) would be the long-run expectation for alarms that are declared randomly. A perfect forecast would have $F = 0$ (no false alarms) and $H = 1$ (all earthquakes are forecast).

The choice of the optimal forecast (preferred value of $B/B_{\max}$) is the subject of extensive discussions in the atmospheric science literature (MASON, 2003). Several quantitative measures find wide use. One of these is *Pierce's skill score*:

$$\text{PSS}_{\text{ROC}} = H - F = \frac{ad - bc}{(a+c)(b+d)}, \quad (3)$$

which is the difference between the observed hit rate H and the value on the diagonal line corresponding to the random forecast. This score is also known as *Hansen-Kuiper's score* or *Kuiper's performance index* (MASON, 2003; HARTE AND VERE-JONES, 2005). The history and applications of this score along with its properties can be found in MASON (2003).

Table 1

Contingency table for earthquakes with $m_T \geq 7.2$ that occurred during the forecast period $T_f = 1/1/2004$ to $T_e = 1/1/2008$ ($\Delta T_{for} = 4$ years)

Forecast	Observed		
	Yes	No	Total
Yes	$a = 23$	$b = 371$	$a + b = 394$
No	$c = 11$	$d = 1098$	$c + d = 1109$
Total	$a + c = 34$	$b + d = 1469$	$a + b + c + d = 1503$

A perfect forecast would be $\text{PSS}_{\text{ROC}} = 1.0$. The values of PSS_{ROC} corresponding to the results in Fig. 2a are given in Fig. 3a. They are plotted as a function of the tuning parameter values $B/B_{\max}$ (Fig. 2a).

The maximum value of $\text{PSS}_{\text{ROC}}^{\max} = \text{PSS}_{\text{ROC}}$ from an ROC diagram can be used to find an optimizing value of the tuning parameter. For large earthquakes $m_T \geq 7.2$ the maximum value of Pierce's skill score is $\text{PSS}_{\text{ROC}}^{\max} = 0.424$ corresponding to the threshold value $B/B_{\max} = 0.0417$. This corresponds to $H = 0.676$ and $F = 0.253$ and was the basis for selecting this threshold value for the example given in Fig. 1. The value of the hit rate $H = 0.676$ (67.6%) gives the probability of having a large earthquake in the alarm cells during the forecast period. The threshold values obtained from retrospective forecasts can be applied a priori in subsequent prospective forecasts. This can be done by analyzing a particular region of study where the magnitude of completeness of a catalog is obtained and the optimal threshold and cell size values were determined from the retrospective analysis.

Another method of measuring the success of a forecast is the *error* or *Molchan's diagram* approach (MOLCHAN, 1991; MOLCHAN and KAGAN, 1992; MOLCHAN and KEILIS-BOROK, 2008). In order to construct the error diagram the miss and alarm rates

$$v = \frac{c}{a+c} = 1 - H, \qquad \tau = \frac{a+b}{a+b+c+d} \quad (4)$$

are introduced. The sum of the hit rate and miss rate is unity, i.e. $H + v = 1$. The alarm rate τ is the fraction of all cells that are alarm cells and defines the probability of a forecast of occurrence in weather forecast verification (MASON, 2003). The parametric plot of v and τ for a range of threshold parameters $B/B_{\max}$ is known as an *error diagram*. The error diagram for the above example is given in Fig. 2b. An analog of the Pierce's skill score for the error diagram is

$$\text{SS}_{\text{ED}} = 1 - v - \tau = \frac{ad - bc}{(a+c)(a+b+c+d)}. \quad (5)$$

From the analysis of the error diagram curve the maximum value of this skill score is $\text{SS}_{\text{ED}}^{\max} = 1 - 0.324 - 0.262 = 0.414$ corresponding to the threshold value $B/B_{\max} = 0.0417$. It is worthwhile to note
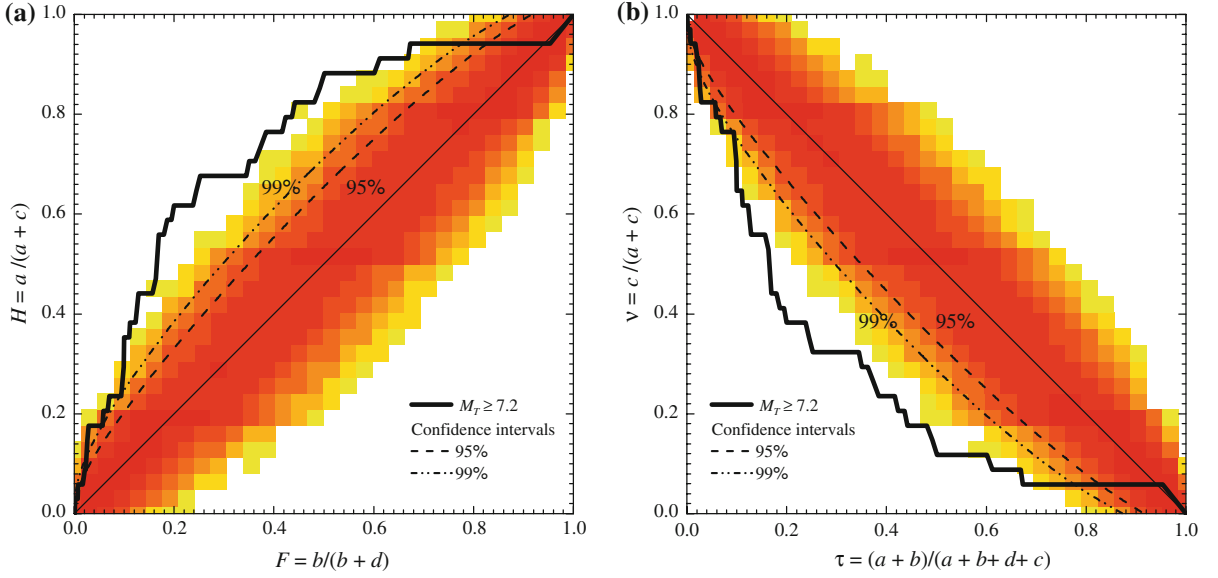
Figure 2

a The ROC diagram for our forecast of global seismicity. **b** The error diagram for our forecast. Also included the results of 150,300 Monte-Carlo simulations (*shaded area*) in which alarm cells have been redistributed randomly and the corresponding confidence intervals
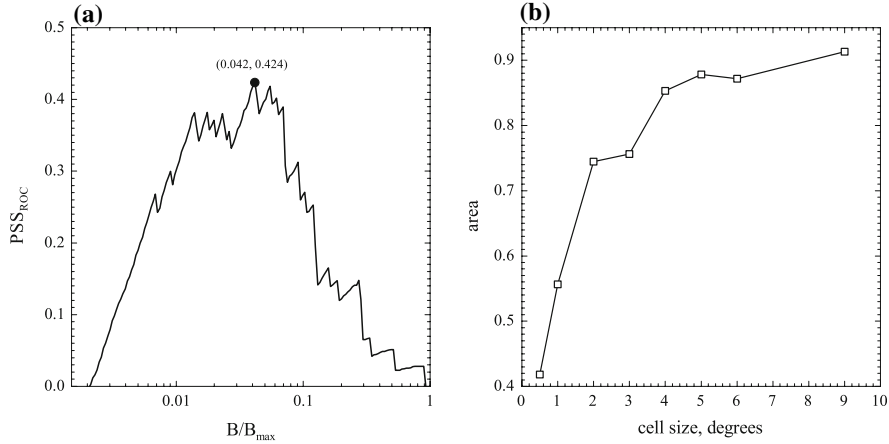

Figure 3

a Pierce's skill scores PSS_{ROC} from the ROC diagram given in Fig. 2a as a function of the threshold values of the normalized cumulative Benioff strain B/B_{max}. The maximum value of the Pierce's skill score is $PSS_{ROC}^{max} = 0.424$ corresponding to the threshold value $B/B_{max} = 0.042$. **b** Dependence of the area under ROC diagram on different cell sizes

that the threshold values corresponding to the maximum values of the skill scores for both ROC and error diagrams are the same. This is due to the fact that the values of F and τ are very close to each other for corresponding threshold values as they are both dominated by large $b + d$ values compared to small $a + c$. As a result the error diagram curve is almost a

mirror image of the ROC curve due to the fact that $v = 1 - H$. In general this is not always the case.

In the above analysis we have used a uniform prior distribution for the seismically active cells and assumed a null hypothesis where alarm cells are redistributed randomly. To reject this null hypothesis in favor of the method presented here we have also

included in Fig. 2a and b the results of 150,300 Monte-Carlo simulations in which alarm cells have been redistributed randomly and the same target earthquakes were used to construct ROC and error diagrams. The confidence intervals were evaluated using the exact analytical result for the binomial distribution given in ZECHAR and JORDAN (2008).

4. Discussion

The primary purpose of this paper is to quantify the extent to which locations of past smaller earthquakes can be used to forecast the locations of future large earthquakes. Our study has been carried out globally. The surface of the Earth has been divided into a grid of $l° \times l°$ cells with $l = 2°$. The objective is to forecast the locations of future large target earthquakes with magnitudes greater than m_T based on the past seismicity. Another important aspect of this work is to introduce different optimization procedures to find the best forecast and optimal values for model parameters. It is also important to note that in the presented analysis the measure of seismicity based on cumulative Benioff strain does not perform better than the predictor based on a long-run rate of occurrence of earthquakes.

The sensitivity of the method to the model parameters is an important issue. The cell size $l = 2°$ has been chosen based on the analysis of the areas under the ROC curves constructed for different cell sizes and different forecasting magnitudes. The analysis shows that the areas tend to increase (better forecast) with increasing cell sizes (Fig. 3b). Bigger cell sizes cover a larger Earth's surface and increase the total success rate. However, bigger cell sizes are less practical in forecasting as they include larger geographical regions and decrease the value of the forecast. In this analysis we assume all earthquakes are point events. Cell size can be an important factor if forecast earthquakes are defined through their rupture areas which can span several cell sizes.

Other important parameters are the durations of the training and forecasting intervals and the lower magnitude cutoff. The analysis of the areas under the ROC curves shows that increasing the training time interval (ΔT_{tr}) usually improves the forecast. As to

the forecasting time interval (ΔT_{for}) changes in the duration of the interval result in different numbers for forecasting earthquakes which, as was shown by ZECHAR AND JORDAN (2008), affects the ROC and error diagrams. To investigate the performance of the method for different forecasting time intervals a contour plot of the values of the areas between the ROC curve and 95% confidence interval for different forecasting years, T_f, and base years, T_b, is given in Fig. 4. Darker areas indicate a better performance with respect to the 95% confidence interval of the random forecast. It shows a prominent increase in the values for intervals less than 4 years and another increase for 15-year interval. This was a reason to use a 4-year forecasting interval in our analysis.

Finally, the dependence of the method on the lower magnitude cutoff, m_c, shows a strong trend in the performance increase of the method with decreasing m_c. However, the completeness of the catalog prevents the use of events smaller than $m < 5.5$. The full analysis of the dependence of the method on the above parameters will be presented elsewhere. By varying the threshold B/B_{max} the method generates a spectrum of forecasts. The use of the Pierce's skill score is probably the most accepted optimization tool, but several others have been proposed (MASON, 2003; HARTE AND VERE-JONES, 2005). An actual optimization should also be based on other
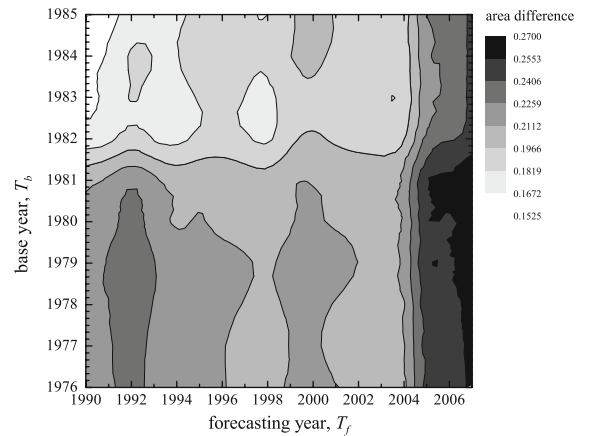

Figure 4
A contour plot of the values of the areas between the ROC curve and the 95% confidence interval for the random forecast for different forecasting (T_f) and base (T_b) years

factors such as available resources, population densities, quality of construction, and other risk factors.

REFERENCES

FIELD, E.H. (2007), *Overview of the working group for the development of Regional Earthquake Likelihood Models (RELM)*, Seismol. Res. Lett. *78*, 7–16.

FRANKEL, A., MUELLER, C., BARNHARD, T., PERKINS, D., LEYENDECKER, E., DICKMAN, N., HANSON, S., and HOPPER, M. (1996), *National seismic-hazard maps*, Tech. Rep. Open-File Report 96–532, USGS.

HARTE, D. and VERE-JONES, D. (2005), *The entropy score and its uses in earthquake forecasting*, Pure Appl. Geophys. *162*, 1229–1253.

HOLLIDAY, J.R., NANJO, K.Z., TIAMPO, K.F., RUNDLE, J.B., and TURCOTTE, D.L. (2005), *Earthquake forecasting and its verification*, Nonlinear Proc. Geophys. *12*, 965–977.

KAGAN, Y.Y. (2007), *On earthquake predictability measurement: Information score and error diagram*, Pure Appl. Geophys. *164*, 1947–1962.

KOSSOBOKOV, V.G., KEILIS-BOROK, V.I., TURCOTTE, D.L., and MALAMUD, B.D. (2000), *Implications of a statistical physics approach for earthquake hazard assessment and forecasting*, Pure Appl. Geophys. *157*, 2323–2349.

KOSSOBOKOV, V.G. and SHEBALIN, P., *Earthquake prediction*. In *Nonlinear Dynamics of the Lithosphere and Earthquake Prediction* (eds. KEILIS-BOROK, V.I., AND SOLOVIEV, A.A.) (Springer-Verlag, Berlin 2003), pp. 141–207.

MASON, I.B., *Binary events*. In *Forecast Verification* (eds. JOLLIFFE, I.T. and STEPHENSON, D.B.) (Wiley, Chichester 2003), pp. 37–76.

MINSTER, J.-B. and WILLIAMS, N., *Systematic global testing of intermediate-term earthquake prediction algorithm*. In *1-st ACES Workshop Proceedings* (ed. MORA, P.) (GOPRINT, Brisbane 1999).

MOLCHAN, G. and KEILIS-BOROK, V. (2008), *Earthquake prediction: Probabilistic aspect*, Geophys. J. Int. *173*, 1012–1017.

MOLCHAN, G.M. (1991), *Structure of optimal strategies in earthquake prediction*, Tectonophysics *193*, 267–276.

MOLCHAN, G.M. and KAGAN, Y.Y. (1992), *Earthquake prediction and its optimization*, J. Geophys. Res. *97*, 4823–4838.

SCHORLEMMER, D., GERSTENBERGER, M.C., WIEMER, S., JACKSON, D.D., and RHOADES, D.A. (2007), *Earthquake likelihood model testing*, Seismol. Res. Lett. *78*, 17–29.

ZALIAPIN, I., KEILIS-BOROK, V., and GHIL, M. (2003), *A Boolean delay equation model of colliding cascades. Part II: Prediction of critical transitions*, J. Stat. Phys. *111*, 839–861.

ZECHAR, J.D. and JORDAN, T.H. (2008), *Testing alarm-based earthquake predictions*, Geophys. J. Int. *172*, 715–724.