The Abdus Salam
International Centre for Theoretical Physics

SMR.1676 - 45

8th Workshop on Non-Linear Dynamics and Earthquake Prediction

3 - 15 October, 2005

Earthquake Prediction : Verification Problem

Vladimir G. Kossobokov
Russian Academy of Sciences
International Inst. of Earthquake Prediction Theory and Matematical Geophysics
117556 Moscow
Russia
&
Institut de Physique du Globe de Paris
4 Place Jussieu 75252 Paris
France

Strada Costiera 11, 34014 Trieste, Italy - Tel. +39 040 2240 111; Fax +39 040 224 163 - sci_info@ictp.it, www.ictp.it

Earthquake prediction: Verification Problem

V. Kossobokov[1,2]

INSTITUT DE PHYSIQUE DU GLOBE DE PARIS

International Institute of Earthquake Prediction Theory and Mathematical Geophysics,
Russian Academy of Sciences,
79-2 Warshavskoye Shosse, Moscow 113556, Russian Federation

Institute de Physique du Globe de Paris,
4 Place Jussieu, 75252 Paris, Cedex 05, France

E-mail: volodya@mitp.ru or volodya@ipgp.jussieu.fr

Precursors and Prejudices

Although many observations reveal unusual changes of geophysical fields at the approach of a large earthquake, most of them report a unique case history and lack a systematic description (*Wyss, 1991*). The later makes an earthquake prediction method hardly reproducible and, therefore, testable by an independent investigator even in cases, when such a method has been proposed long time ago and its post the fact applications are subject of numerous publications in prestigious scientific journals.

“The analysis of data inevitably involves some trafficking with the field of *statistics*, that gray area which is not quite a branch of mathematics - and just as surely not quite a branch of science. In the following sections, you will repeatedly encounter the following paradigm:

- apply some formula to the data to compute "a statistic"
- compute where the value of that statistic falls in a probability distribution that is computed on the basis of some "null hypothesis"
- if it falls in a very unlikely spot, way out on a tail of the distribution, conclude that the null hypothesis is *false* for your data set

If a statistic falls in a *reasonable* part of the distribution, you must not make the mistake of concluding that the null hypothesis is "verified" or "proved". That is the curse of statistics, that it can never prove things, only disprove them! At best, you can substantiate a hypothesis by ruling out, statistically, a whole long list of competing hypotheses, every one that has ever been proposed. After a while your adversaries and competitors will give up trying to think of alternative hypotheses, or else they will grow old and die, and *then your hypothesis will become accepted*. Sounds crazy, we know, but that's how science works!”

(William H. Press et al., *Numerical Recipes*, p.603)

What is earthquake prediction?

The United States National Research Council, Panel on Earthquake Prediction of the Committee on Seismology suggested the following definition (1976, p.7):

"An earthquake prediction must specify the expected magnitude range, the geographical area within which it will occur, and the time interval within which it will happen with sufficient precision so that the ultimate success or failure of the prediction can readily be judged. Only by careful recording and analysis of failures as well as successes can the eventual success of the total effort be evaluated and future directions charted. Moreover, scientists should also assign a confidence level to each prediction."

Stages of earthquake prediction

- Term-less prediction of earthquake-prone areas
- Prediction of time and location of an earthquake of certain magnitude

Temporal, ***in years***		Spatial, ***in source zone size L***	
Long-term	10	Long-range	up to 100
Intermediate-term	1	Middle-range	5-10
Short-term	0.01-0.1	Narrow	2-3
Immediate	0.001	Exact	1

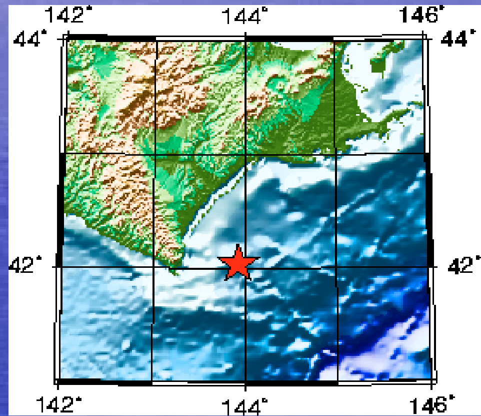
- Moreover, the Gutenberg-Richter law suggests limiting magnitude range of prediction to about one unit. Otherwise, the statistics would be essentially related to dominating smallest earthquakes.

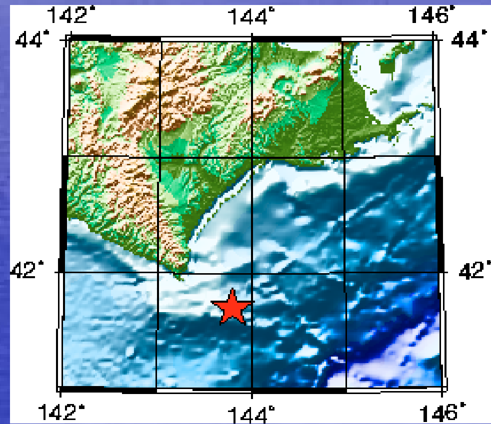
PLANETS ALIGN:
On Wednesday morning, September 24th, 2003 a lovely trio appeared in the eastern sky: **Jupiter, the crescent moon and Mercury…**

Is it a coincidence or a law?

Two days later …

防災科研Hi-net暫定処理による震源位置





本震
2003年9月26日04時50分11秒
北緯42.0度
東経143.9度
深さ25km
M7.7

最大余震
2003年9月26日06時08分03秒
北緯41.8度
東経143.9度
深さ35km
M7.4

Seismic Roulette

Consider a roulette wheel with as many sectors as the number of events in a sample catalog, a sector per each event.

- Make your bet according to prediction: determine, which events are inside area of alarm, and put one chip in each of the corresponding sectors.
- Nature turns the wheel.
- If seismic roulette is not perfect…

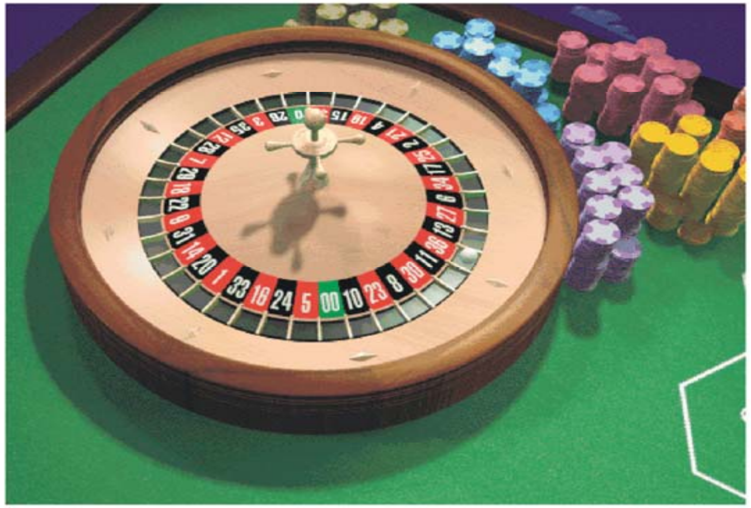
then ***systematically*** you can win! ☺

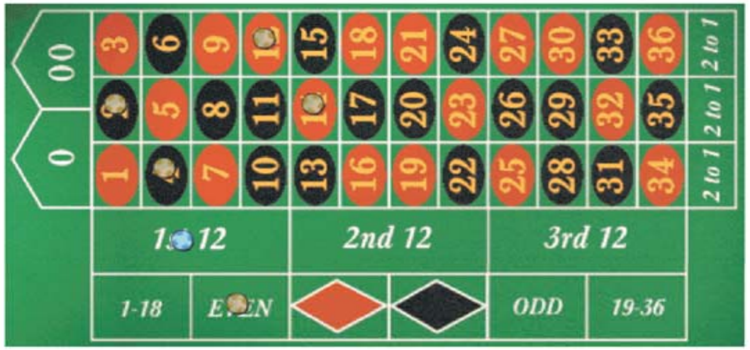
and lose … ☹

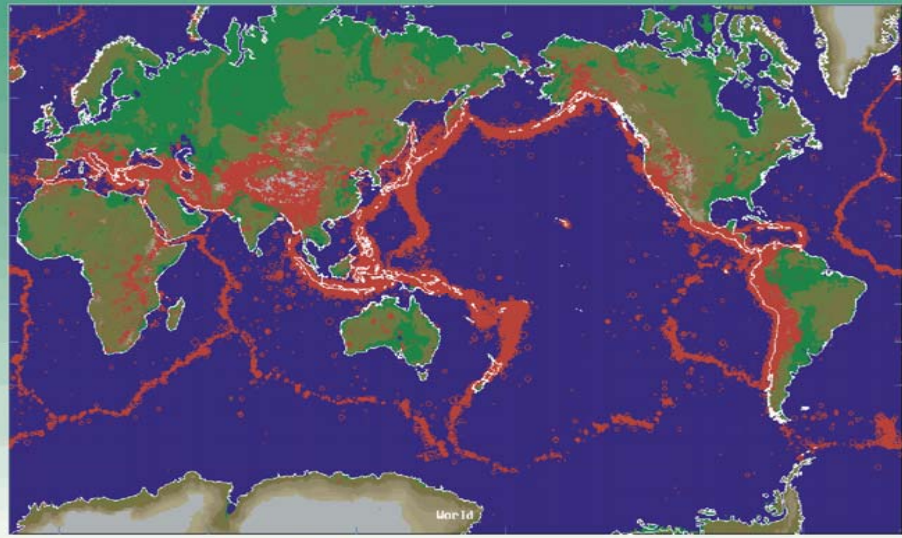
If you are smart enough and your predictions are effective ------

the first will outscore the second! ☺ ☺ ☹ ☺ ☺ ☺ ☹ ☺ ☺ ☺

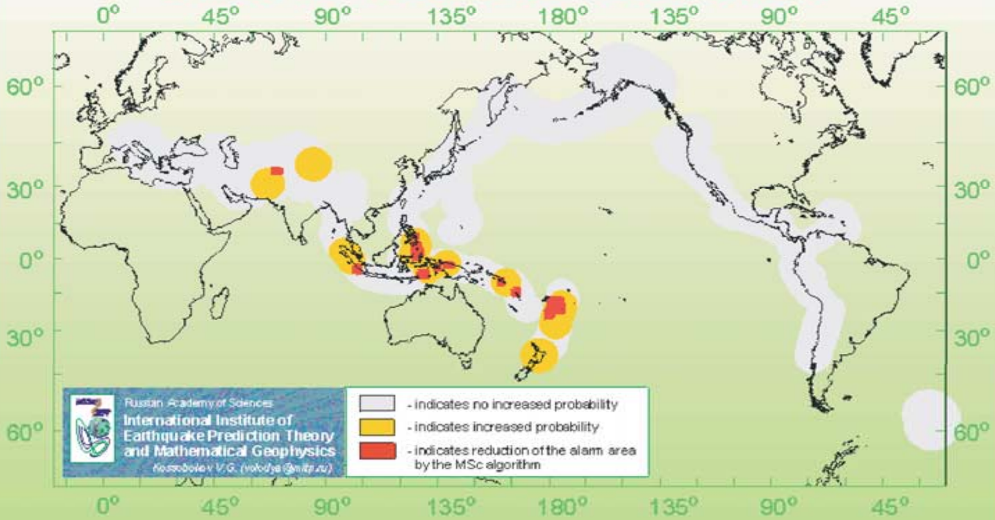
Seismic Roulette



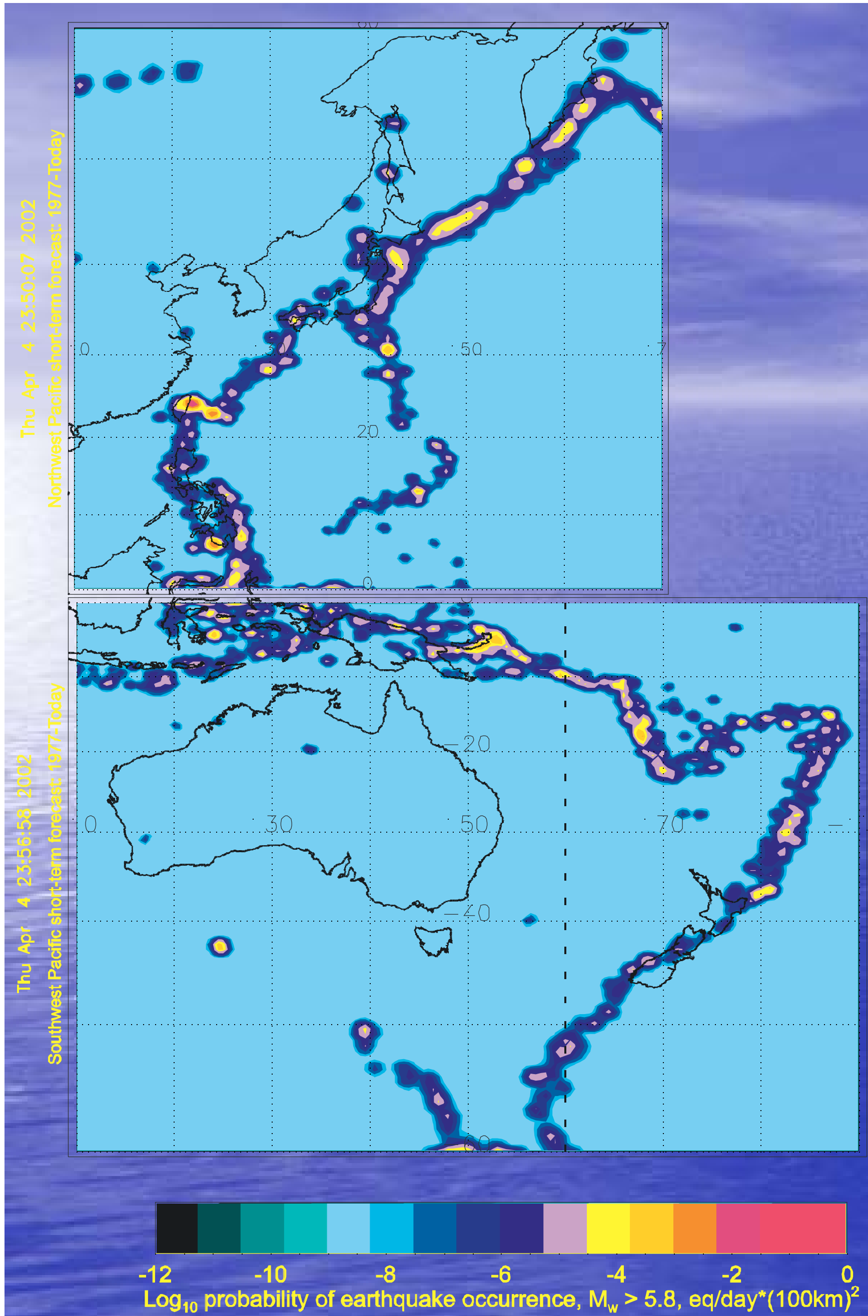




Regions of Increased Probability of Magnitude 8.0+ Earthquakes
as on July 1, 2000 (*subject to update on January 1, 2001*)



West Pacific short-term forecast



Is it prediction?

- No, unless you are a little bit more specific…

Is it short-term?

- Seems not…

Can anyone evaluate its effectiveness?

- Yes, if provided a track record of the experiment.

Jackson and Kagan "Testable earthquake forecasts for 1999", Seism. Res. Lett., 70, 393-403, 1999

Kagan and Jackson (2000) "Probabilistic forecasting of earthquakes", Geophys. J. Int., 143, 438-453

We have analyzed the predictions arising from setting a threshold probability or a threshold probability ratio on top the daily updated Short-term forecasts for NW and SW Pacific in April 2002 - September 2004 (http://scec.ess.ucla.edu/~ykagan/predictions_index.html; Kagan and Jackson, 2000. Probabilistic forecasting of earthquakes, Geophys. J. Int., 143, 438-453) and the catalog of earthquakes for the same period and have come to the following conclusion:

The predictions based on the Yan Y. Kagan and David D. Jackson forecasts are hardly better than random guessing, when main shocks are considered, and could be used for effective prediction of aftershocks only.

The conclusion is based on the prediction outcome achieved for 218 shallow (with depth less than 70 km) earthquakes of MwHRV = 5.8 or more. According to the definition from (*Keilis-Borok et al., 1980*), there are 67 aftershocks and 151 main shocks.

The territory of West Pacific short-term forecast is coarse-grained into cells, 0.5 by 0.5 degree each. Making a “bet” on a cell C, we pay n(C), which is the number of earthquakes from the sample catalog. Each target earthquake E defines the threshold value - p(E) (or p/P(E)) - being the value of short-term probability p (or the value of probability ratio p/P) determined in advance for the day of the earthquake.

In its turn the threshold defines the minimal cost of a bet required for successful prediction of the target earthquake, N(E), which is the sum of all bets n(C) over the union of cells with p equal or above p(E) (same for the ratio p/P). The track record of the experiment provides the set of bets {N(E)} associated with target earthquakes that happened.

Denote μ being the bet sum normalized to the total sum of n(C) and ν being the number of failures-to-predict normalized to the total number of target earthquakes that happened in the course of testing. The ν vs. μ diagram characterize the effectiveness of the prediction method, e.g., random prediction performance is associated with the diagonal that connects "optimist's" {1,0} and "pessimist's" {0,1} strategies (Molchan, G. M.. Earthquake Prediction as a Decision-making Problem, Pure Appl. Geophys., 149, 233-247, 1997).

Given -

(1) the track record of the West Pacific short-term forecasts in the period from April 10, 2002 to September 13, 2004;

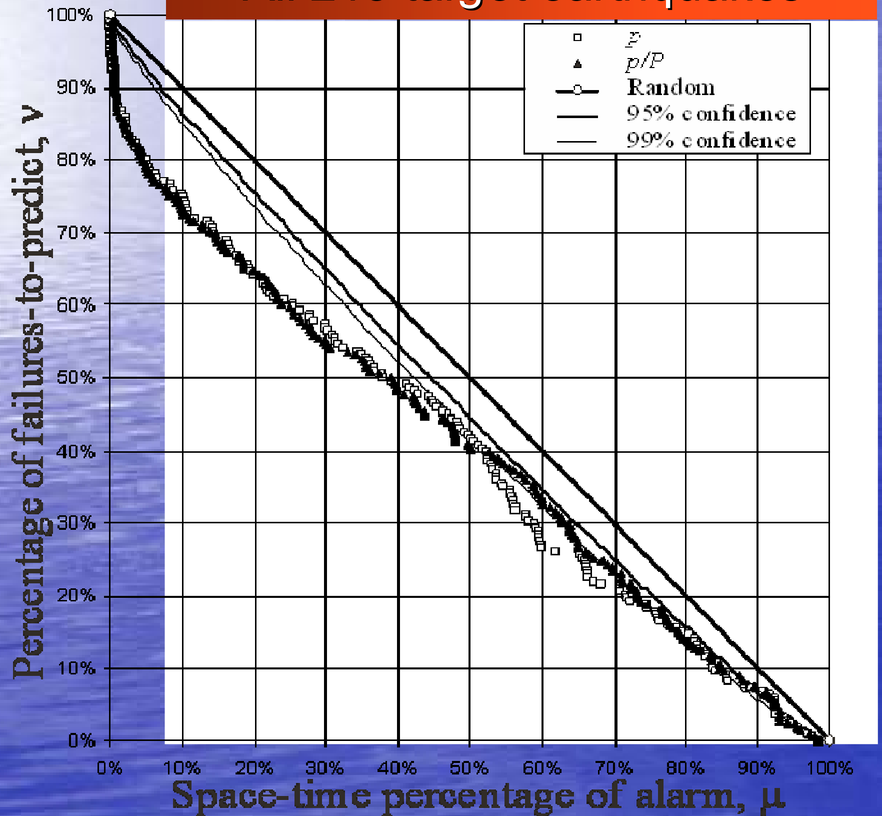
(2) the Harvard CMT catalog for the same period of time;

(3) the counts of n(C) based on the NEIC catalog of shallow earthquakes -

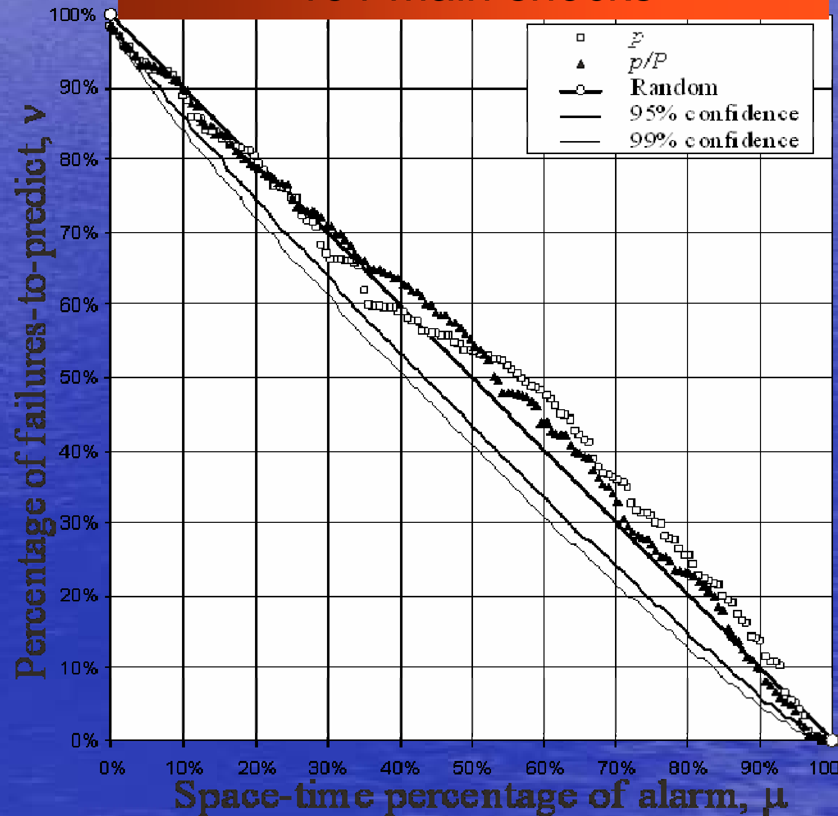
we plotted several ν vs. μ diagrams.

The two figures show the performance of predictions based on p or p/P in the test period from April 10, 2002 to September 13, 2004. The total of 218 earthquakes of magnitude Mw = 5.8 or more with the depth of 70 km or shallower occurred in the West Pacific. According to definition from (*Keilis-Borok et al., 1980*), 67 of them are aftershocks and 151 main shocks.

All 218 target earthquakes

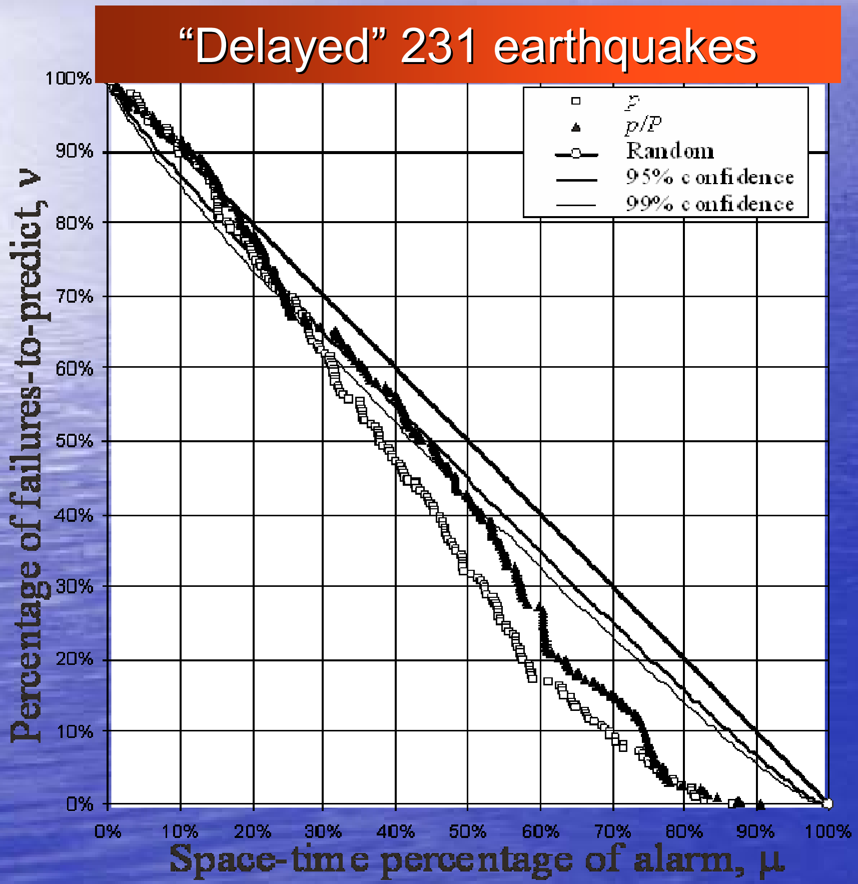


151 main shocks



The outcome of an "absurd" prediction:

The percentage of the failures-to-predict ν versus the percentage of the alerted space-time volume μ: $\{\mu_p(E), \nu_p(E)\}$ and $\{\mu_{p/P}(E), \nu_{p/P}(E)\}$ generated by "prediction" of the 231 earthquakes with magnitude MwHRV ≥ 5.8 and depth ≳ 70 km in April 10, 1992-September 13, 1994 using the *p* and *p/P* maps computed for April 10, 2002-September 13, 2004.



The observed deviation from the diagonal is about the same or better than in the real-time applications. Thus, we cannot reject random coincidence and may conclude that (i) the effectiveness of the Jackson-Kagan "probabilistic" method for predicting large earthquakes is rather doubtful, and that (ii) the applicability of the underlying ETAS model is superstitious.

USGS Web Site May Mislead Californians

On 19 May 2005, the United States Geological Survey began a public web site with forecasts of expected ground shaking for 'tomorrow' and *Nature* published the underlying work by Gerstenberger *et al*[1]. Since that time, two earthquakes of intensity VI in California[2] have occurred in the areas of the web-site's lowest-risk. This should not surprise Californians: Gerstenberger et al[1] overturned the outcome of the primary verification and used a method developed by the Regional Earthquake Likelihood Models (RELM) group[3], which has fundamental flaws that neither define the forecast precision nor allow a means to judge the ultimate success or failure in specific cases.

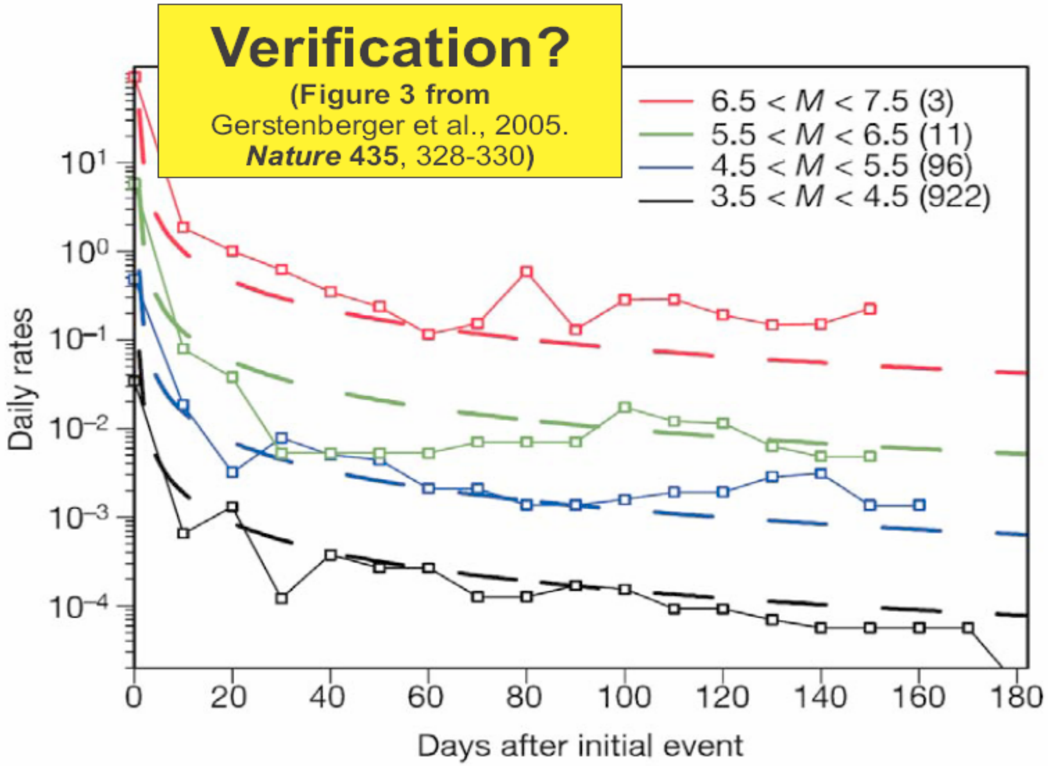
1. Gerstenberger, M. C., Wiemer, S., Jones, L. M. & Reasenberg, P. A. Real-time forecasts of tomorrow's earthquakes in California. *Nature* **435**, 328-331 (19 May 2005)

2. The earthquakes on June 12, 2005 near Anza and June 16, 2005 near Yucaipa produced ground shaking with Modified Mercalli intensity VI (pasadena.wr.usgs.gov/shake/ca).

3. Schorlemmer, D., Gerstenberger, M., Wiemer, S. & Jackson D. Earthquake Likelihood Model Testing (manuscript in preparation, February 7, 2005)

LETTERS NATURE|Vol 435|19 May 2005

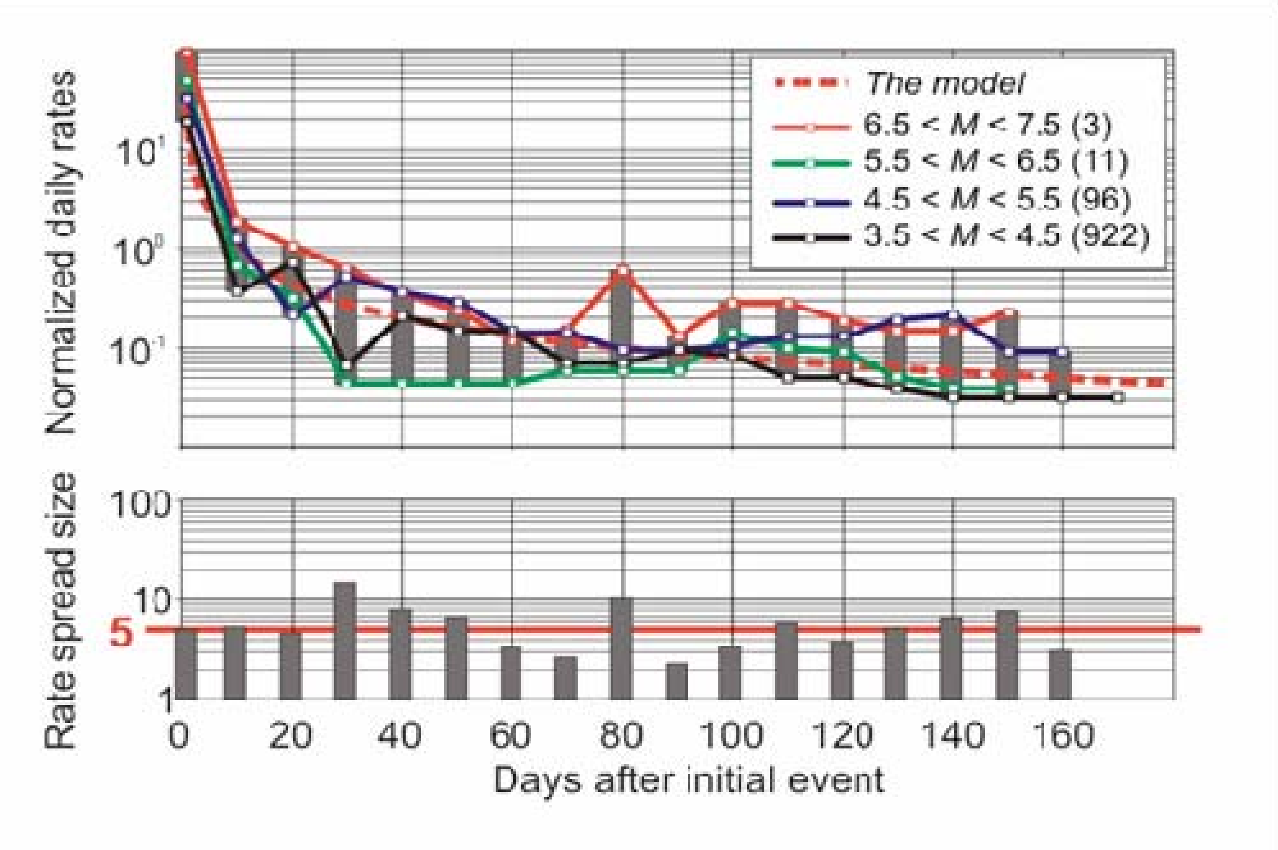
Figure 3 | Calculated and observed rates of events $M \geq 4$ in 24-hour intervals following mainshocks occurring between 1988 and 2002 in southern California. Dashed lines show the rates forecasted by the generic California clustering model (without cascades) for the mainshock magnitude (M) shown. For this test a simple circular aftershock zone implementation (solid lines) gives the observed rates of $M \geq 4.0$ aftershocks following all mainshocks with magnitude within 0.5 units of M. The aftershock zones are defined as the areas within one rupture length of the mainshock epicentre.



Soliciting misuse of statistics?

"As a first test, we verified that the generic clustering model describes the average clustering activity of California reasonably well. Using data from 1988−2002, after the period used to initially develop the model and thus independent data, we compute the average daily rate of events following an earthquake of a given size (Fig. 3)."

Calculated and observed rates of events $M \geq 4$ in 24-hour intervals following mainshocks occurring between 1988 and 2002 in southern California.



Dashed line shows the rate forecasted by the generic California clustering model for the initial mainshock of magnitude $6.5 < M < 7.5$; solid lines display the observed rates of $M \geq 4$ aftershocks following all mainshocks with magnitude within 0.5 units of M, normalized to the rate of the mainshock of magnitude $6.5 < M < 7.5$. Grey bars stretch from the minimal to the maximal value of the observed rates; their size is about a factor of 5.

Analyzing the figure by means of the well-known Kolmogoroff-Smirnoff criterion, an experimentalist would be led to reject the hypothesis that the random variable "Time after initial event" in different magnitude ranges of the initial event has the same statistical distribution.

Proof: Normalised by condition that the total integral of the p.d.f. (probability density function) increments equals 1, each of the four plots provides the minimum of positive p.d.f. increments, which are by definition either 1/N or its integer multiple (e.g., 2/N, 3/N, etc.). These are about 0.0012, 0.0008, 0.0025, and 0.0015, which values imply the sample sizes about 846, 1250, 401, and 665 or integer multiples of these values. The probability of a smaller value of the Kolmogoroff-Smirnoff statistic D than that for the two samples used to plot the daily rates after $5.5 < M < 6.5$ (green plot in Figure 3) event and after $3.5 < M < 4.5$ (black plot) event is larger than 97%. Therefore, the hypothesis that these two samples are drawn from the same distribution can be rejected at significance level of 0.03. ■

(A skilful experimentalist would easily recognize the sample size in the order of a thousand just from the range of the empirical distribution of rates, about three decimal orders, in Figure 3, while a skilful observer would grasp 922 that signifies the number of aftershock about magnitude 4. Moreover, giving a look at Figure 3, he or she, even without any statistical testing, would say that the data does not support the model.)

Aftershocks and their epoch analysis

Aftershock sequences of southern California are extremely different – e.g. the total number of M2.0+ aftershocks in 100 days can be 0 for some main shocks up to magnitude 5.0 (about 10-25% of the total for different magnitudes) and can differ by a factor 10 or more for magnitude 6.0 main shocks (for Whittier Narrows, 1987, M6.2, the number of M2.0+ aftershocks is about one hundred, while for Joshua Tree, 1992, M6.1, it is above 19 hundred). For M7.0+, the recent Landers, 1992, M7.3, has about 8.5 thousand, while Hector Mine, 1999, M7.1, has only 4.6 thousand of M2.0+ aftershocks. Therefore, epoch analysis of the aftershock series, in fact, is analogous to measuring of the average patients' temperature in a clinic, while "an average behavior of the seismicity" in the region is analogous to crossing the pond through the middle of its waters, which is the average of walking around it, by turning to the left or to the right.

Aftershock sequences of the great shocks: Summary from Romashkova and Kossobokov (2001)

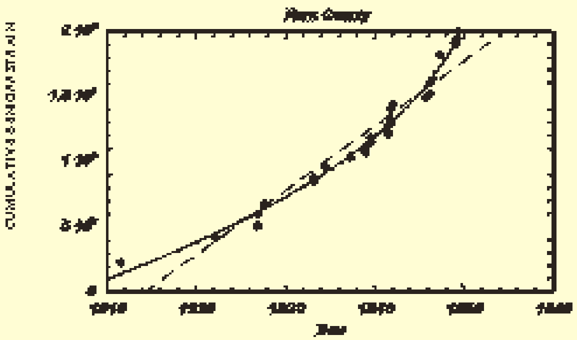
Ромашкова, Л.Л., В.Г. Кособоков. Динамика сейсмической активности до и после сильнейших землетрясений мира, 1985-2000. Проблемы динамик литосферы и сейсмичности. М.: Геос, 2001, 162-189 (Вычислительная сейсмология, Выпуск 32)

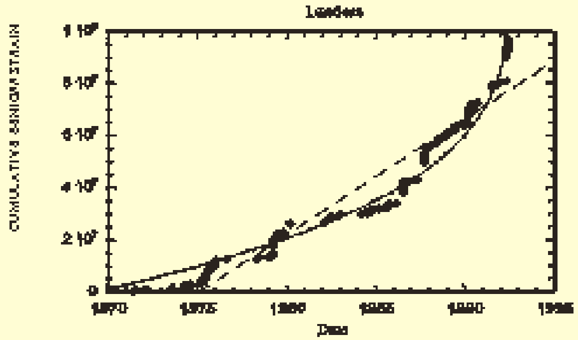
Date	Number 100 days	Number 3 years	Aftershocks decay 100 d	Aftershocks decay 3 y	Relaxation time, years
1985/09/19	29	65	Omori Law	Modified OL 3	284 days
1986/10/20	151	205	Modified OL 3	Modified OL 3	100 days, =1.5
1989/05/23	36	54	Omori Law	Modified OL 2	1.3 years, >3
1993/08/08	121	247	Modified OL 2	Modified OL 3	65 days, >1.5
1994/06/09	5	5	-	-	-
1994/10/04	515	919	Modified OL 2	Modified OL 3	2 years, >2.5
1995/04/07	52	302	Modified OL 2	Modified OL 2	14 days, >2
1995/12/03	311	483	Modified OL 2	Modified OL 3	1 year
1996/02/17	357	427	Modified OL 2	Modified OL 2	2 years, >2.5
1998/03/25	38	47	Omori Law	Modified OL 2	140 days
2000/06/04	278	799	Modified OL 2	Modified OL 2	2 years, >1.7

Thus, the Omori's law for aftershocks is hardly a solidly documented fact. Some earthquakes are not followed by any comparable shocks. Some do cluster in space and time but individual clusters could be of different kinds, among which seismologists distinguish "swarms" and "foreshocks-main shock-aftershocks" series. The absence of commonly accepted definitions and classifications results in controversies.

Other evident cases of misuse of Statistics

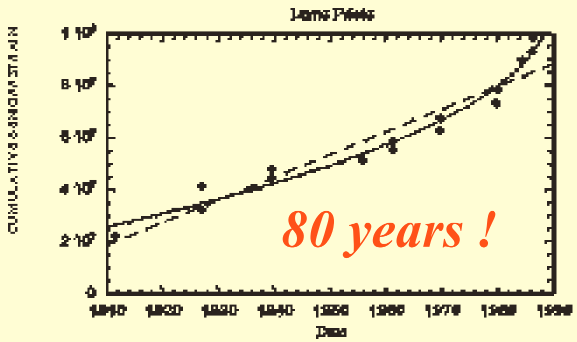
Bowman, Ouillon, Sammis, Sornette, & Sornette, 1998

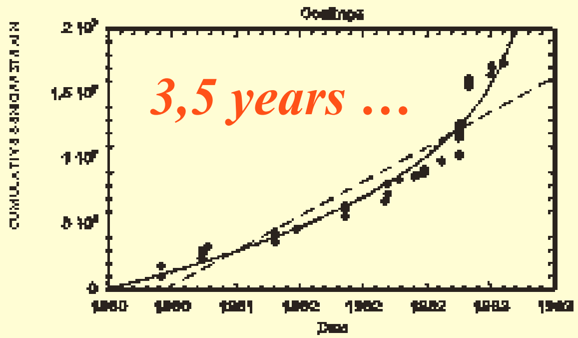


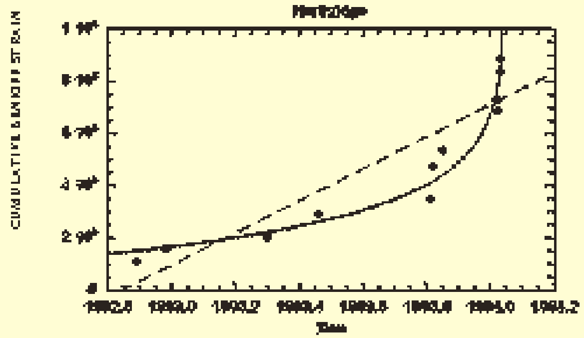


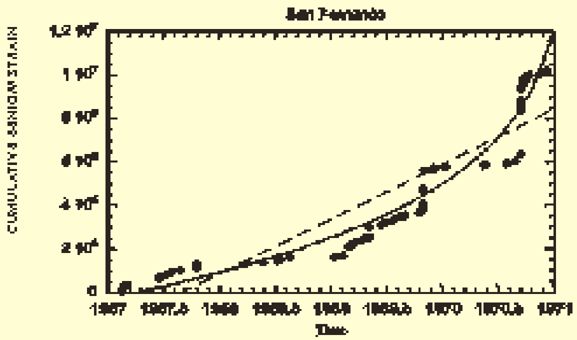
It is still unclear if “the best fit” is random… ?

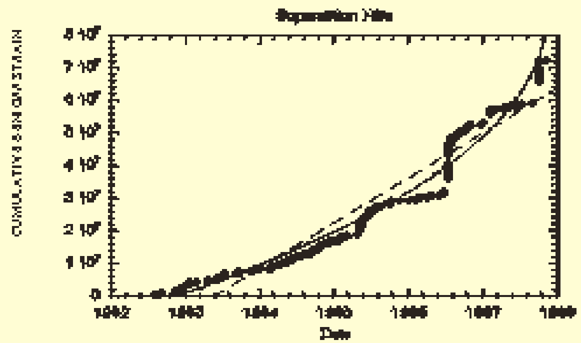
Free parameters: dT, Mc, aftershocks

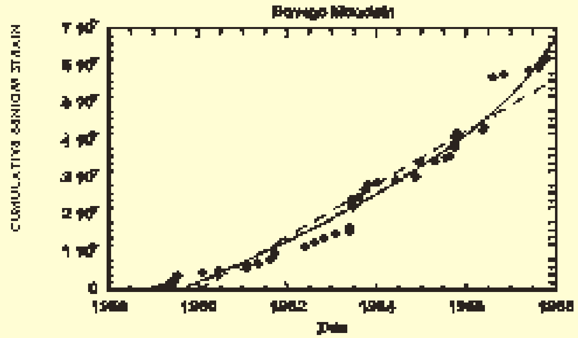


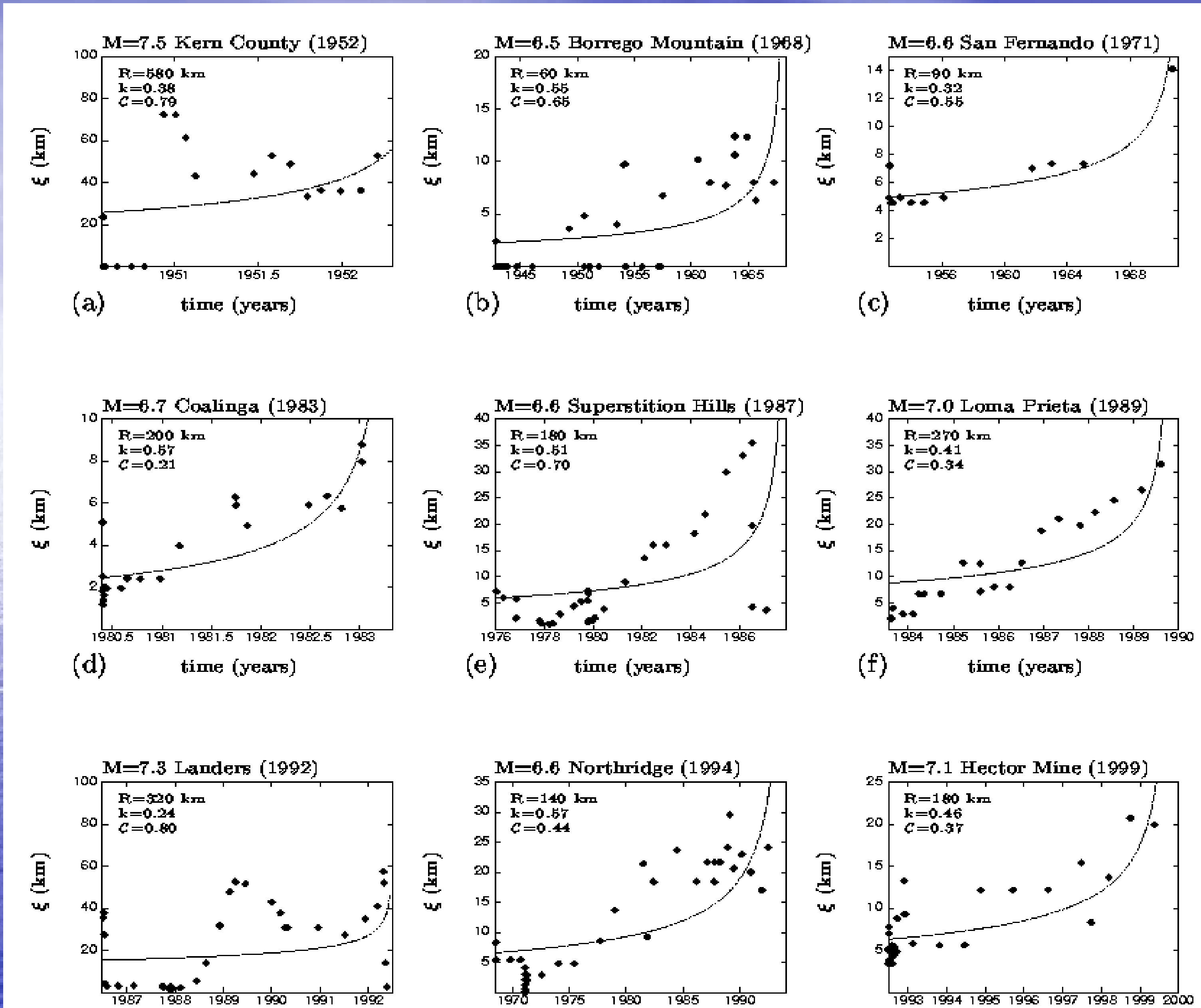












Zoller, Hainzl & Kurths, 2001

Does "the best fit" fit the data at all ?

Verified “Precursors”

- The simple seismicity patterns – Σ and “burst of aftershocks” – were given unambiguous reproducible definitions and their predictive value was validated by the prospective worldwide tests. **However, it** is not clear yet whether some single simple premonitory pattern may compete in performance with prediction algorithms that combine several traits describing the dynamics of seismic region at the approach of a large earthquake.

Worldwide performance of earthquake prediction algorithms M8 and M8-MSc: Magnitude 8.0+.

Test period	Large earthquakes Total	Predicted by M8	Predicted by M8-MSc	Measure of alarms, % M8	Measure of alarms, % M8-MSc	Confidence level, % M8	Confidence level, % M8-MSc
1985-present	11	9	7	33.24	17.14	99.87	99.92
1992-present	9	7	5	28.42	14.37	99.69	99.54

The significance level estimates use the most conservative measure of the alarm volume accounting for empirical distribution of epicenters.

To drive the achieved confidence level below 95%, the Test should encounter four failures-to-predict in a row.

Worldwide performance of earthquake prediction algorithms M8 and M8-MSc: Magnitude 7.5 or more.

Test period	Large earthquakes Total	Predicted by M8	Predicted by M8-MSc	Measure of alarms,% M8	Measure of alarms,% M8-MSc	Confidence level, % M8	Confidence level, % M8-MSc
1985-present	53	30	16	34.35	11.05	99.93	99.98
1992-present	40	19	10	28.77	10.45	99.07	99.31

The significance level estimates use the most conservative measure of the alarm volume accounting for empirical distribution of epicenters.

The prediction for M7.5+ is less effective than for M8.0+. Nevertheless, we continue testing the algorithms for this and smaller magnitude ranges.

Some References

- *Allen, C.R. (Chaiman), W. Edwards, W.J. Hall, L. Knopoff, C.B. Raleigh, C.H. Savit, M.N. Toksoz, and R.H. Turner, 1976. Predicting earthquakes: A scientific and technical evaluation – with implications for society. Panel on Earthquake Prediction of the Committee on Seismology, Assembly of Mathematical and Physical Sciences, National Research Council, U.S. National Academy of Sciences, Washington, D.C.*
- *D. D. Bowman, G. Ouillon, G.G. Sammis, A. Sornette, and D. Sornette, An observational test of the critical earthquake concept, J. Geophys. Res., 103, 24,359-24,372, 1998*
- *C.G. Bufe, and D.J. Varnes. Predictive modeling of the seismic cycle of the greater San Francisco Bay region, J. Geophys. Res., **98**, 9,871-9,883, 1993*
- *Geller, R.J., 1997. Earthquake prediction: A critical review. Geophys. J. Int. **131**: 425–450*
- *R.J. Geller, D. D. Jackson, Y. Y. Kagan, and F. Mulargia. Earthquakes cannot be predicted, Science **275**, 1616-1619, 1997.*
- *Gerstenberger, M. C., Wiemer, S., & Jones, L. M. Real-time Forecasts of Tomorrow's Earthquakes in California: a New Mapping Tool. U. S. Geological Survey OFR-2004-1390 (2004).*
- *D. Jackson, and Y. Kagan. Testable earthquake forecasts for 1999. Seism. Res. Lett., **70**, 393-403, 1999.*
- *Y. Kagan, and D. Jackson. Probabilistic forecasting of earthquake, Geophys. J. Int., **143**, 438-453, 2000.*
- *H. Kanamori, and E.H. Brodsky, 2001. The Physics of earthquakes, Physics Today, June 2001, 34-40*
- *Kossobokov, V.G., V.I. Keilis-Borok, L.L. Romashkova, and J.H. Healy, 1999. Testing earthquake prediction algorithms: Statistically significant real-time prediction of the largest earthquakes in the Circum-Pacific, 1992-1997. Phys. Earth Planet. Inter., **111**, 3-4: 187-196.*
- *Lehmann, E. L. Testing statistical hypotheses. 2nd ed. New York: John Wiley & Sons (1986)*
- *G.M. Molchan, O.E. Dmitrieva, I. M. Rotwain, and J. Dewey. Statistical analysis of the results of earthquake prediction, based on bursts of aftershocks, Phys. Earth Planet. Int. **61**, 128–139, 1990.*
- *G.M. Molchan. Earthquake Prediction as Decision-making Problem. PAGEOPH, vol. 149, 233-247, 1997.*
- *Nonlinear Dynamics of the Lithosphere and Earthquake Prediction. Keilis-Borok & Soloviev (Eds) Springer, Heidelberg, 141-207, 2003.*
- *Omori. On the after-shocks of earthquakes. J. Coll. Sci. Imp. Univ. Tokyo, **7**: 111-200, 1894.*
- *Wyss, M. (editor), 1991. Evaluation of Proposed Earthquake Precursors. AGU, Washington, D.C.*
- *Wyss, M., 1997a. Cannot earthquakes be predicted? Science, **278**: 487–488*
- *Wyss, M., 1997b. Second round of evaluation of proposed earthquake precursors, Pure Appl. Geophys., **149**, 3-16*
- *Zöller, G., Hainzl, S., and Kurths, J. Observation of growing correlation length as an indicator for critical point behavior prior to large earthquakes. J. Geophys. Res., **106**(B2):2167-2176.*