Systematic global testing of intermediate-term earthquake prediction algorithms

Jean-Bernard Minster and Nadya Williams

Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California Sand Diego, La Jolla California, 92093-0225 (e-mail: jbminster@ucsd.edu; phone (619) 534-5650; fax (619) 534-2902).

Abstract

The systematic testing of intermediate-term earthquake prediction algorithms is faced with the difficulties of (1) the limited time spanned by instrumental catalogs of seismicity in most areas of the world, and (2) the relative arbitrariness of the selection of the region and time interval over which to apply the algorithm. We offer a systematic method which uses worldwide seismicity to compute the *Receiver Operating Characteristic* (ROC) curve of such an algorithm and permits a quantitative test of the null hypothesis that the algorithm samples space time no better than a random process with uniform probability distribution. We use the "M8" algorithm as an example.

Introduction

The *Receiver Operating Characteristic* (ROC) curve for an arbitrary detector is simply a plot of a measure of success of the detector in terms of the fraction of signals correctly detected, as a function of the volume of parametric space in which a signal is claimed to exist. For a prediction algorithm we plot the percentage of earthquakes for which an alarm had been declared, versus the total volume of space-time in which a state of alarm exists. Clearly, it is desirable to predict successfully as large a percentage of earthquakes in as small a space-time volume as possible.

The "M8" algorithm, originally developed for intermediate-term prediction of large events, uses a catalog of mainshocks to identify large scale seismicity patterns before large earthquakes in a given region (e.g. *Gabrielov et al.*, 1986; *Keilis-Borok et al.* 1988; *Keilis-Borok et al.*, 1990; *Updyke et al.,* 1989; *Healy et. al*, 1992; *Kossobokov et al.*, 1992, *Keilis Borok and Rotwain,* 1994; *Kossobokov and Mazhkenov*, 1994). We have developed an approach to measure the stability of the results, and to test specific hypotheses (*e.g. Minster and Williams*, 1992, 1994, 1995, 1996), and used it to assess the performance of M8 intermediate-term earthquake prediction algorithm. The parameter space of the algorithm is explored in a random way by repeated application of the algorithm. To test null-hypotheses we apply the same algorithm to randomized situations. Thus we define time- and space-dependent bootstrap likelihood functions, and test the "true" catalog results against them. We applied the algorithm to Poissonian catalogs derived by re-sampling of the "true" catalog (with replacement). These catalogs allow to preserve a spatial structure

and temporal micro-structure, while destroying any long-term trends or possible systematic catalog errors.

In this report, we have extended the analysis to the entire worldwide catalog used by *Healy et al.* (1995). This catalog has been "de-clustered", that is aftershocks have been counted and removed by these authors, and we use the resulting mainshock catalog without any modification.

Objective

We seek to construct the algorithm's *Characteristic Receiver Operating Curve* (ROC), as an objective measure of success. The ROC is defined as the proportion of successful predictions in the chosen region as a function of the volume of space-time during which an alarm is in effect. We compute it for the entire world catalog.

Methodology

We proceed in successive steps:
1) For this test, we used the worldwide catalog of mainshocks (1963-1998), with a magnitude range $M \geq 4.0$, which permits us to test "M8" for the earthquakes of magnitude $M_0 \geq 7.5$.

2) We tile the world with a dense rectangular grid on equal area projection, select a target date, and run "M8" for each grid node, using both observed and randomized catalogs. For each node where "M8" may be executed, we then perform N additional runs (e.g. $N = 5$ or 10), randomizing the circle center over one grid cell, randomizing the radius by $\pm 10\%$, and randomizing the start-time for the algorithm in a 2-year window. The circles cover the whole Earth in such a way that each point at the surface of the Earth is covered by approximately 900 circles, 80 percent of which have slightly randomized parameters.

3) For each node, we define a likelihood - the number of circles with a current *Time of Increased Probability (TIP)* at the target date, as a fraction of the number of circles which include this node.

4) We then use a decision criterion for prediction, and test it against a null hypothesis.

Approach

1) In order to produce the *ROC* curve, it is necessary to define carefully the "space" covered by the predictions. Upon careful consideration, we argue that "space" must be defined as a complete set of points included in at least one circle where the algorithm can actually be executed. "Time" is the period 1985-1998, over which the algorithm has stabilized.

2) *Hypothesis Test:* In order to formulate a hypothesis test, we first settle on a *Decision Rule* R. A prediction is successful if the current likelihood $\mathbf{L}(t)$ exceeds a threshold $\mathbf{L_o}$. The *Null Hypothesis* is then stated as follows:

- *Given the decision rule* R, the algorithm samples the space-time volume according to a uniform probability density.
- The *Alternate Hypothesis* may then be expressed in the form:
- *Given the decision rule* R, the algorithm samples the space-time volume according to a non-uniform (perhaps very concentrated if the algorithm is efficient) probability density.

3) *ROC Curve:* The desired characteristic shape of the *ROC* curve has upward curvature above the prime diagonal, indicating good prediction performance for a small fraction of the space-time volume. The *ROC* is constructed by reckoning the prediction success rate as a function of the space-time volume $V(L > L_o)$ during which an alarm is declared.

Testing the algorithm

It is clear that, if one selects a very strict decision rule, such that only an infinitesimal fraction of space-time is in a state of alarm, then a very small fraction of events will be predicted. Similarly, if one declares an alarm everywhere in space and time, then all events are predicted. These two extreme situations are useless for practical purposes, but serve to demonstrate that the *ROC* must join the origin to the upper right hand corner of the graph.

Under the null hypothesis of random sampling of space-time according to a uniform probability density, the fraction of earthquakes correctly predicted is statistically proportional to the fraction of space-time in which an alarm is declared, and the *ROC* should lie close to the prime diagonal of the graph.

It is an easy matter to show that the distribution that governs departures from the diagonal under repeated realizations is a negative binomial distribution, which permits straightforward construction of confidence regions. If the observed *ROC* based on the actual catalog remains within the confidence region, then the null hypothesis cannot be rejected.

The test of "M8" over the worldwide catalog is in progress. Each run takes approximately 20 hours of computer time on a high-end workstation capable of executing the algorithm about 50 times per second. Figures 1 and 2 show preliminary results presented by Minster and Williams (1996) for the period 1985-96. In order to reduce computer time, these runs neglect the addition of random circles whenever the seismicity is too low to run M8. This leads to a speed-up by a factor of 4 for the entire world, but has the unfortunate consequence of leading to over estimates of the likelihoods in regions of low likelihood, biasing the test toward a better apparent performance of the algorithm. Nonetheless, the figures illustrate several major characteristics that seem robust.

There were 38 events worldwide in the catalog, with $M_0 \geq 7.5$ for this 11 year period. For three of these events, the surrounding seismicity was too low to even permit execution of M8, and these events were not included in the statistics. Figure 1 shows the "space" dimension used in the test as defined above, and Figure 2 shows the *ROC* curve, and the 95% confidence bounds determined from the hypergeometric distribution. We find 3 regimes for remaining 35 events:

1) With $L_0 \geq 20\%$, M8 performs better than random at nearly 95% confidence level for 11 events; this represents a 31% success rate, over 25% of space-time. At these

thresholds, M8 performs temporarily better than random in the short run, but acceptance of the algorithm requires more data in the long run.

2) For 17 events, for values $L_0 < 20\%$, a 31-80% success rate over 25-75% of space-time, the null-hypothesis cannot be rejected, and M8 performs no better than random.

3) For 7 events no successful prediction is possible at *any* likelihood threshold. These events pull the *ROC* curve in the rejection region at large thresholds: one has to declare an alarm for all space and all time in order to predict these events successfully.

This suggests that the performance of "M8" is affected by factors not recognized explicitly by the algorithm, and therefore not incorporated in our test. Examination of the moment tensor solutions does not reveal any systematic relationships between the events and their position on the graph. Additional testing in progress will shed light on possible biases in this assessment.
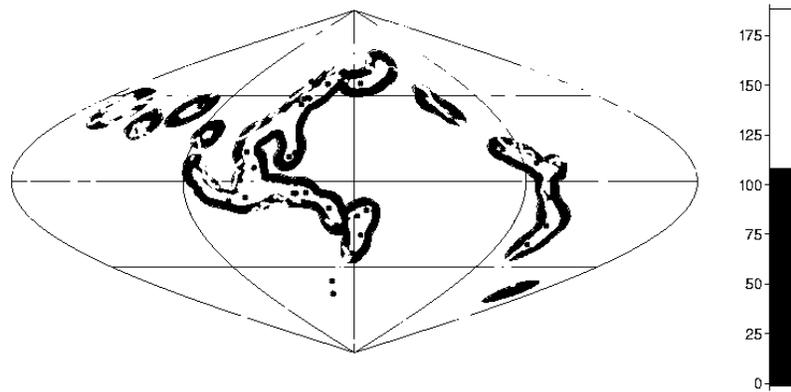

Figure 1: Equal area map showing the 38 large events comprising the worldwide test of M8 for 1985-96. The colored area shows the ensemble of points that are included in the "space" component of the "space-time" domain used to test M8. The two events along the Macquarie Ridge and one on Sakhalin Island (shown as red circles) are removed from the statistical sample, since they are not included in the "space".
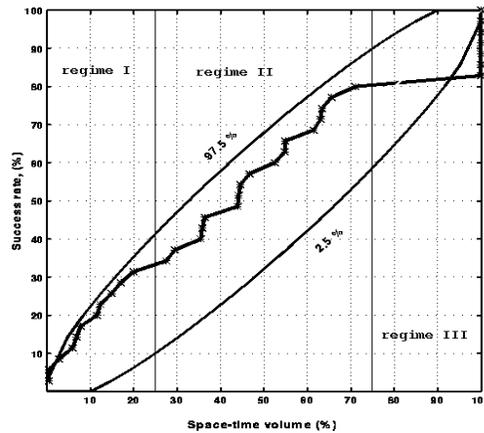
494

Figure 2: Receiver Operating Characteristic Curve calculated for "M8" for the 35 events with $M \geq 7.5$ (1985-1996) shown in Figure 1. The two smooth curves labeled 2.5% and 97.5% define a 95% confidence interval based on the null hypothesis that the algorithm samples the space-time according to a uniform probability density. An *ROC* curve excursion above this region is diagnostic of an algorithm that performs better than random. Similarly, an *ROC* curve that wanders below this region indicates an algorithm performance worse than random. This diagram shows the population of events used here to be tri-modal in terms of M8 performance. For regime I (11 events) M8 performs somewhat better than random; for regime II (17 events) M8 performs no better than random, and for regime III (7 events) no successful prediction could be made.

Acknowledgments

References

[1] Gabrielov A.M. et al., *Algorithms of long-term earthquake prediction, Int. School for research oriented to earthquake prediction algorithms*, software and data handling, CERESIS, Lima, 1986

[2] Gardner J., and L. Knopoff, *Is the sequence of earthquakes in southern California with aftershocks removed Poissonian? Yes*, Bull. Seismol. Soc. Amer., 64, 1363-1367, 1974.

[3] Healy, J.H., V.G. Kossobokov, and J.W. Dewey, *A test to evaluate the earthquake prediction algorithm, M8*, USGS Open-file report, in press, 1992

[4] Keilis-Borok, V.I., L. Knopoff, V.G. Kossobokov, and I.M. Rotvain, *Intermediate-term prediction in advance of the Loma Prieta earthquake*, Geophys. Res. Lett., 17, 1461-1464, 1990.

[5] Keilis-Borok, V.I., and I.M. Rotwain, *Diagnosis of times of increased probability (TIPs) for strong earthquakes in Northern Appalachians*, Comp. Seismol. and Geodyn., 1, 1-5, 1994.

[6] Kossobokov, V.G. and M.G. Mazhkenov, *Times of increased probability of large earthquake in the eastern Tien Shan diagnosed by the M8 algorithm*, Comp. Seismol. and Geodyn., 1, 16-19, 1994.

[7] Kossobokov, V.G., J.H. Healy, and V.I. Keilis-Borok, *Loma Prieta earthquake prediction by the M8 algorithm*, (Preprint) 1992

[8] Kossobokov V.G., and J.H. Healy, *Testing an earthquake prediction algorithm*, Science submitted, 1994

[9] Keilis-Borok, V.I., L. Knopoff, and I.M. Rotvain, *Bursts of aftershocks long term precursors of strong earthquakes*, Nature, 283, 259-263, 1980.

[10] Minster, J.B., and N.P. Williams, *The "M8" intermediate-term earthquake prediction algorithm: an independent assessment*, EOS Am. Geophys. Union Trans. Oct. 1992.

[11] Minster, J.B., and N.P. Williams, *A systematic approach to assess M8-class intermediate-term earthquake prediction algorithms*, IASPEI 27th General Assembly, Wellington, N.Z. Abstr. S6-21 1995a.

[12] Minster, J.B., and N.P. Williams, *Performance of the M8 Intermediate-term earthquake prediction algorithm, assessed from bootstrap estimates*, IUGG XXI Gen. Assembly, Abstr. A-401. 1995

[13] Minster, J.B., and N.P. Williams, *Worldwide performance of a seismicity-based intermediate-term prediction algorithm: The M8 algorithm, M>7.5, 1985-94*, SCEC Annual Meet. p.74 Oct. 1995

[14] Minster, J.B., and N.P. Williams,, *Intermediate-term earthquake prediction algorithms*, Southern California Earquake Center Annual meeting, 1996.

[15] Molchan G.M., and O.E. Dmitrieva, *Aftershock identification: methods and new approaches*, Geophys. J. Intern., 109, 501-516,1992.

[16] Sornette, D. and C.G. Sammis, *Universal log-periodic correction to renormalization group scaling for regional seismicity: Implications for earthquake prediction.* SCEC annual meeting, 1994, preprint.

[17] Updyke et al., Proc. NEPEC June 6-7, 188, Reston VA, USGS Open-file report 89-144, 1989