Earthquake Forecast/Prediction:
Verification, Accuracy, Limitations

Vladimir G. Kossobokov

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

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

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

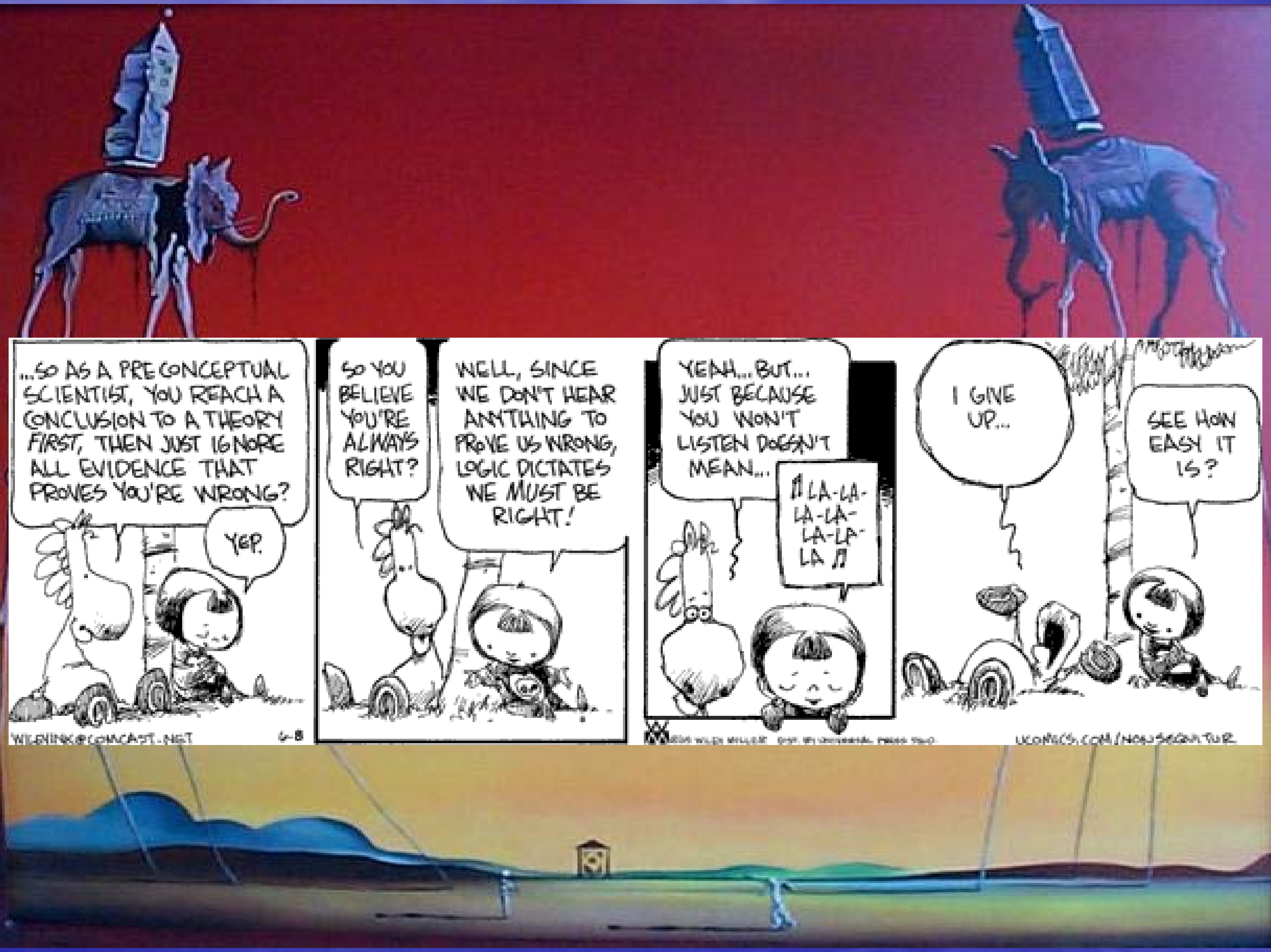
IPGP
INSTITUT DE PHYSIQUE
DU GLOBE DE PARIS

МИТПАН
IIEPT

International Conference on Global Warming

Government of Pakistan
National Centre for Physics
Quaid-i-Azam University Campus,Islamabad

13-14 November
2006


...SO AS A PRECONCEPTUAL SCIENTIST, YOU REACH A CONCLUSION TO A THEORY FIRST, THEN JUST IGNORE ALL EVIDENCE THAT PROVES YOU'RE WRONG?
YEP.
SO YOU BELIEVE YOU'RE ALWAYS RIGHT?
WELL, SINCE WE DON'T HEAR ANYTHING TO PROVE US WRONG, LOGIC DICTATES WE MUST BE RIGHT!
YEAH... BUT... JUST BECAUSE YOU WON'T LISTEN DOESN'T MEAN...
♫ LA-LA-LA-LA-LA-LA-LA ♫
I GIVE UP...
SEE HOW EASY IT IS?
WILDINK@COMCAST.NET
6-8
UCOMICS.COM/NONSEQUITUR

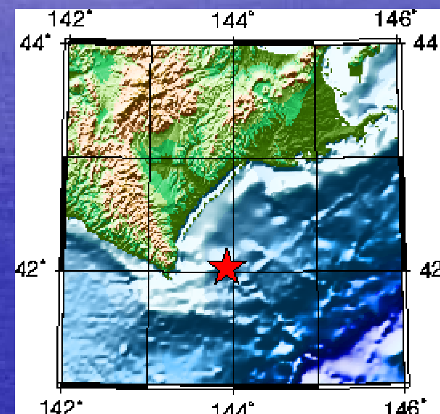
Seismology is juvenile and its appropriate statistical tools to-date may have a "medieval flavor" for those who hurry up to apply a fuzzy language of a highly developed probability theory. To become "quantitatively probabilistic" earthquake forecasts/predictions must be defined with a scientific accuracy. Following the most popular objectivists' viewpoint on probability, we cannot claim "probabilities" adequate without a long series of "yes/no" forecast/prediction outcomes. Without "antiquated binary language" of "yes/no" certainty we cannot judge an outcome ("success/failure"), and, therefore, quantify objectively a forecast/prediction method performance.

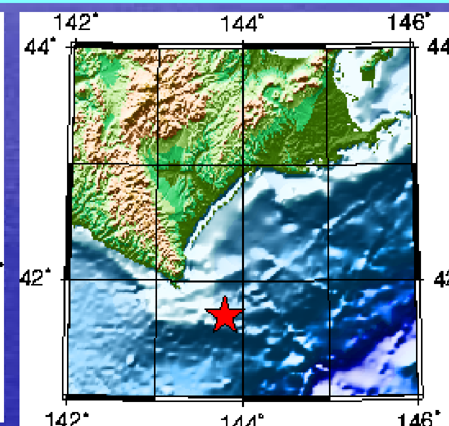
PLANETS ALIGN:
On Wednesday morning, September 24th, 2003 a lovely trio appeared in the eastern sky: **Jupiter, the crescent moon and Mercury…**

Is it a coincidence or a law?

Two days later …

Hi-net

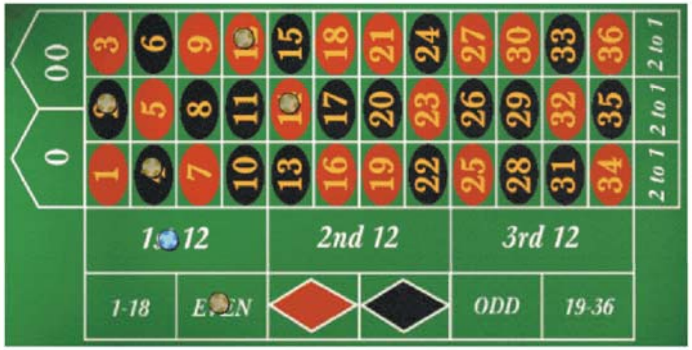


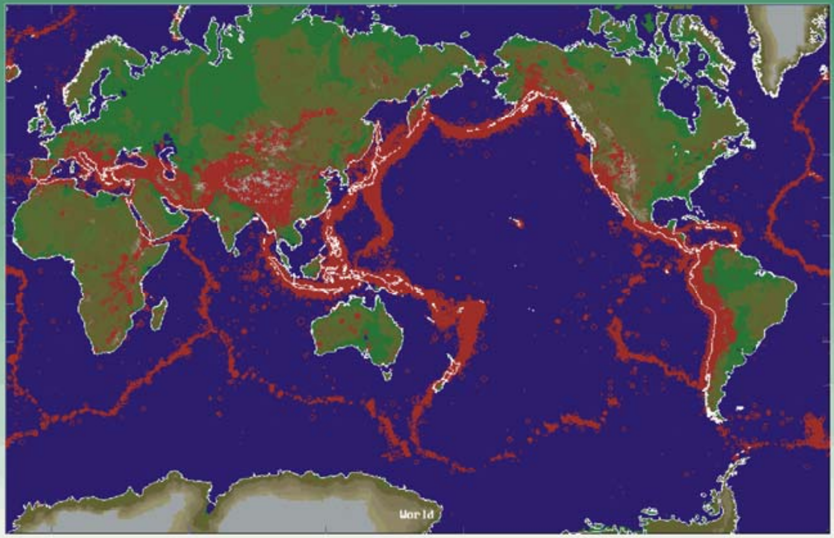


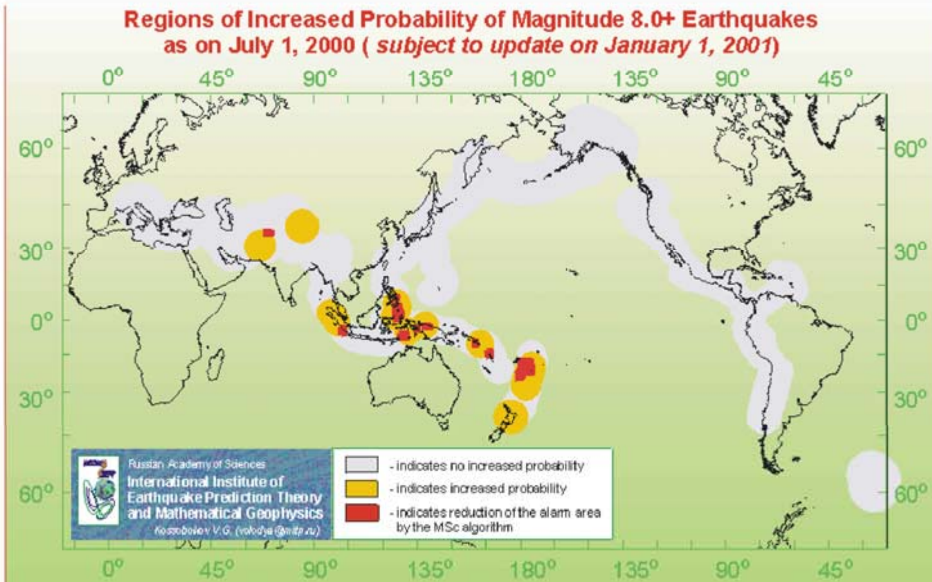
2003□9□26□04□50□11□
□□42.0□
□□143.9□
□□25km
M7.7

2003□9□26□06□08□03□
□□41.8□
□□143.9□
□□35km
M7.4

Seismic Roulette







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…

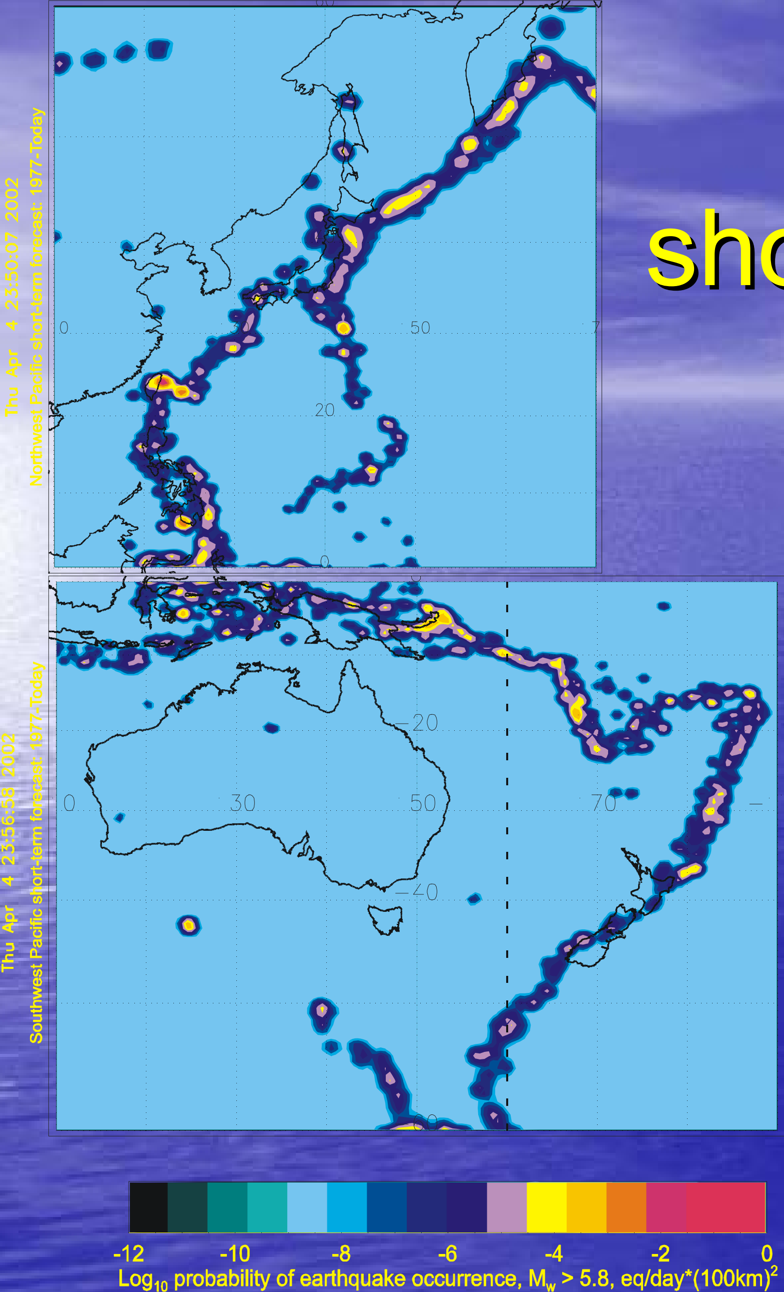
then ***systematically*** you can win! ☺

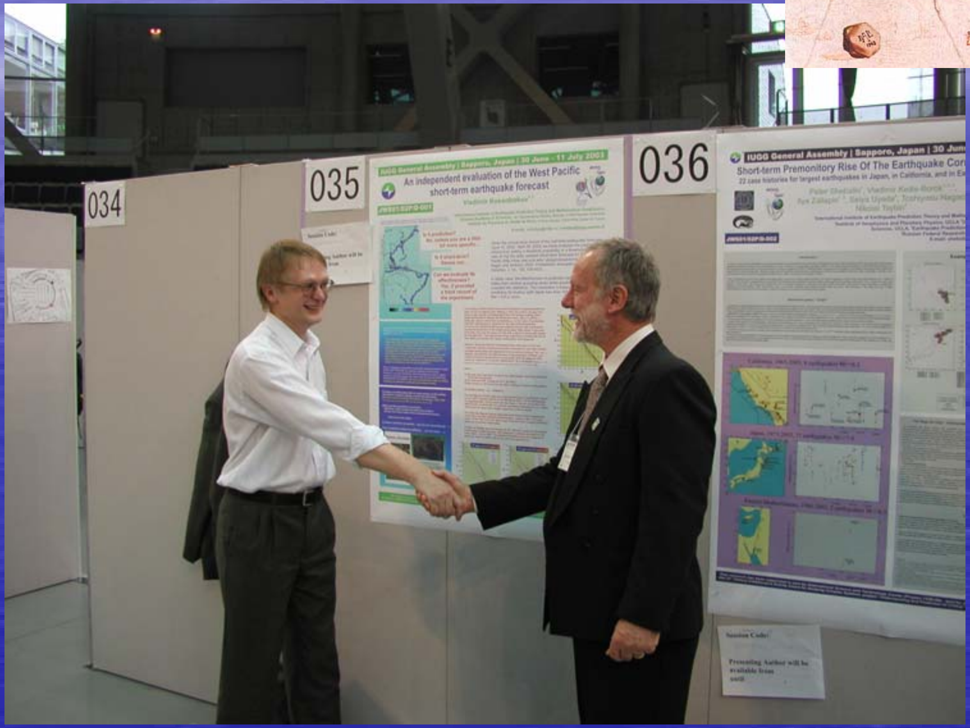
and lose … ☹

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

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

West Pacific short-term forecast





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;

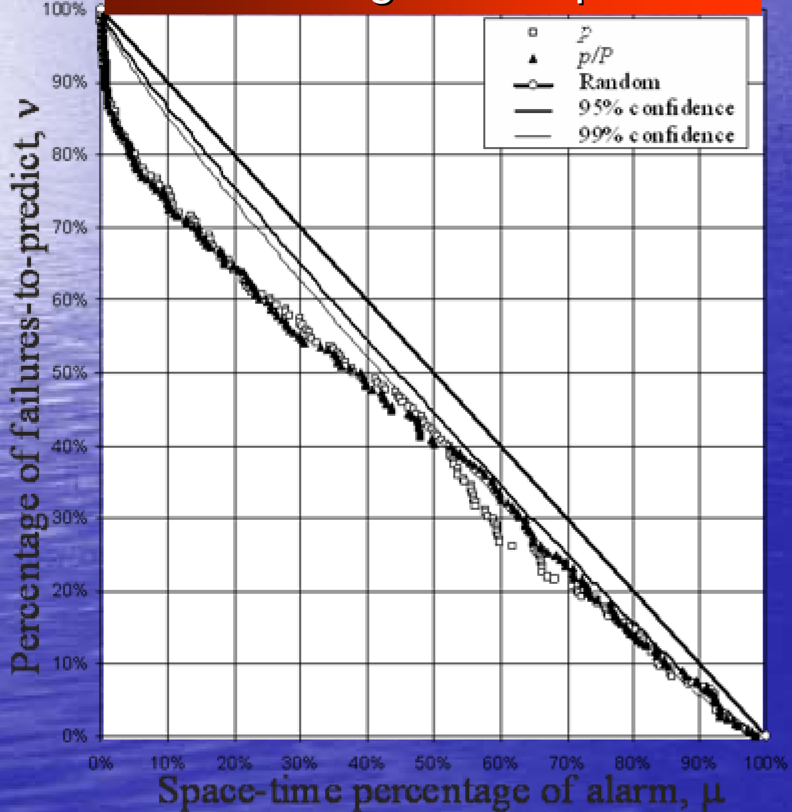
(2) the Harvard CMT catalog for the same period of time;

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

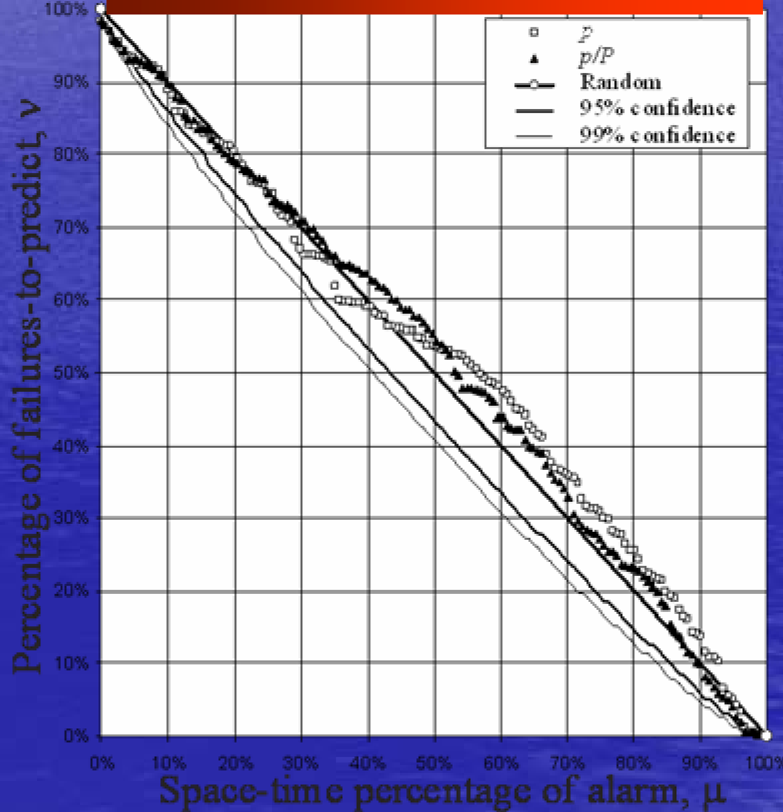
we plotted several ν vs. μ diagrams.

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



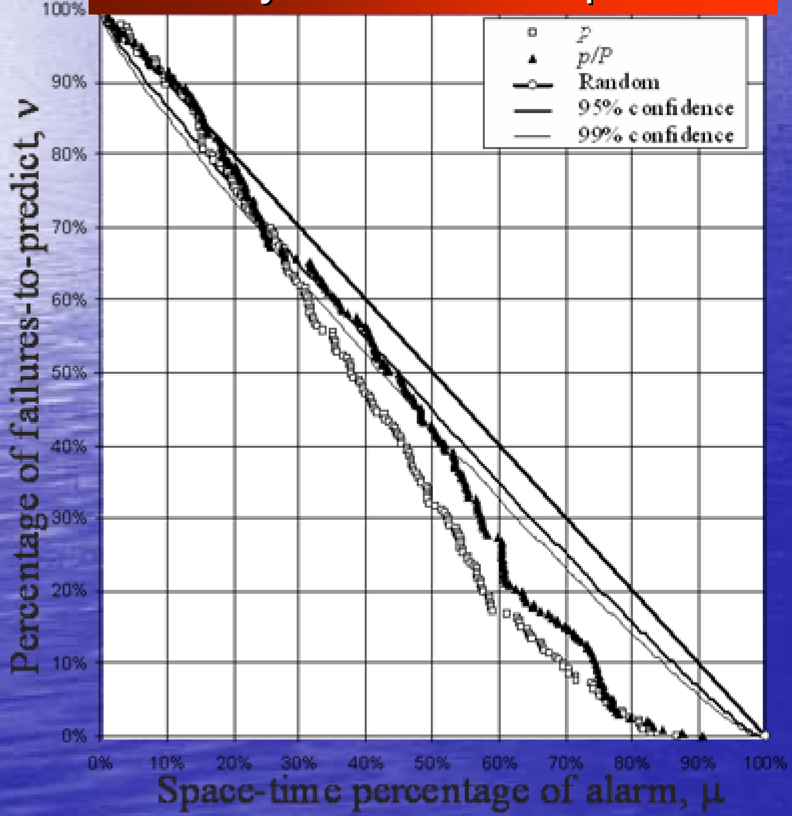
151 main shocks



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.

“Delayed” 231 earthquakes



The observed deviation from the diagonal is about the same or larger than in the real-time applications.

Thus, we cannot reject random nature of the Jackson-Kagan “probabilistic” method and may conclude that

(i) its effectiveness for predicting large earthquakes is doubtful, and

(ii) the applicability of the underlying ETAS model is an ingrained bigotry.

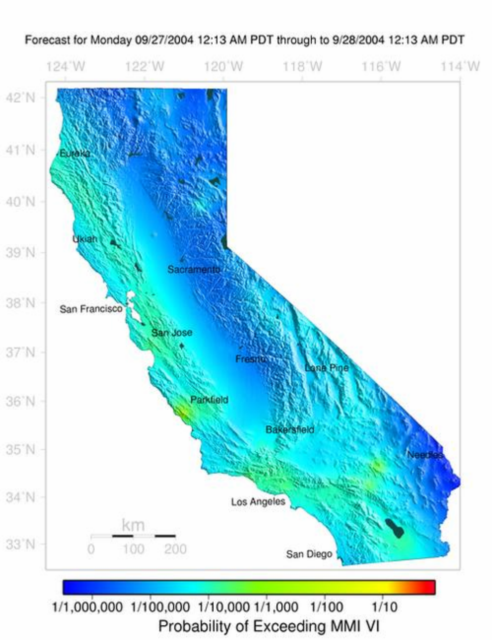
“Hierarchical evidence is a house of cards.

Pull out your primary assumption, and everything gets shaky.”

Regional Earthquake Likelihood Models:

A realm on shaky grounds?

Likelihood scoring is one of the delicate tools of Statistics, which could be worthless or even misleading when inappropriate probability models are used. This is a basic loophole for a misuse of likelihood as well as other statistical methods on practice. The flaw could be avoided by an accurate verification of generic probability models on the empirical data. It is not an easy task in the frames of the Regional Earthquake Likelihood Models (RELM) methodology, which neither defines the forecast precision nor allows a means to judge the ultimate success or failure in specific cases. Hopefully, the RELM group realizes the problem and its members do their best to close the hole with an adequate, data supported choice.



Regretfully, this is not the case with the erroneous choice of Gerstenberger et al., who started the public web site with forecasts of expected ground shaking for `tomorrow' (*Nature* 435, 19 May 2005).

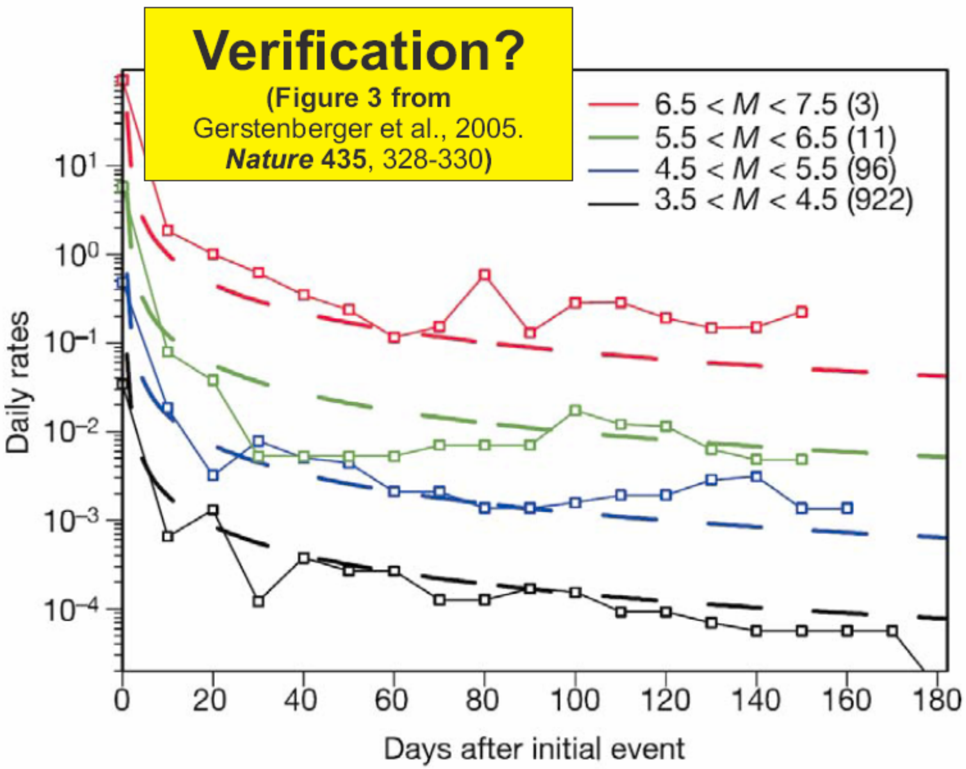
Gerstenberger et al. **HAVE INVERTED THE CRITICAL EVIDENCE OF THEIR STUDY,** i.e., the 15 years of recent seismic record accumulated just in one key figure, which suggests rejecting with confidence above 97% "the generic California clustering model" used in automatic calculations.

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)

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

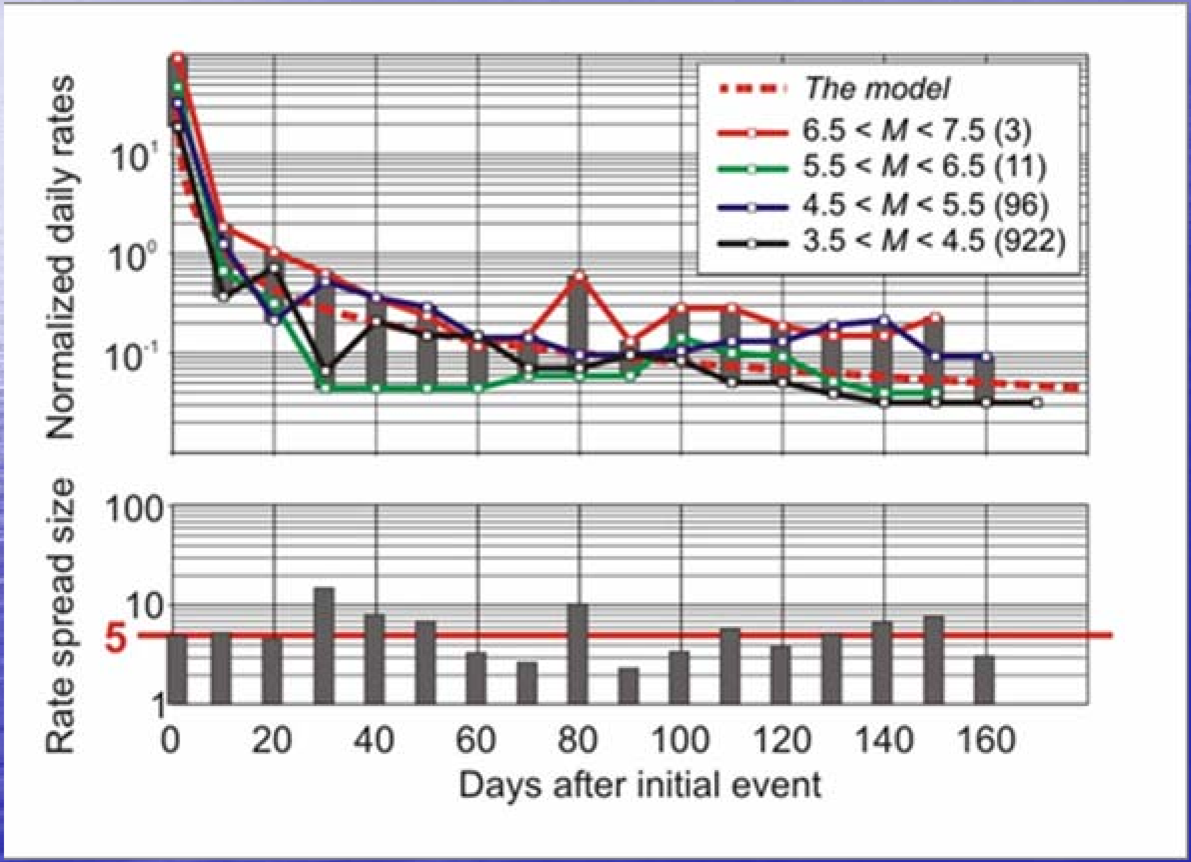
Figure 3 | Calculated and observed rates of events $M \geq 4$ in 24-hour intervals following mainshocks occurring between 1988 and 2002 in southern California. Dashed lines show the rates forecasted by the generic California clustering model (without cascades) for the mainshock magnitude (M) shown. For 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.

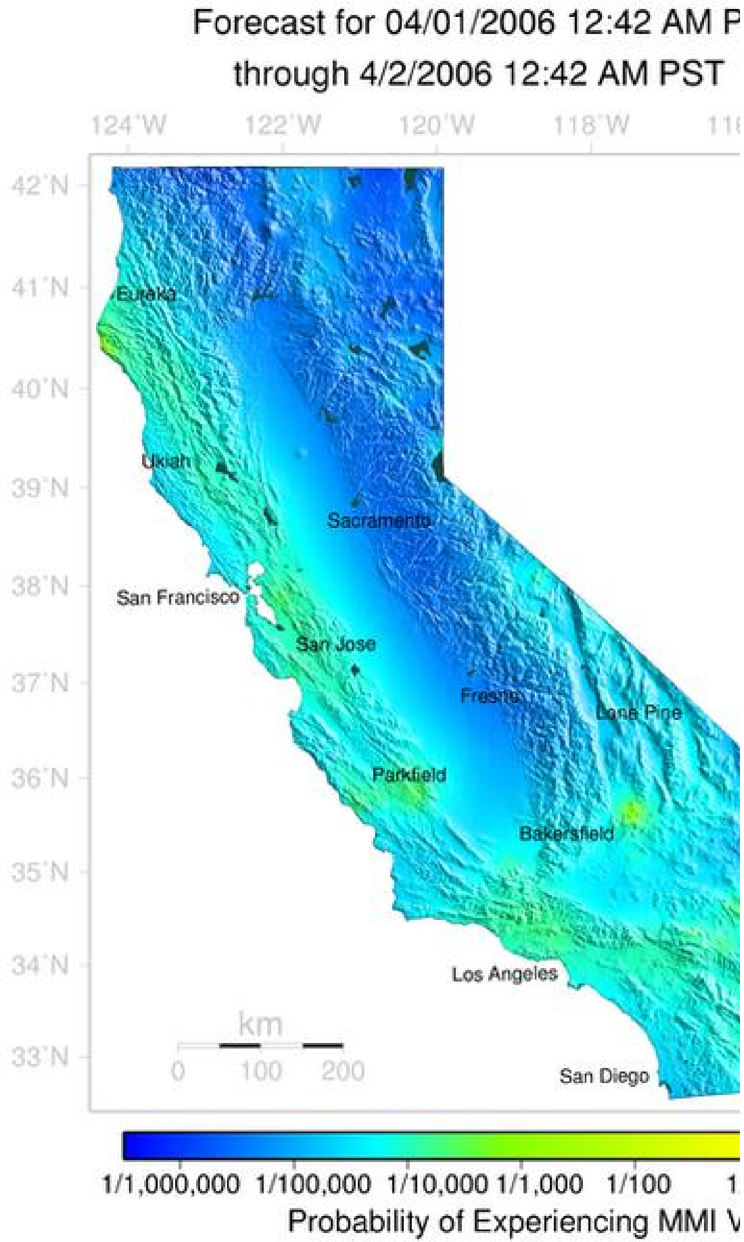


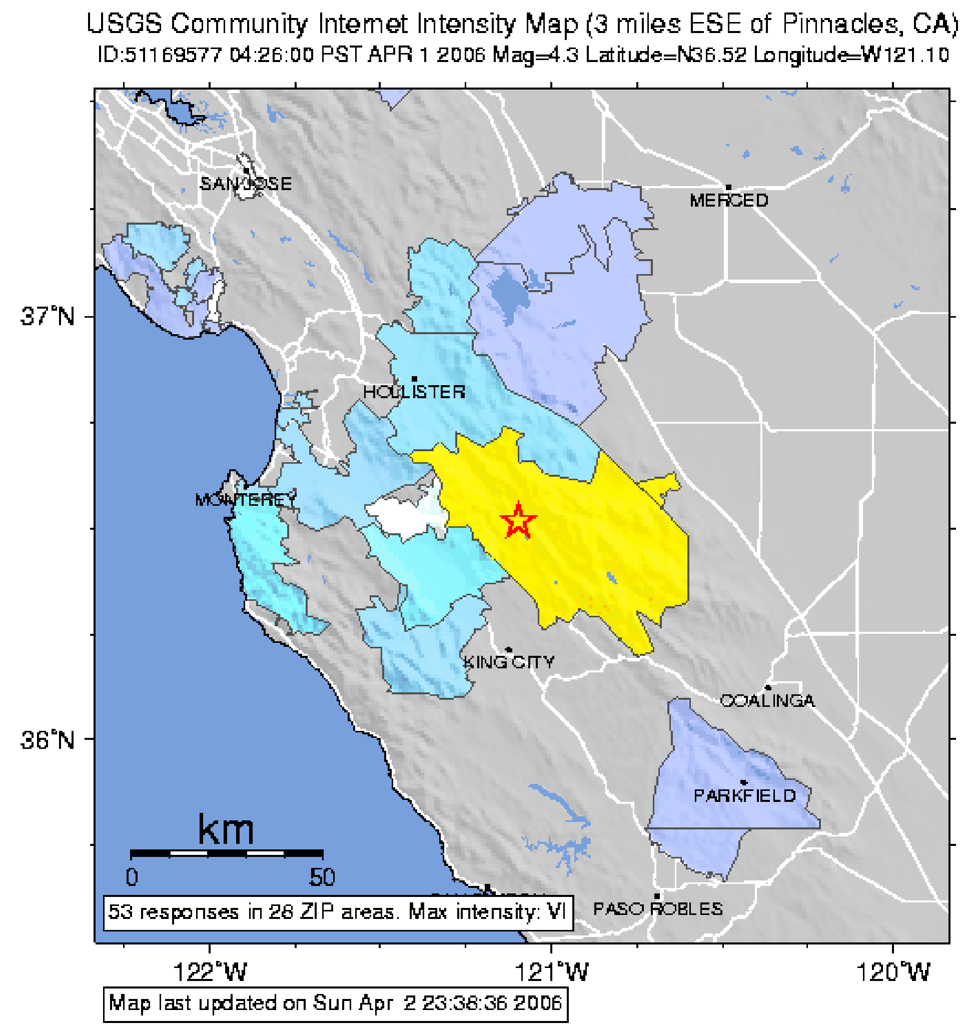
Dashed line shows the rate forecasted by the generic California clustering model for the initial mainshock of magnitude $6.5 < M < 7.5$; solid lines display the observed rates of $M \geq 4$ aftershocks following all mainshocks with magnitude within 0.5 units of M, normalized to the rate of the mainshock 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 (i.e., $D = 0.07 \cdot (N_1 N_2/(N_1+N_2))^{1/2} \geq 2.12$) 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. ■

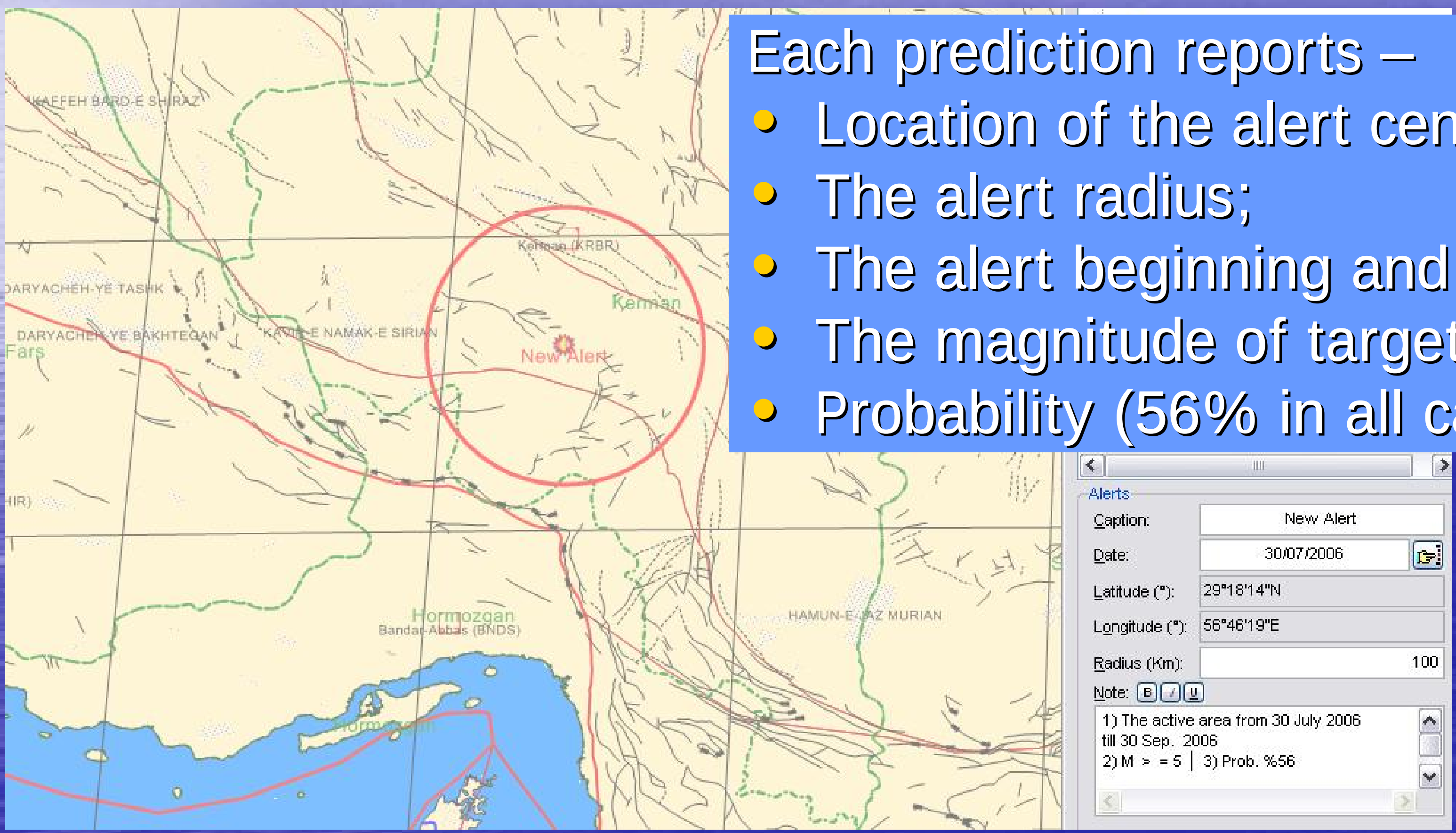
(A skilful experimentalist would easily recognize the sample size in the order of a thousand just from the range of the empirical distribution of rates, about three decimal orders, in Figure 3, while a skilful observer would grasp 922 that signifies the number of events about magnitude 4. Moreover, giving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





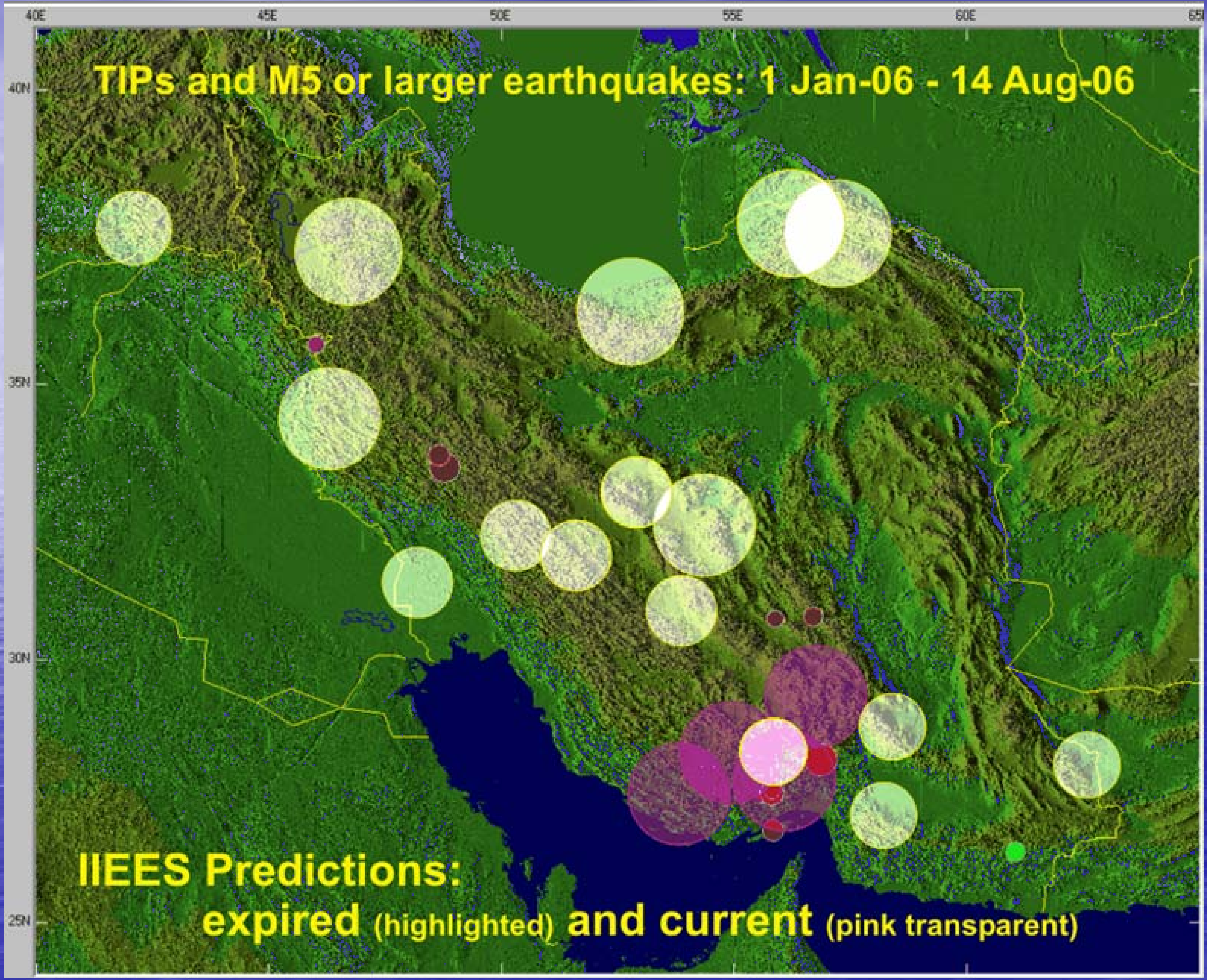
INTENSITY	I	II-III	IV	V	VI	VII	VIII	IX	X+
SHAKING	Not felt	Weak	Light	Moderate	Strong	Very strong	Severe	Violent	Extreme
DAMAGE	none	none	none	Very light	Light	Moderate	Moderate/Heavy	Heavy	Very Heavy

Recently, IIEES did set up a website of restricted access (ftp://www.iiees.ac.ir/eqprediction), which we have a chance to visit systematically since March 8, 2006.

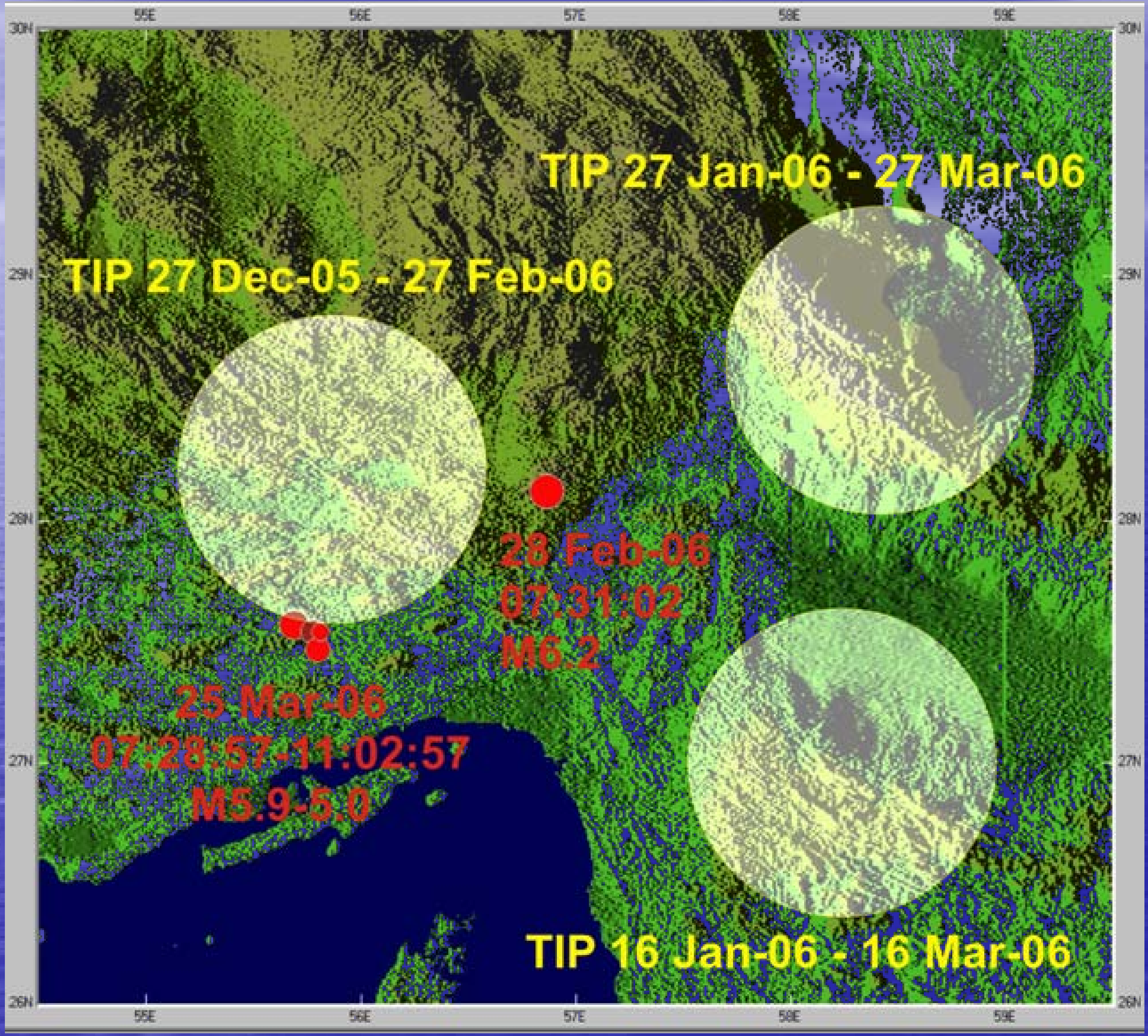


Each prediction reports –

- Location of the alert center;
- The alert radius;
- The alert beginning and end;
- The magnitude of target event;
- Probability (56% in all cases)


40E
45E
50E
55E
60E
65E
40N
35N
30N
25N
TIPs and M5 or larger earthquakes: 1 Jan-06 - 14 Aug-06
IIEES Predictions:
expired (highlighted) and current (pink transparent)

Zoom to alerts and target earthquakes in the South

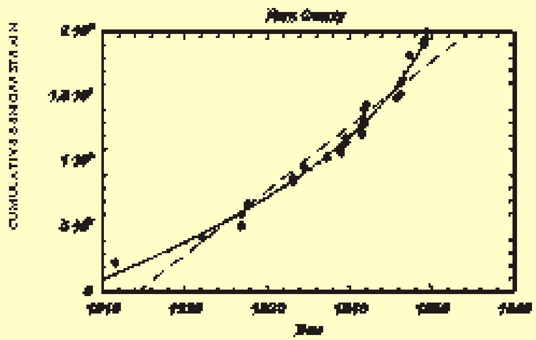


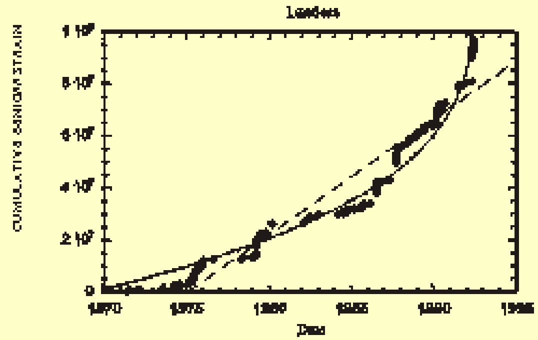
IIEES predictions:

- We continuously observe no success;
- Evidently, this highly contradicts the expected number $P \cdot N = 56\% \cdot 21 = 11.76$ (presumably, P is an estimate of probability of success);
- The IIEES predictions are misleading and their dissemination to the public, emergency planners and the media should not be done;
- The underlying theory is either erroneous or applied in a wrong way.

Other evident cases of misuse of Statistics

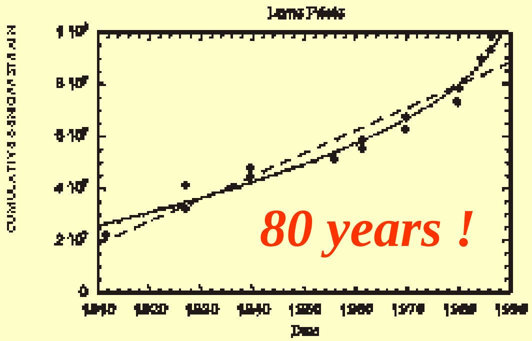
Bowman, Ouillon, Sammis, Sornette, & Sornette, 1998

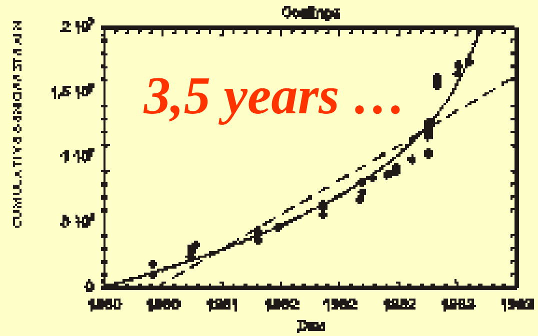


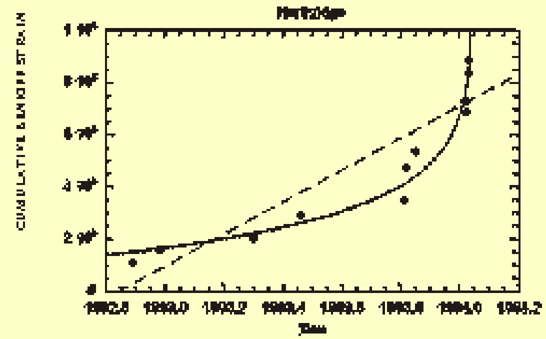


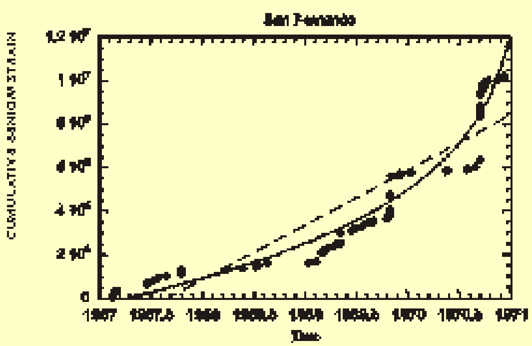
It is still unclear if "the best fit" is random… ?

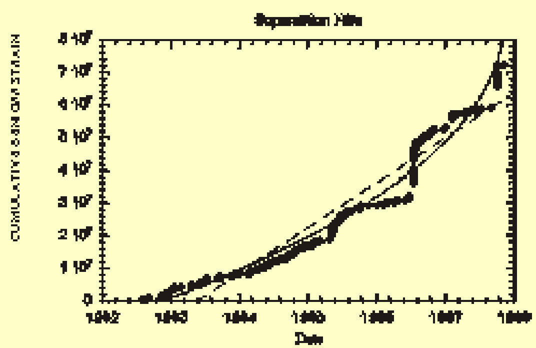
<u>**Free parameters**</u>:
dT, Mc, aftershocks

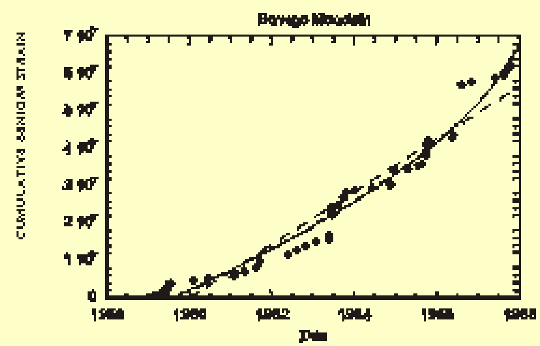


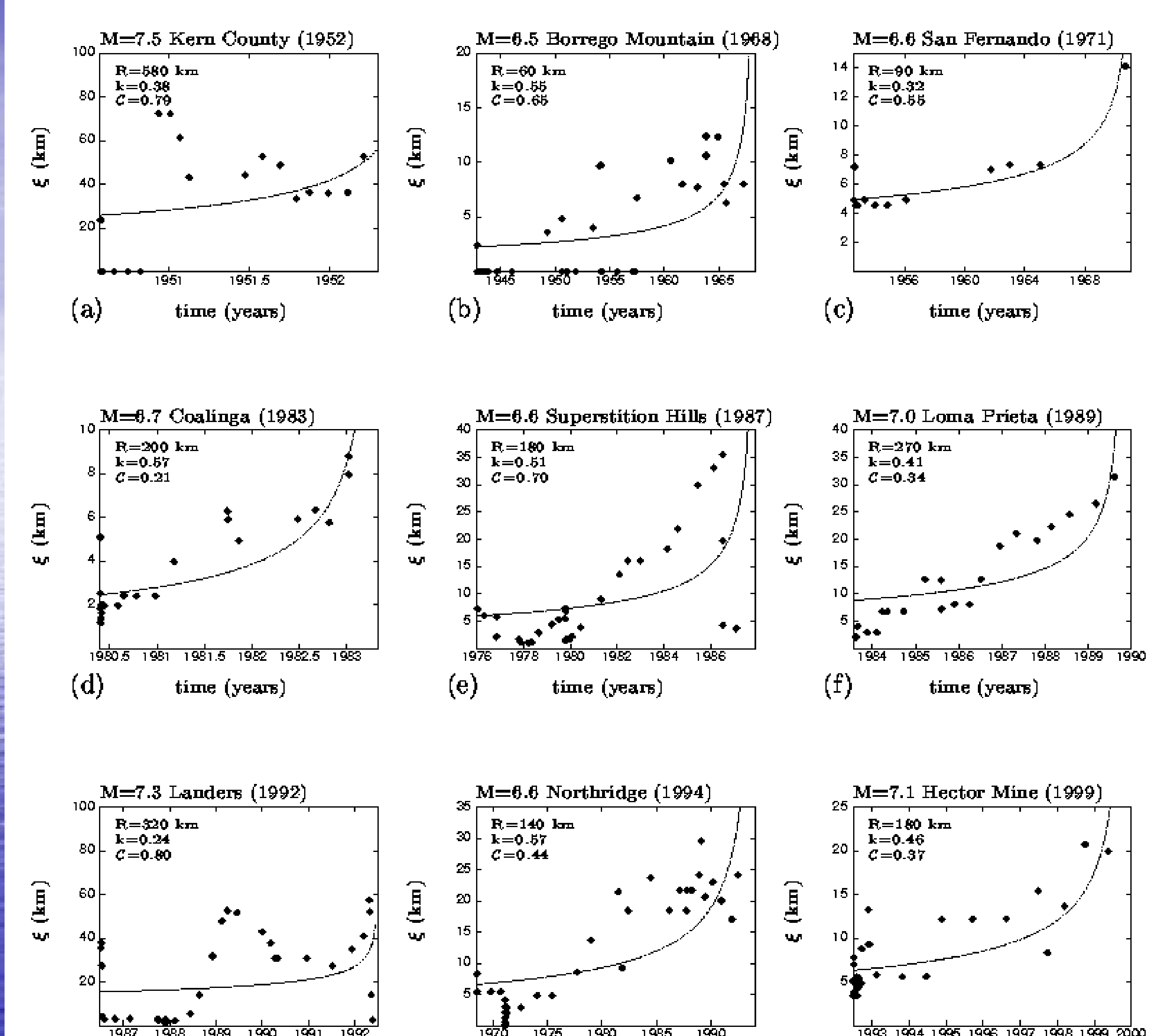












Zoller, Hainzl & Kurths, 2001

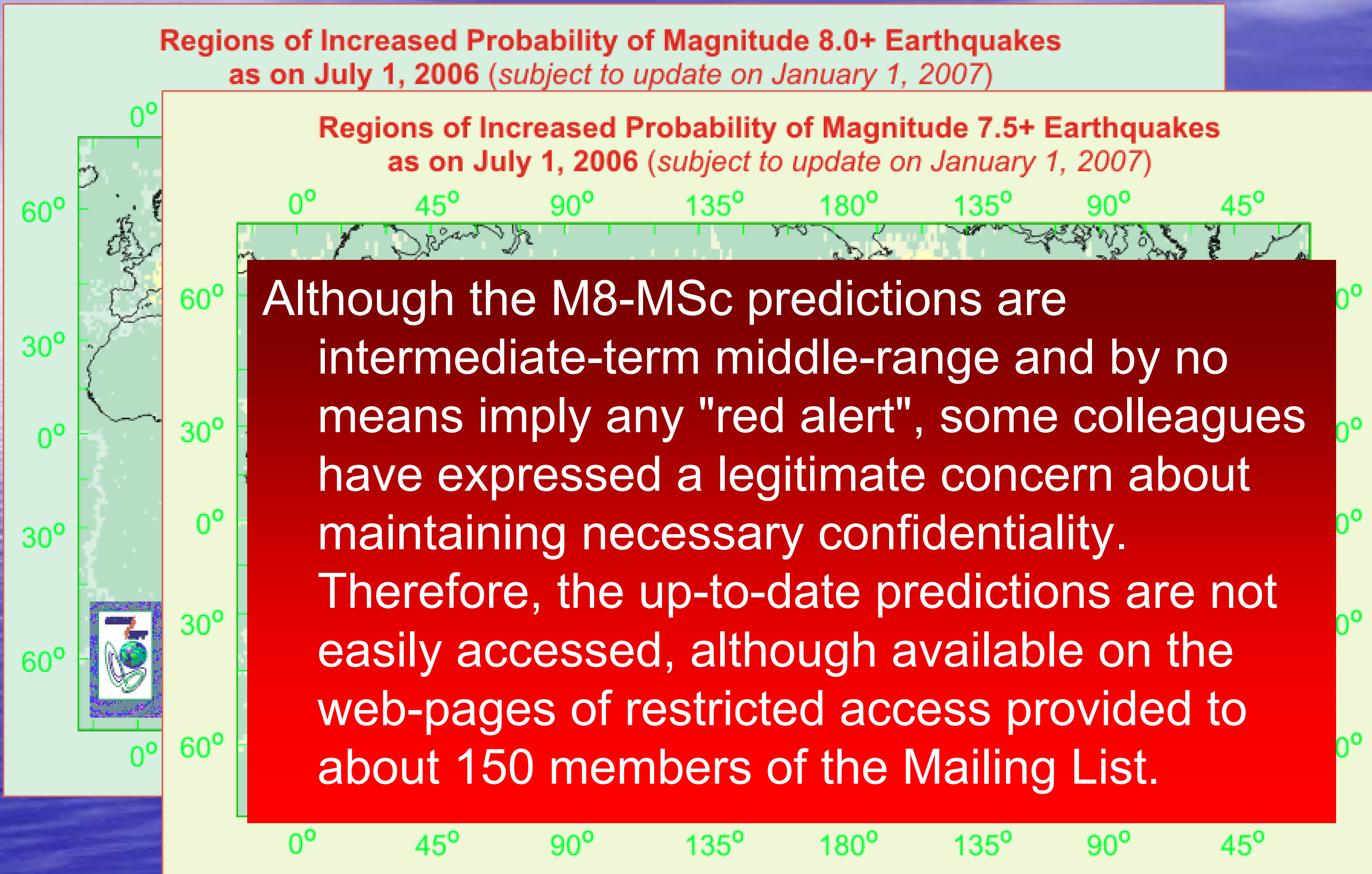
Does "the best fit" fit the data at all ?

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, i**t 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.

Real-time prediction of the world largest earthquakes

(http://www.mitp.ru or http://www.phys.ualberta.ca/mirrors/mitp)



Although the M8-MSc predictions are intermediate-term middle-range and by no means imply any "red alert", some colleagues have expressed a legitimate concern about maintaining necessary confidentiality. Therefore, the up-to-date predictions are not easily accessed, although available on the web-pages of restricted access provided to about 150 members of the Mailing List.

"Undue precision of computations is the first symptom of mathematical illiteracy"

N.Krylov, famous Russian mathematician

The accuracy of an earthquake prediction method is essentially predefined by the accuracy of the data available, which is far from ideal. The unavoidable natural difficulties in observing seismic events as well as in correlating them with other geophysical phenomena and fields complicates the design and testing of a new generation of earthquake prediction technique.

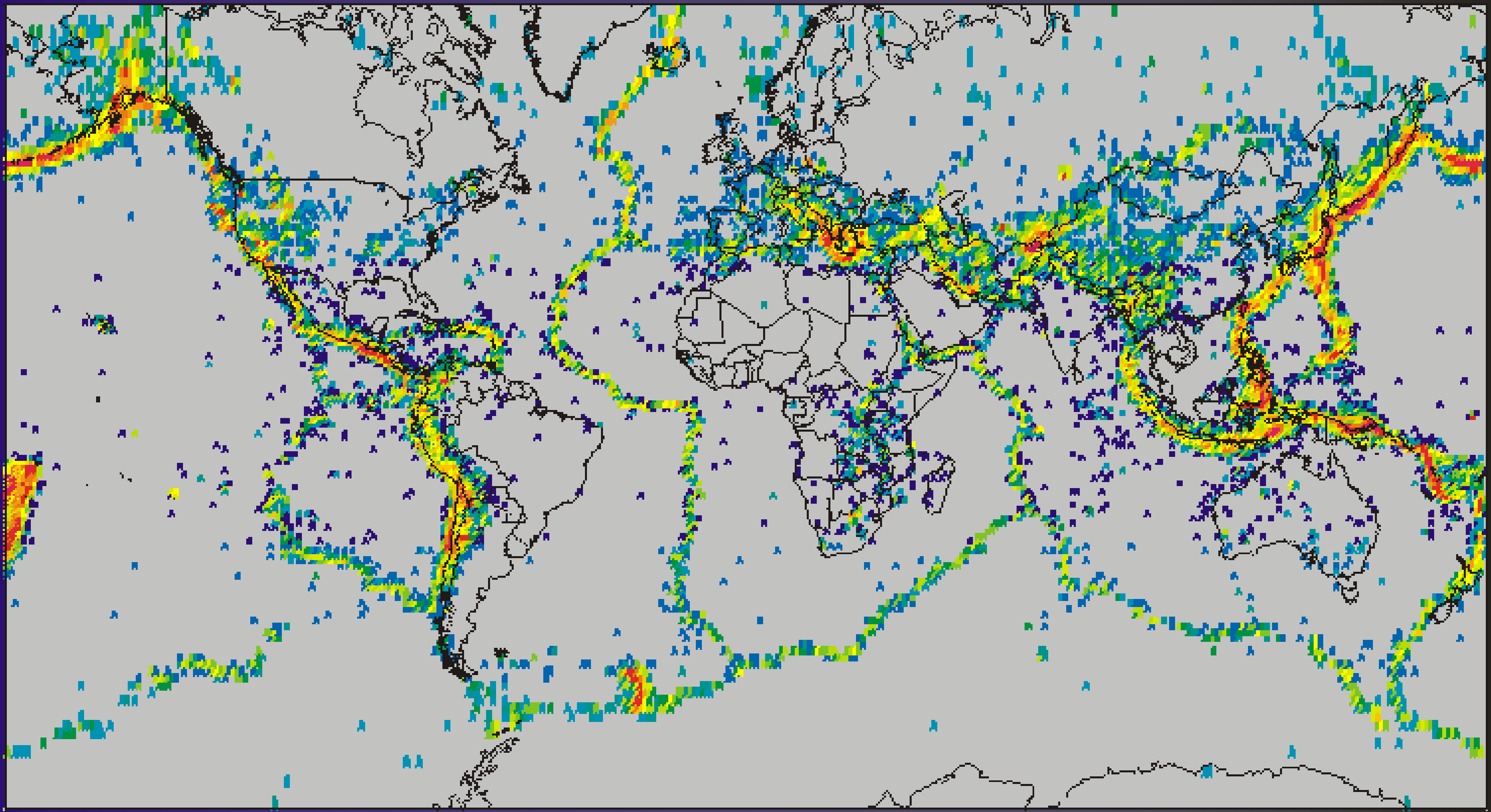
The accumulated case-histories of predicted and not predicted earthquakes provide us unique and so far very limited information that may help understanding the ultimate limits of seismic predictability.

Stages of earthquake prediction

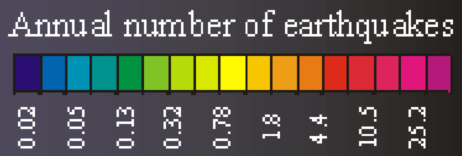
- Term-less prediction of earthquake-prone areas
- Prediction of time and location of an earthquake of certain magnitude

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.

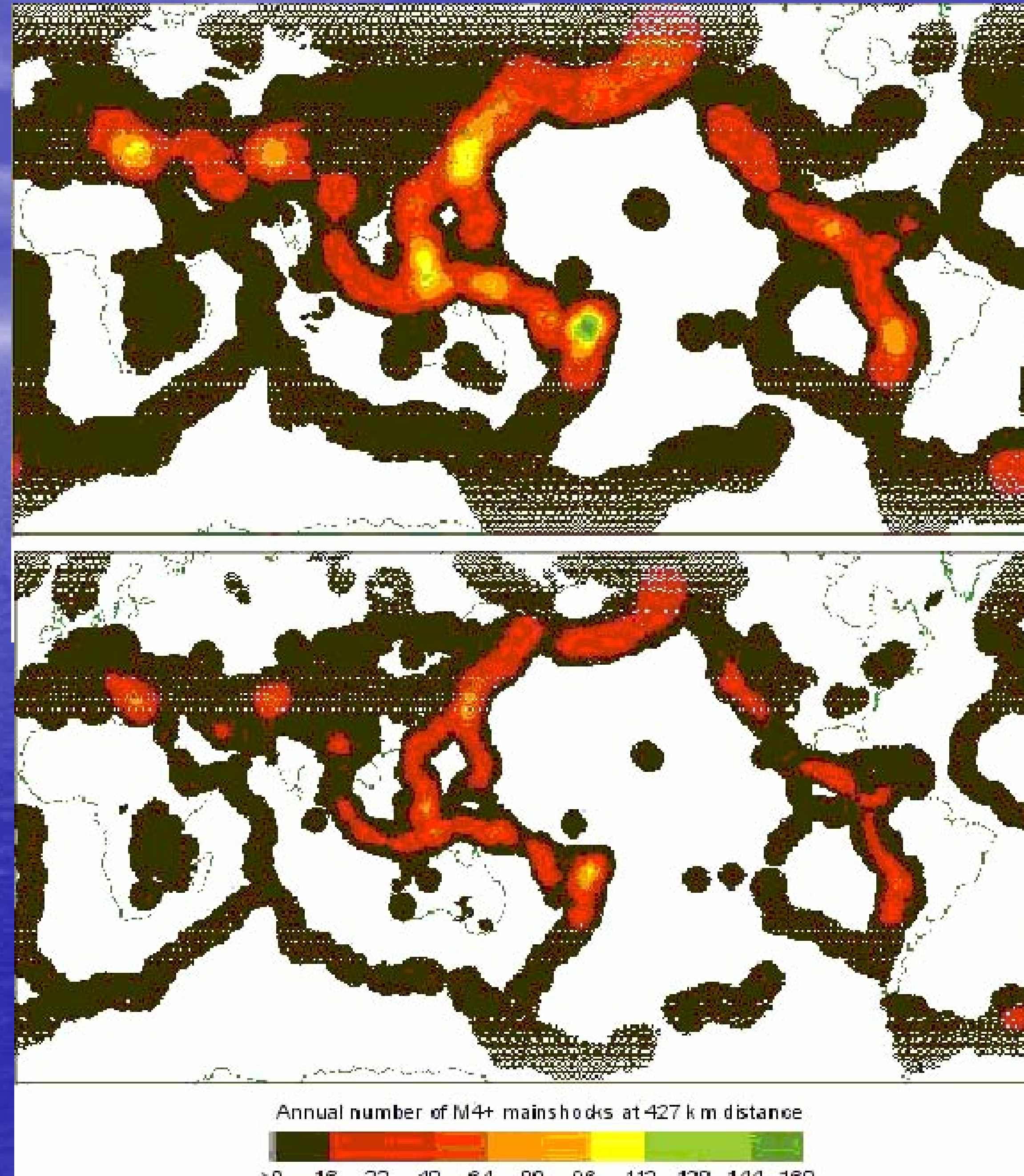


Average annual number of magnitude 4.0 or greater earthquakes at a 1°×1° cell (*normalized to its area on equator*)



Earthquakes are rare events. Therefore, the application of the M8 algorithm is limited to the areas where reported earthquakes are large enough in number.

The color on the maps signifies the annual average number of earthquakes with magnitude 4 or larger in the 667-km (above) and 427-km (below) circles centered at the point.



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

Test period	Large earthquakes: Total	Large earthquakes: Predicted by M8	Large earthquakes: 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	52	30	16	34.35	11.05	99.95	99.99
1992-present	40	20	10	28.77	10.45	99.34	99.43

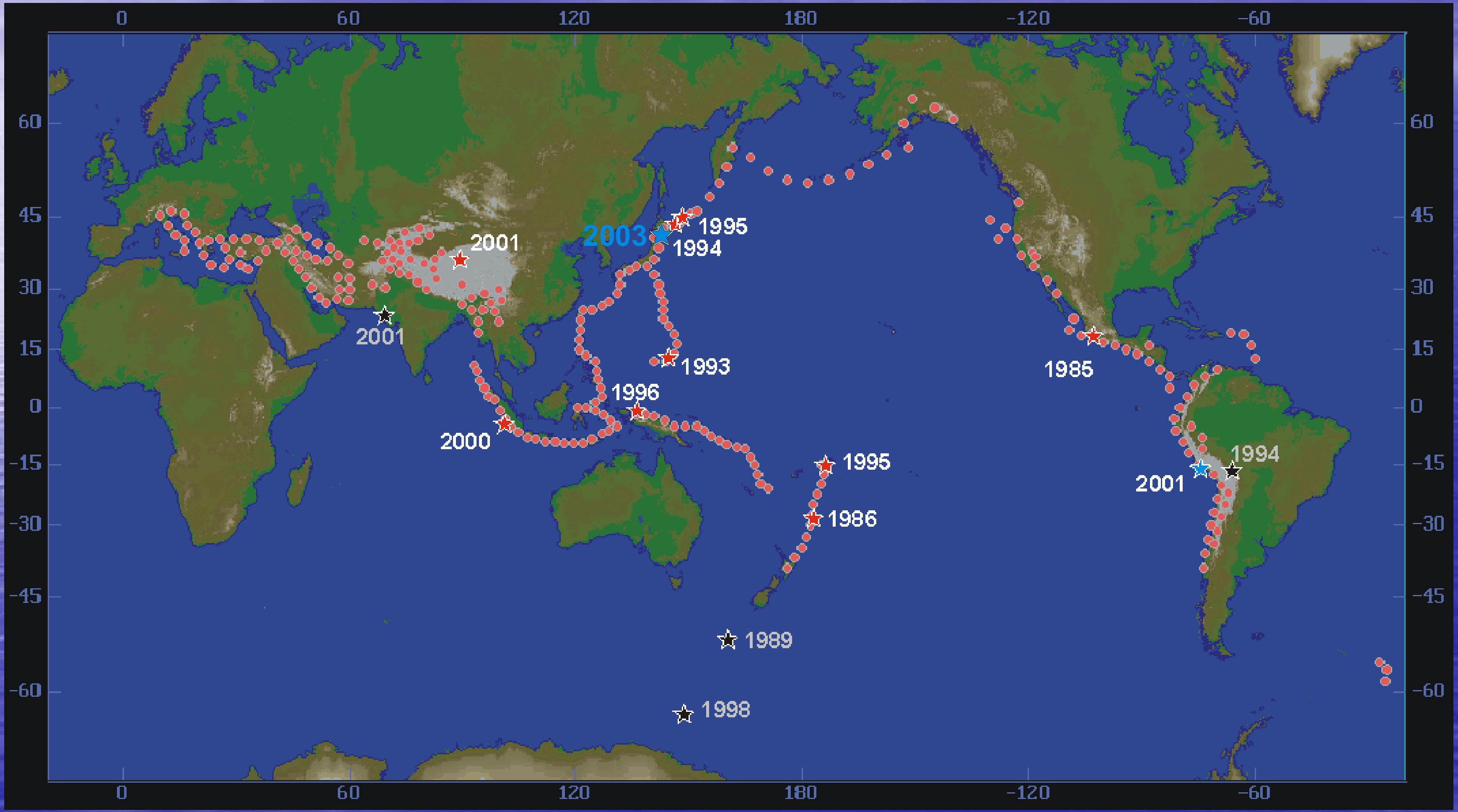
The significance level estimates use the most conservative measure of the alarm volume accounting for empirical distribution of epicenters.

The prediction for M7.5+ is less effective than for M8.0+.

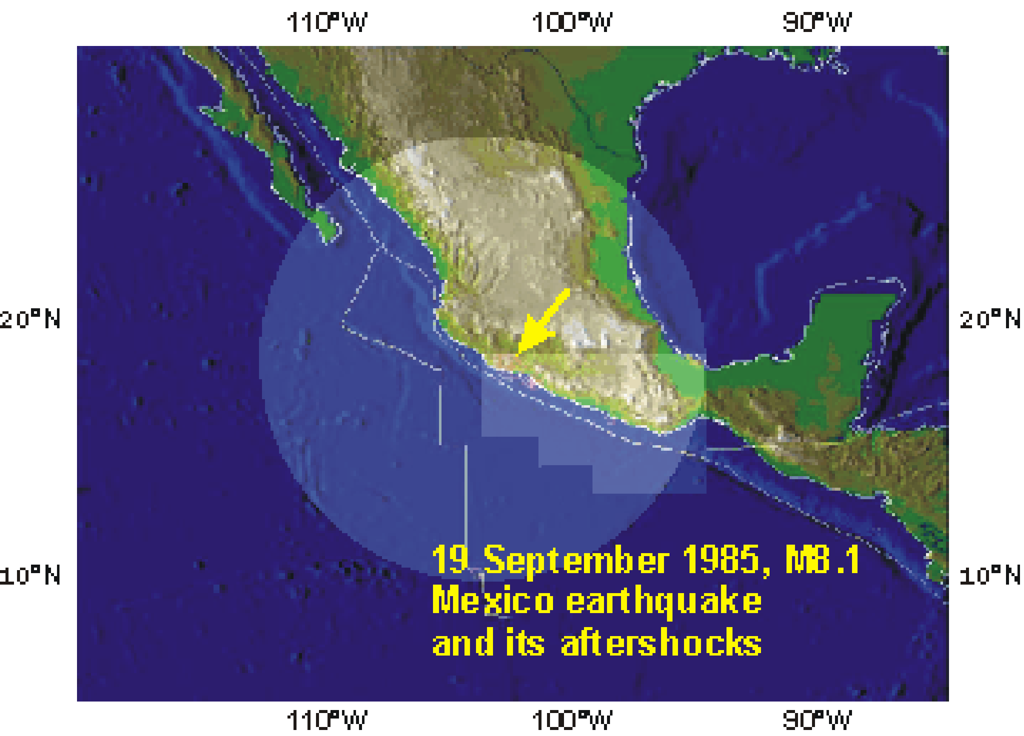
Nevertheless, we continue testing the algorithms for this and smaller magnitude ranges.

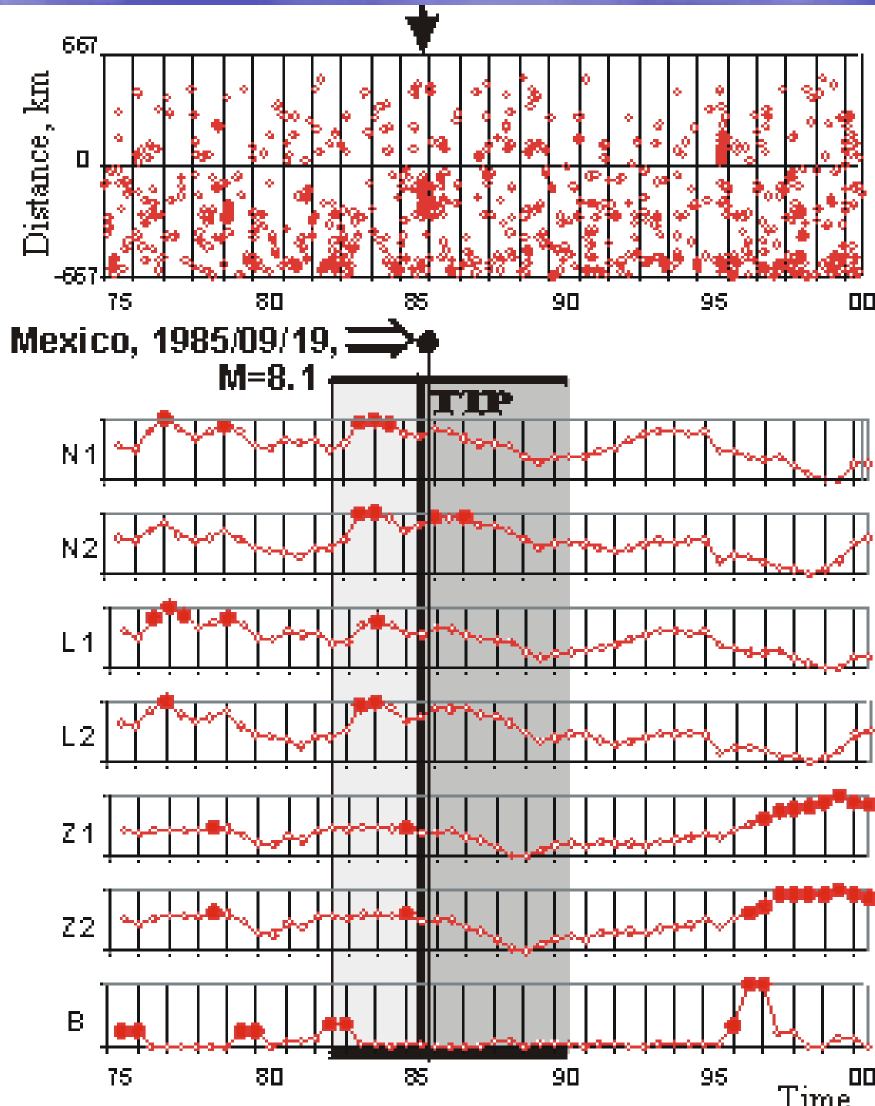
Real-time monitoring (http://www.mitp.ru or http://www.phys.ualberta.ca/mirrors/mitp):

Centers of CI's and Great Earthquakes, 1985-2003

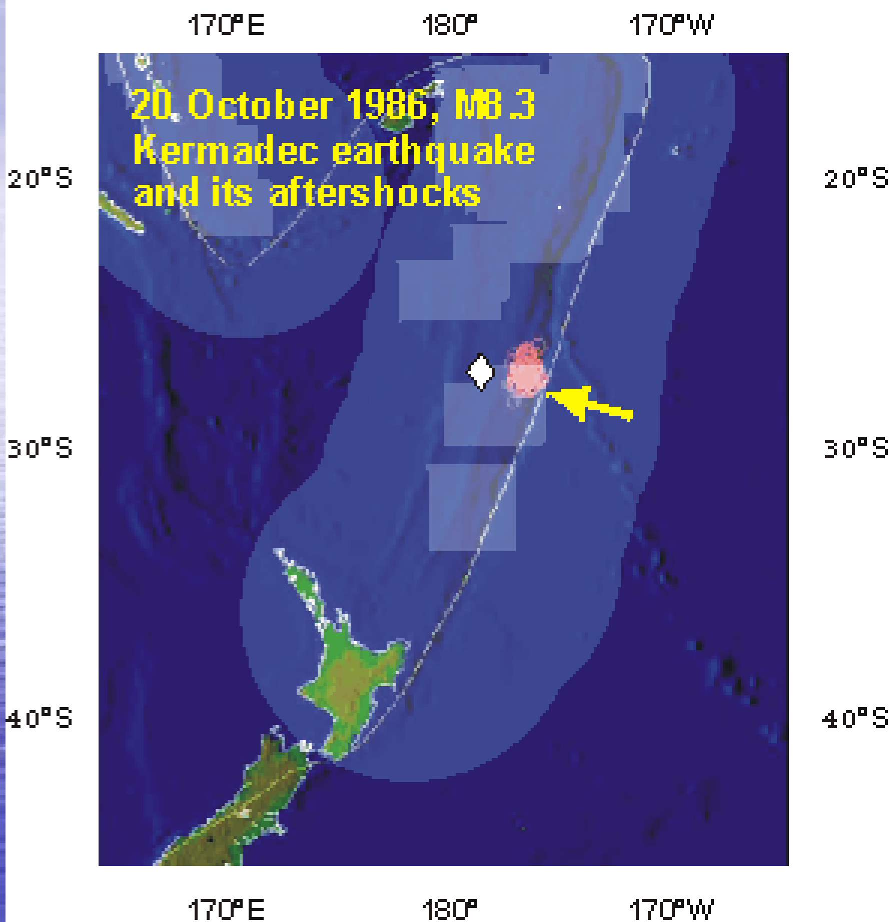


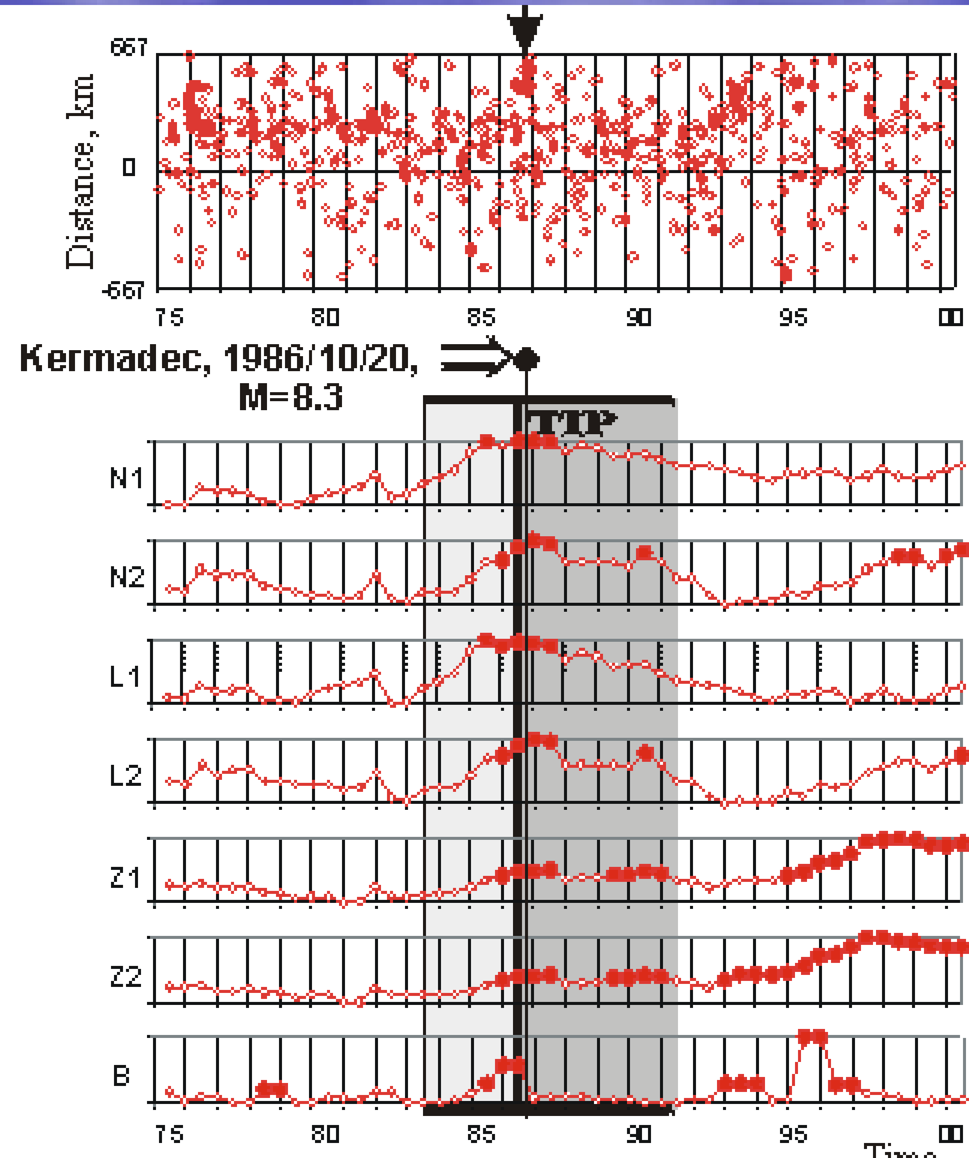
19/09/1985 Mexico Earthquake





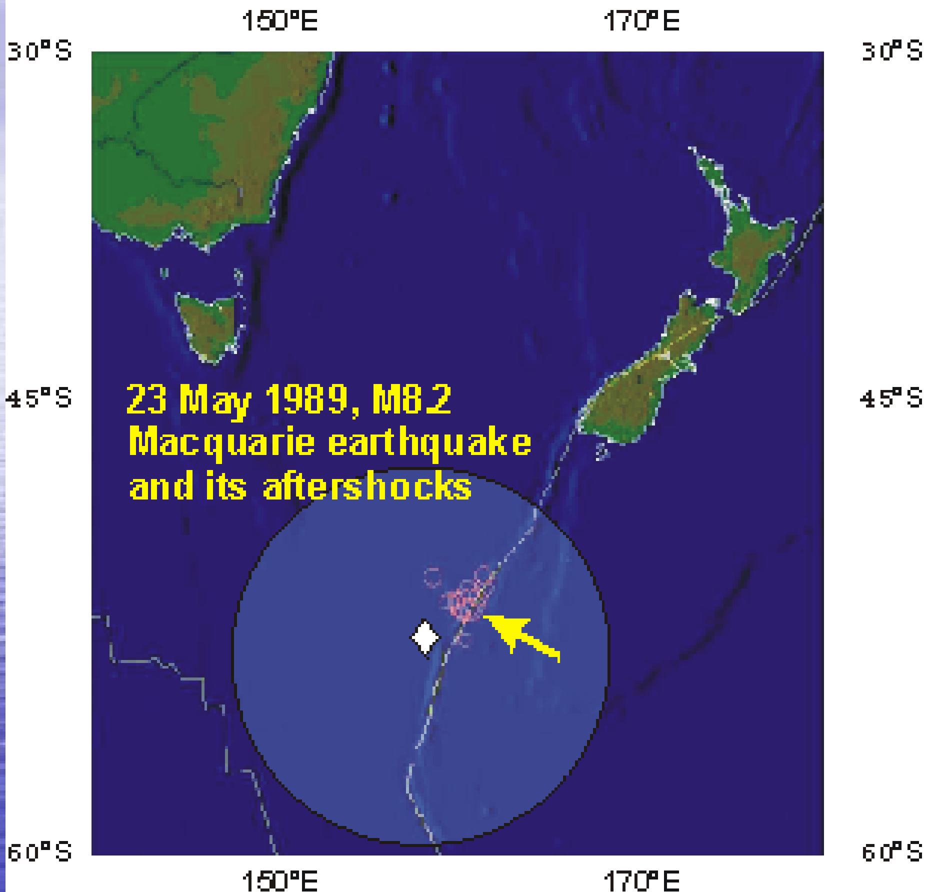
20/10/1986 Kermadek Earthquake

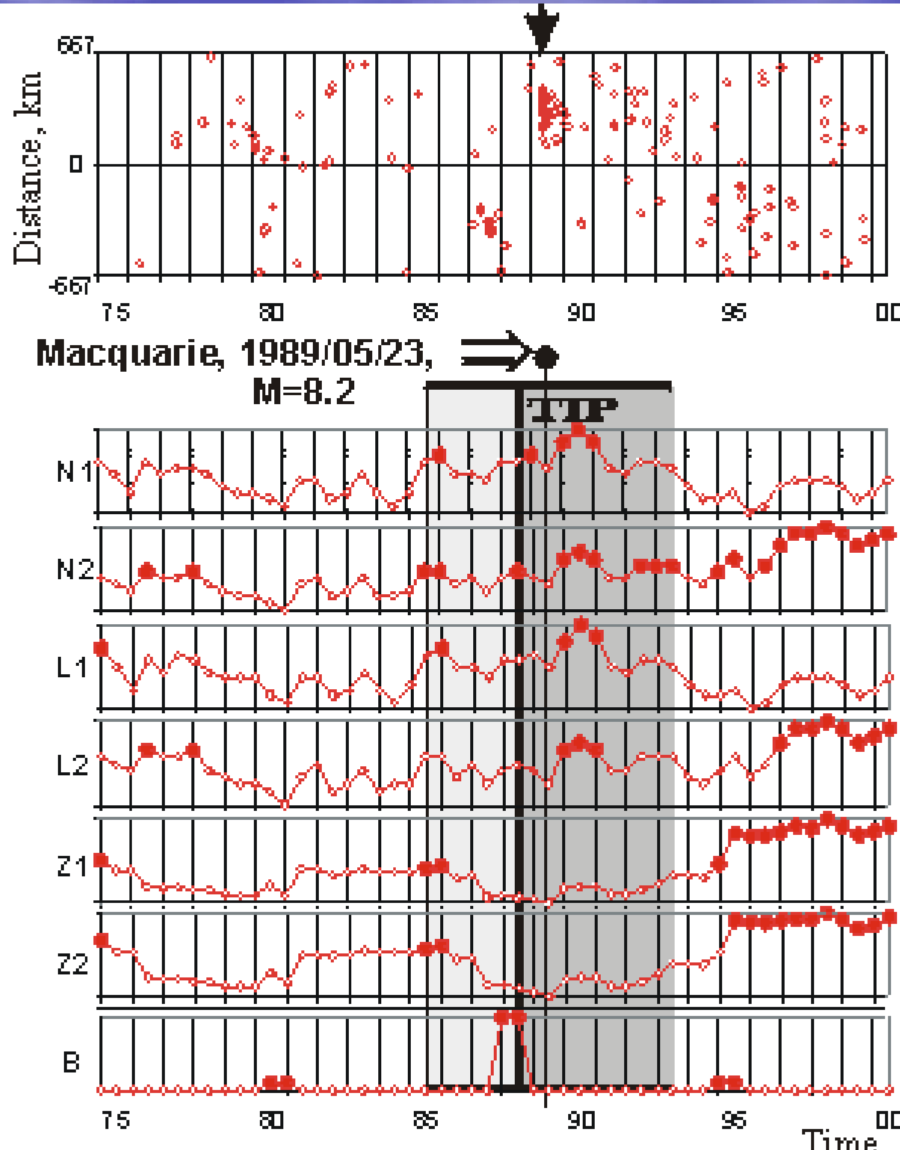




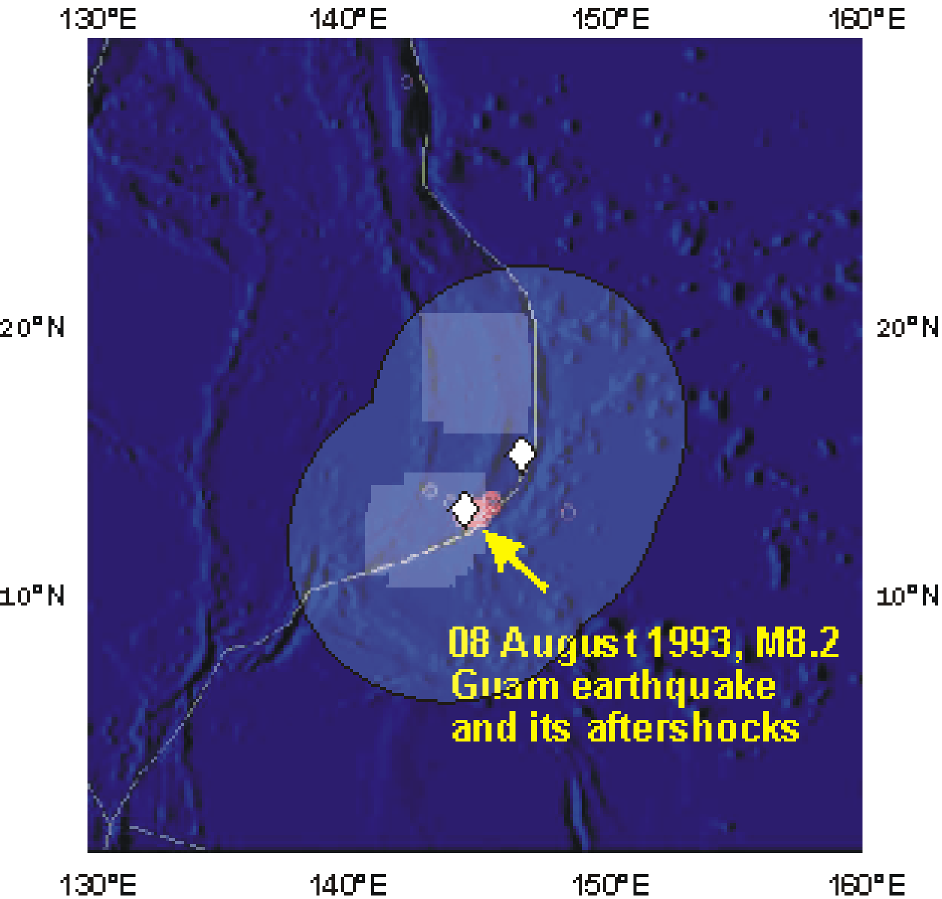
Outside Test Area, NOT COUNTED in the overall statistics

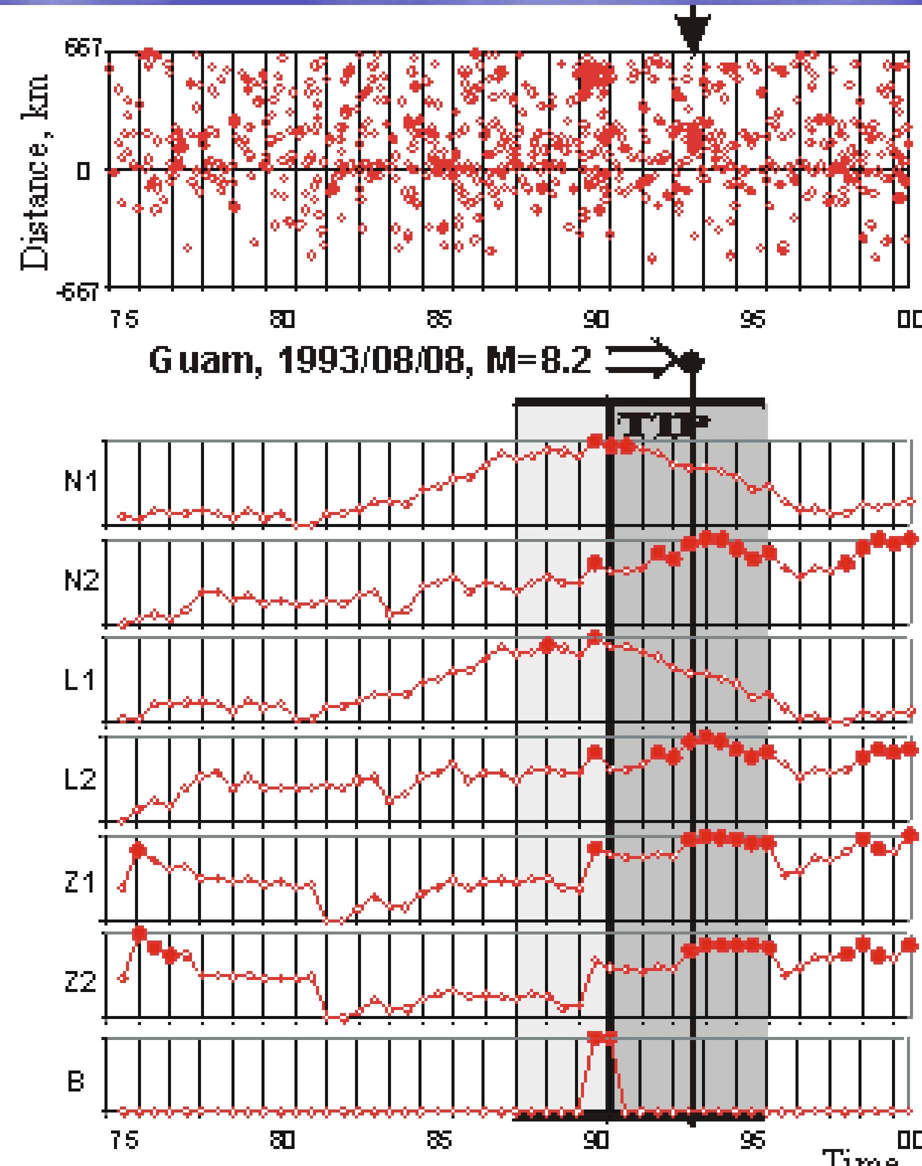
23/05/1989 Macquarie Earthquake





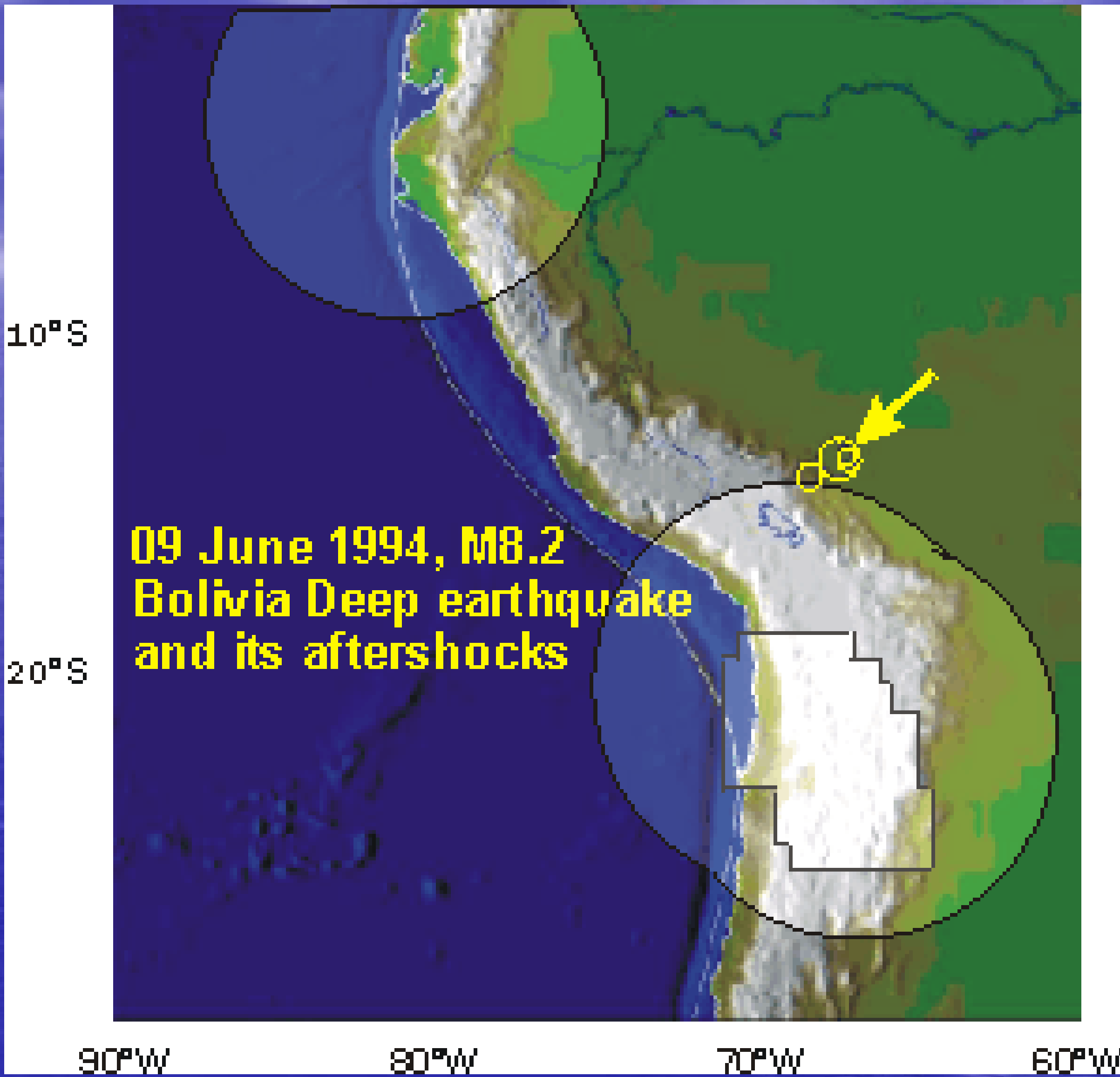
08/08/1993 Guam Earthquake





Outside Test Area, NOT COUNTED in the overall statistics

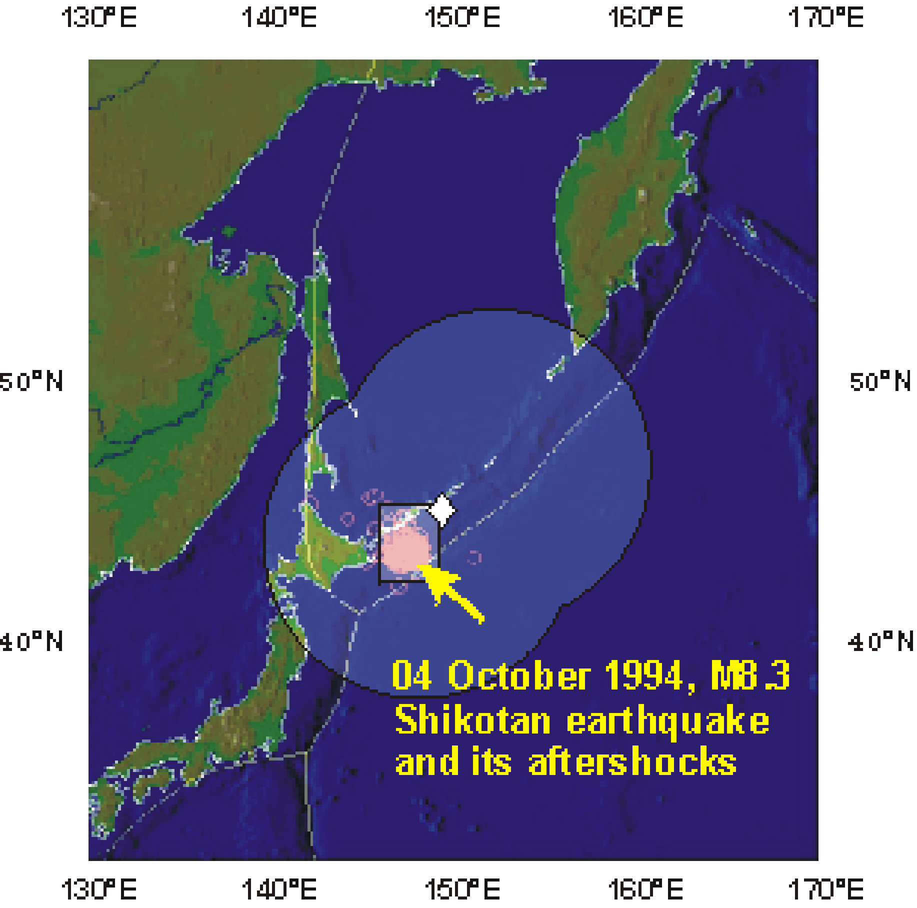
09/06/1994 Bolivia Deep Earthquake

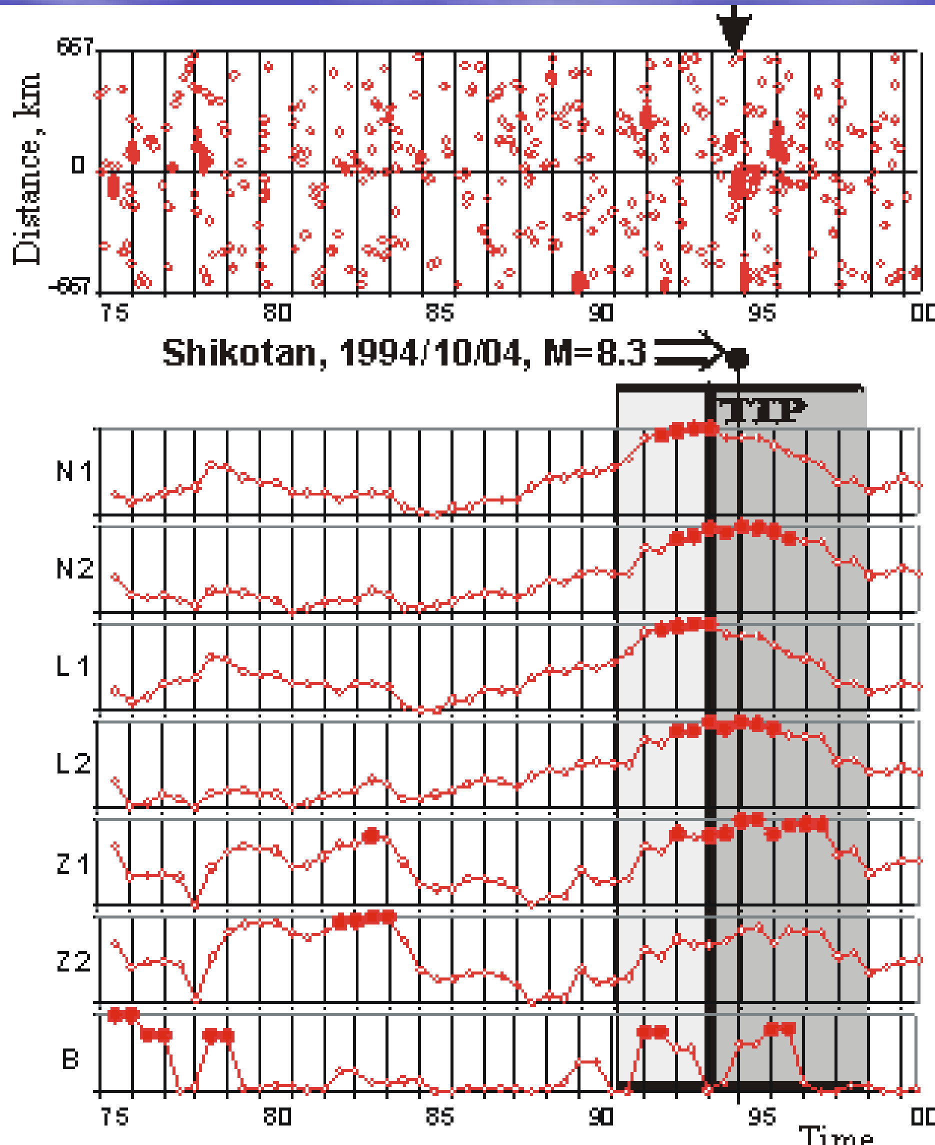


- The Great Deep Bolivia earthquake did occur after the January 10, 1994, magnitude 6.9, depth 595 km earthquake at distance of about 250 km.

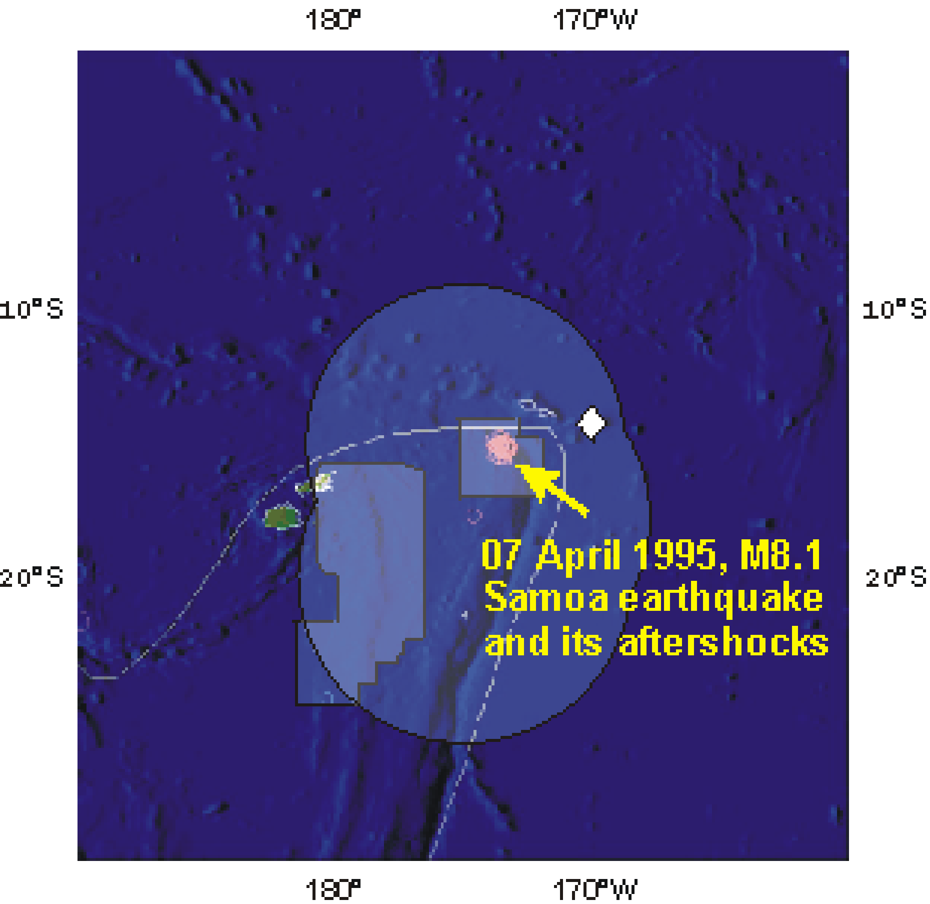
The previous earthquake that deep happened here in 1963.

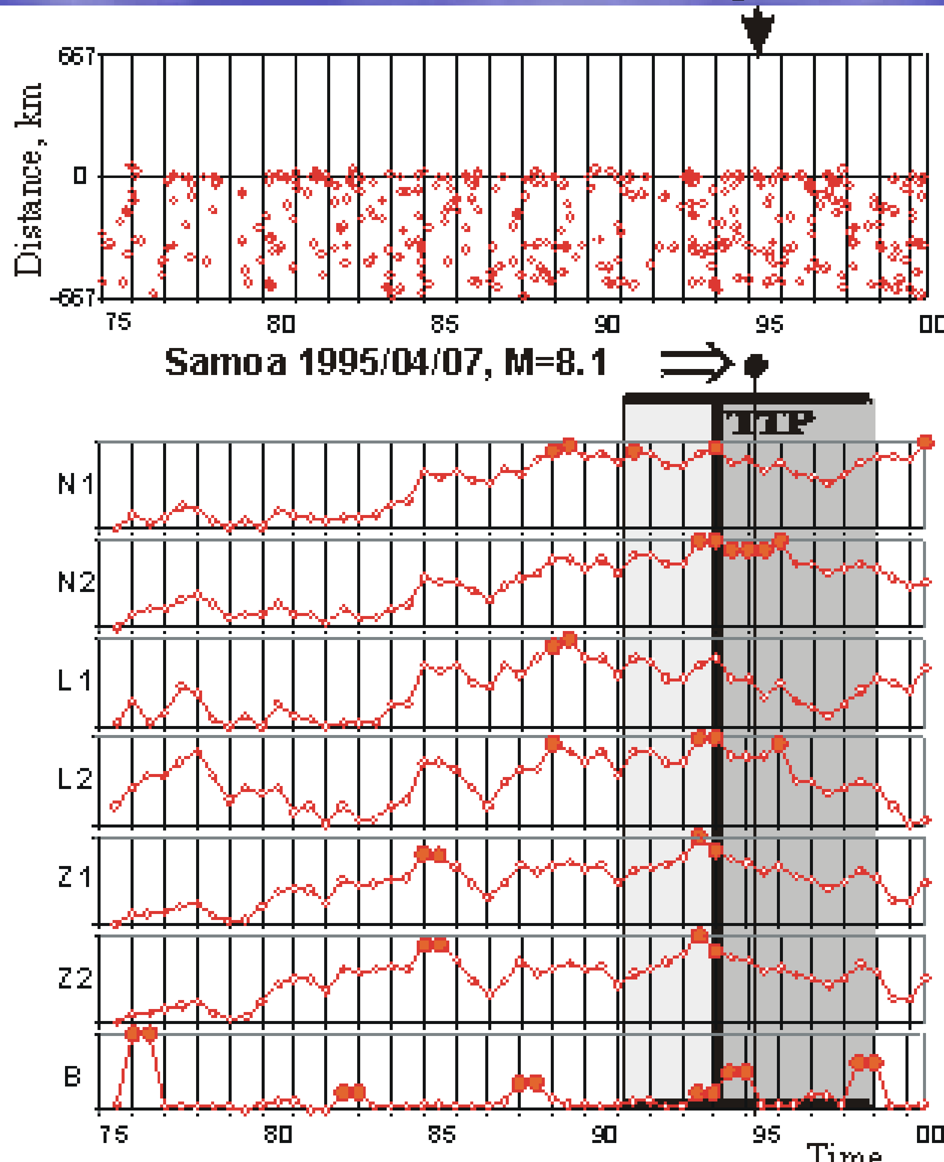
04/10/1994 Shikotan Earthquake



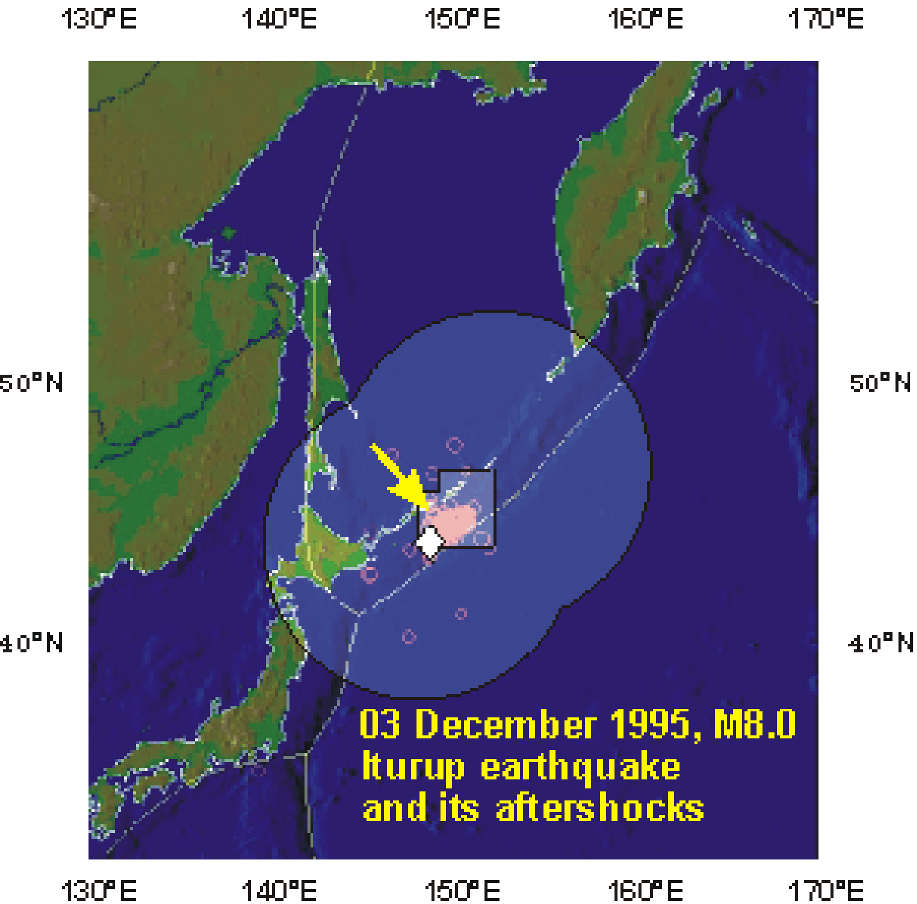


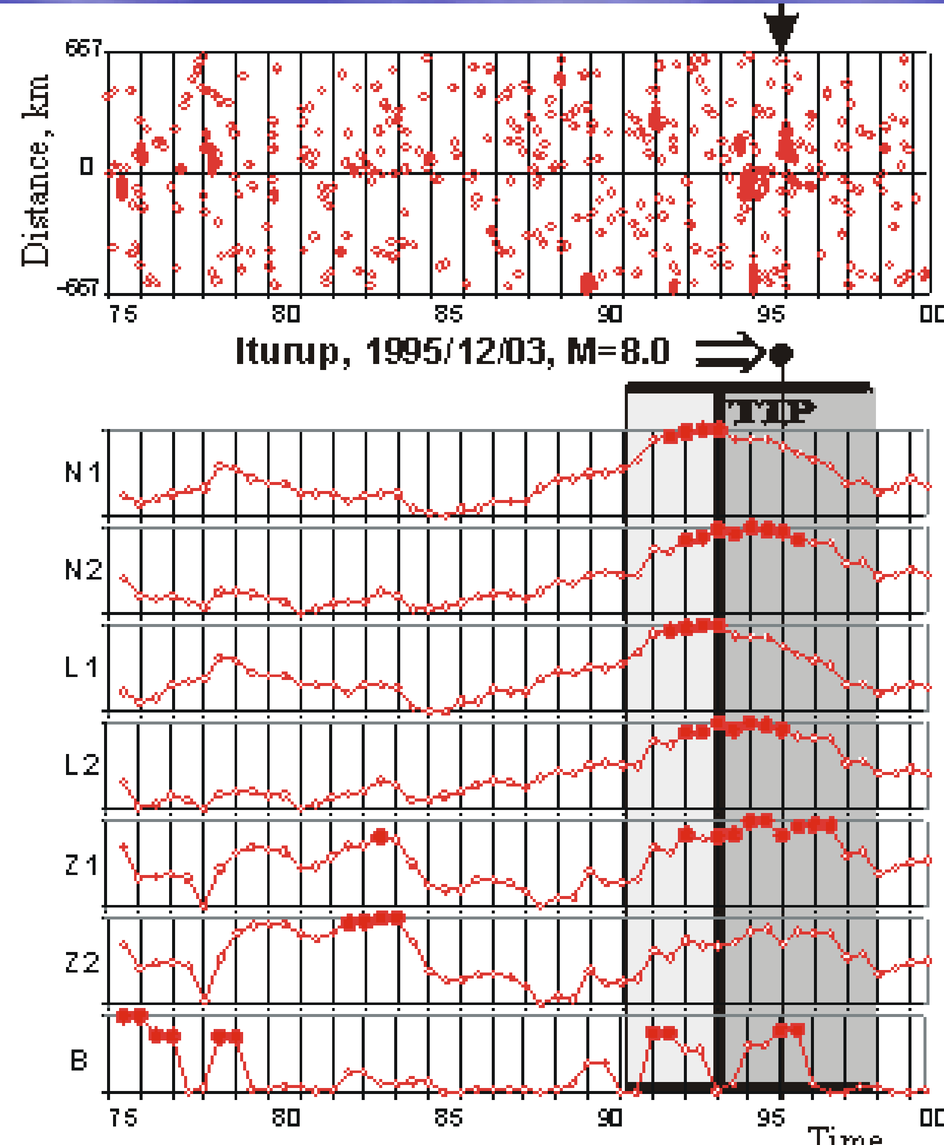
07/04/1995 Samoa Earthquake



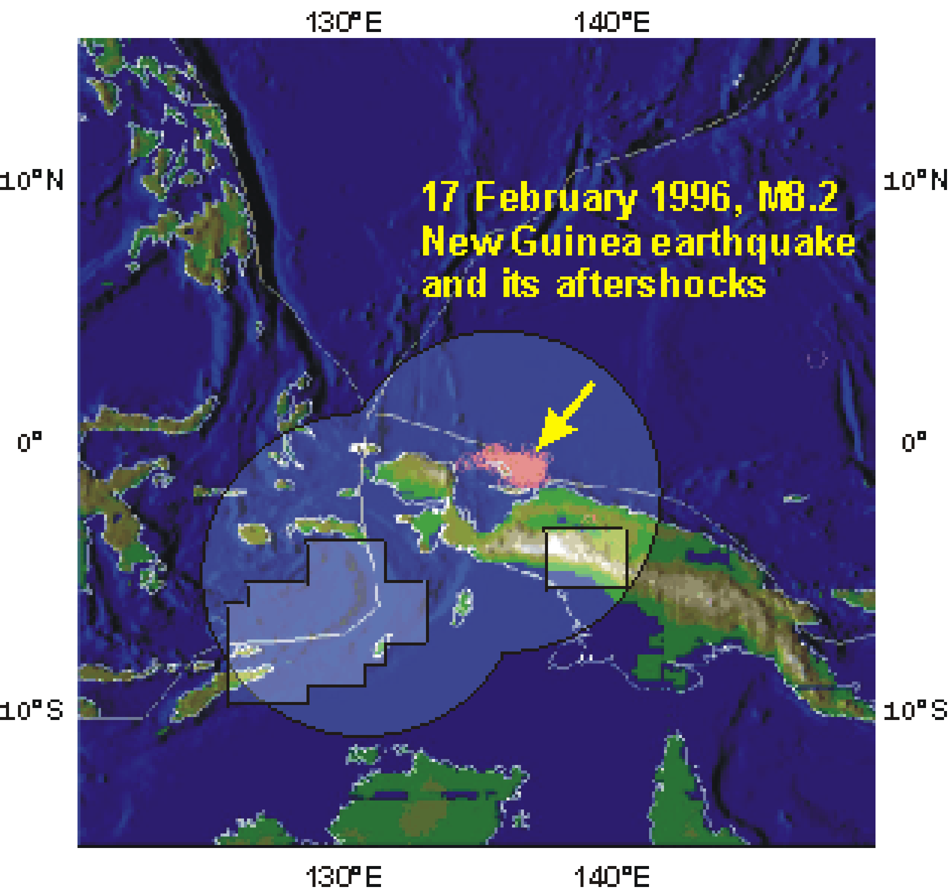


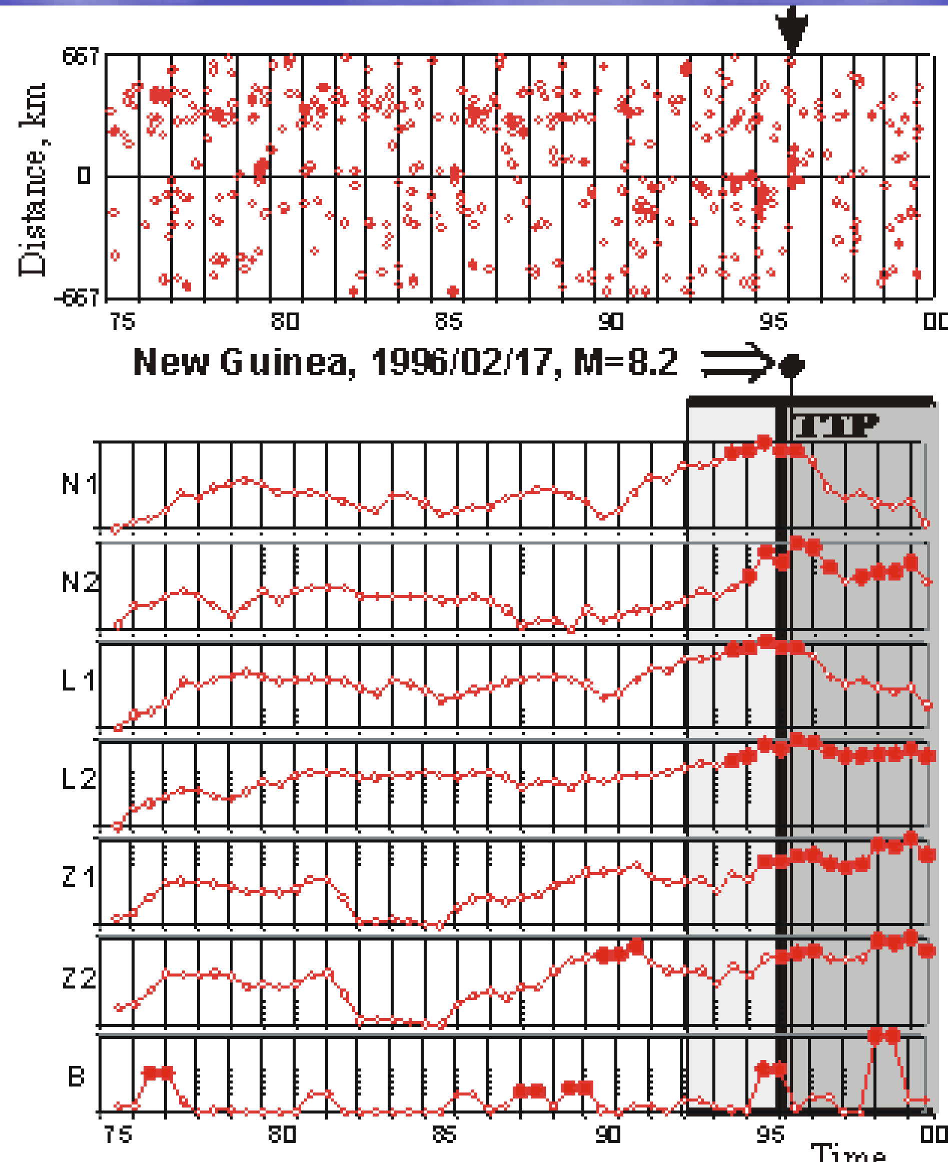
03/12/1995 Iturup Earthquake





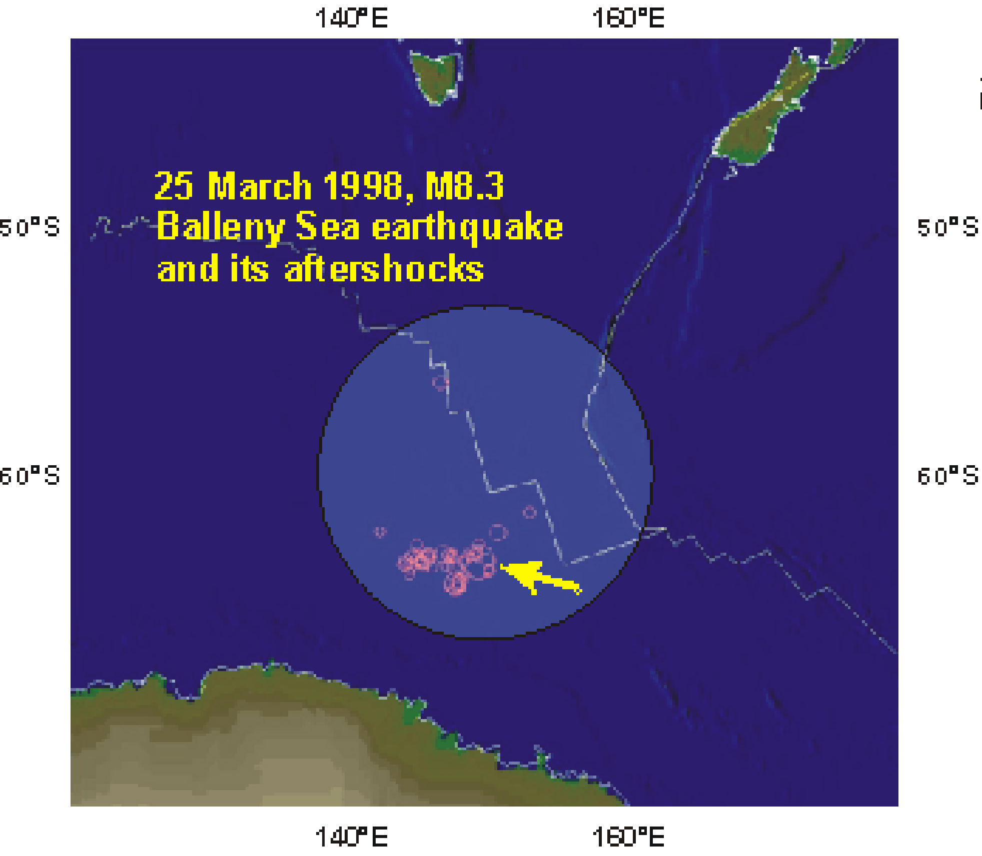
17/02/1996 New Guinea Earthquake

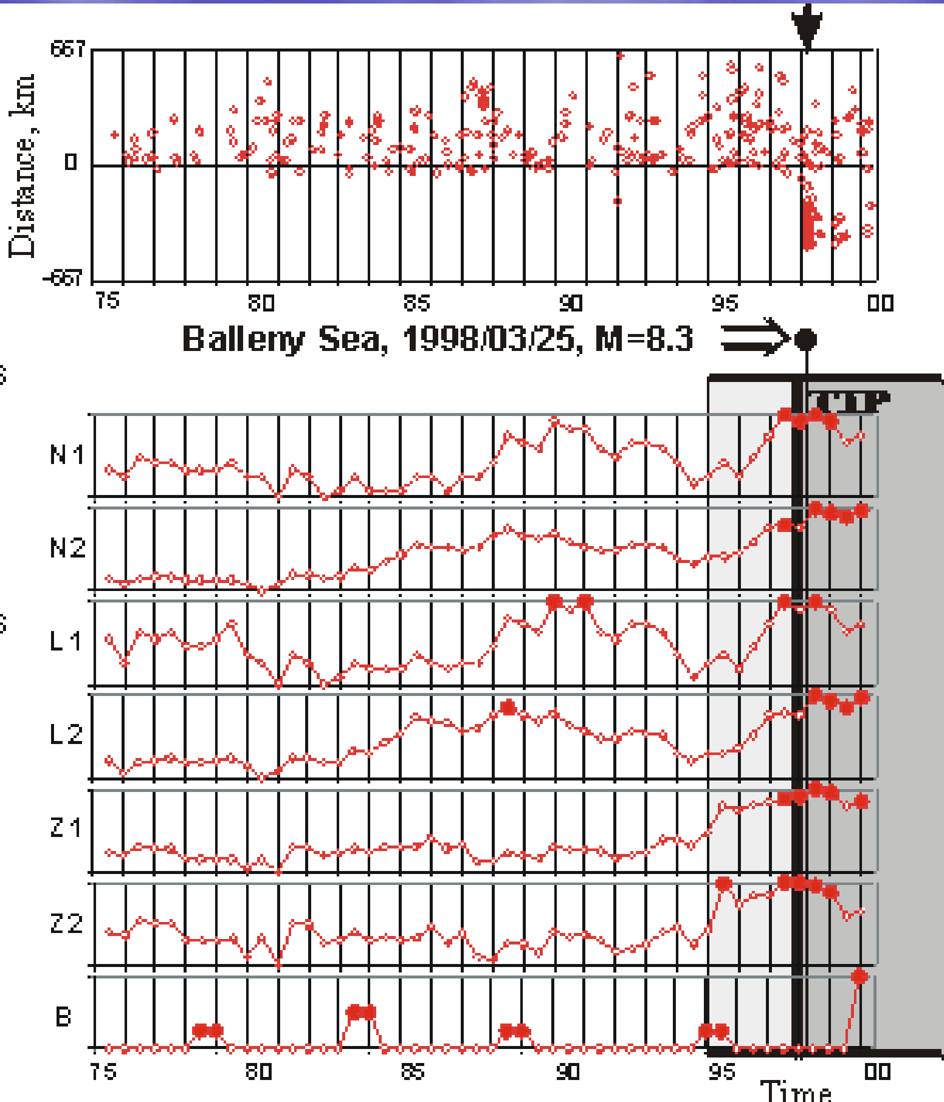




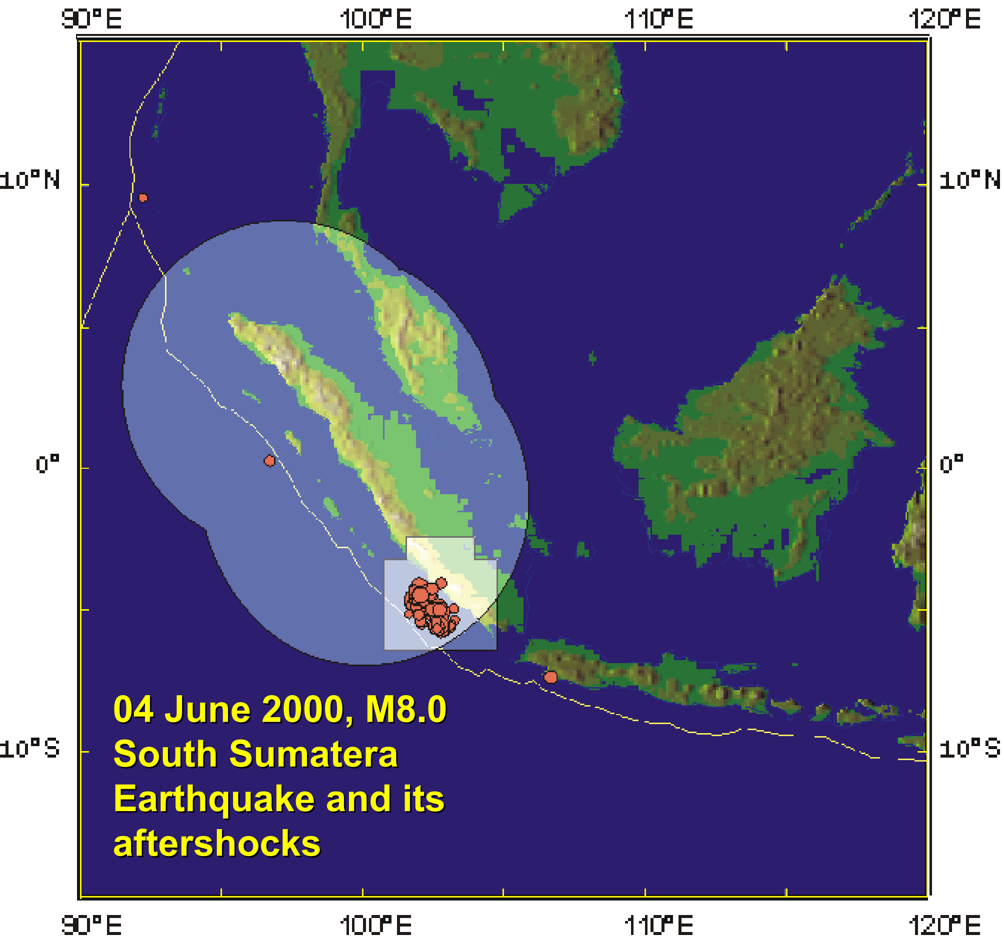
Outside Test Area, NOT COUNTED in the overall statistics

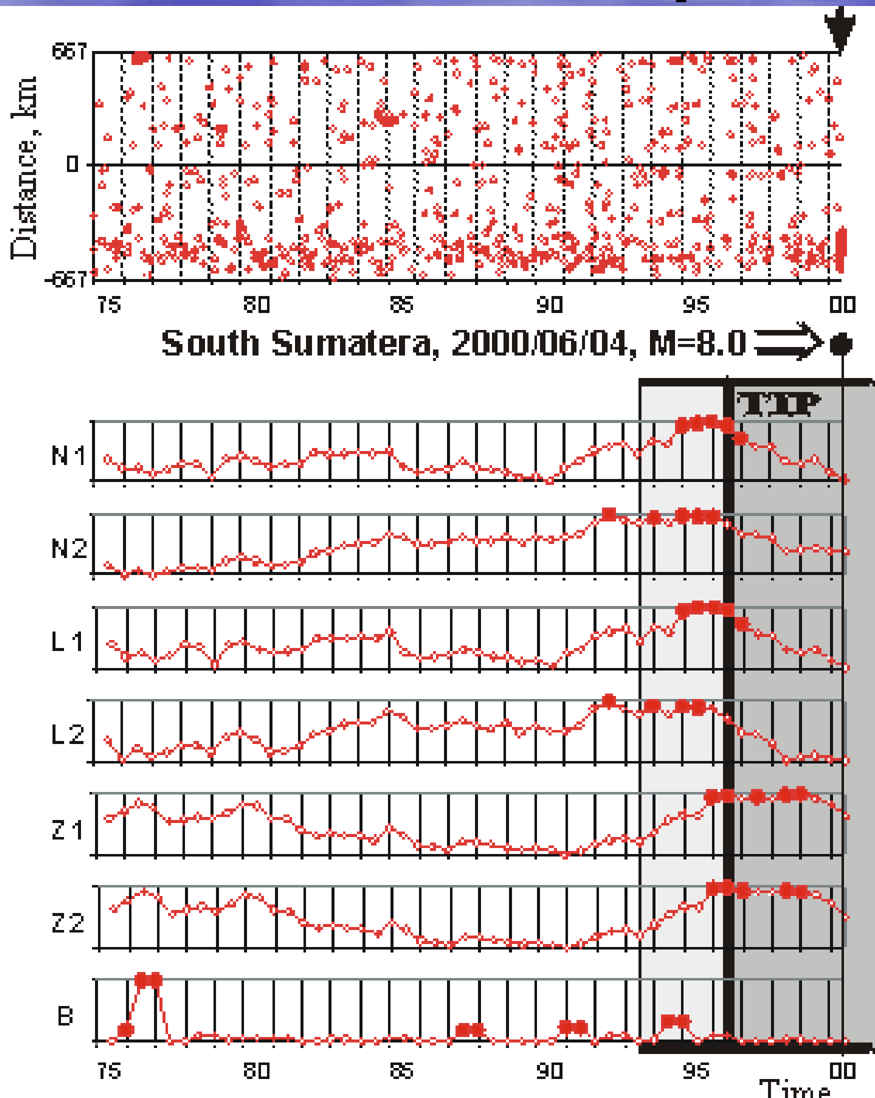
25/03/1998 Balleny Sea Earthquake



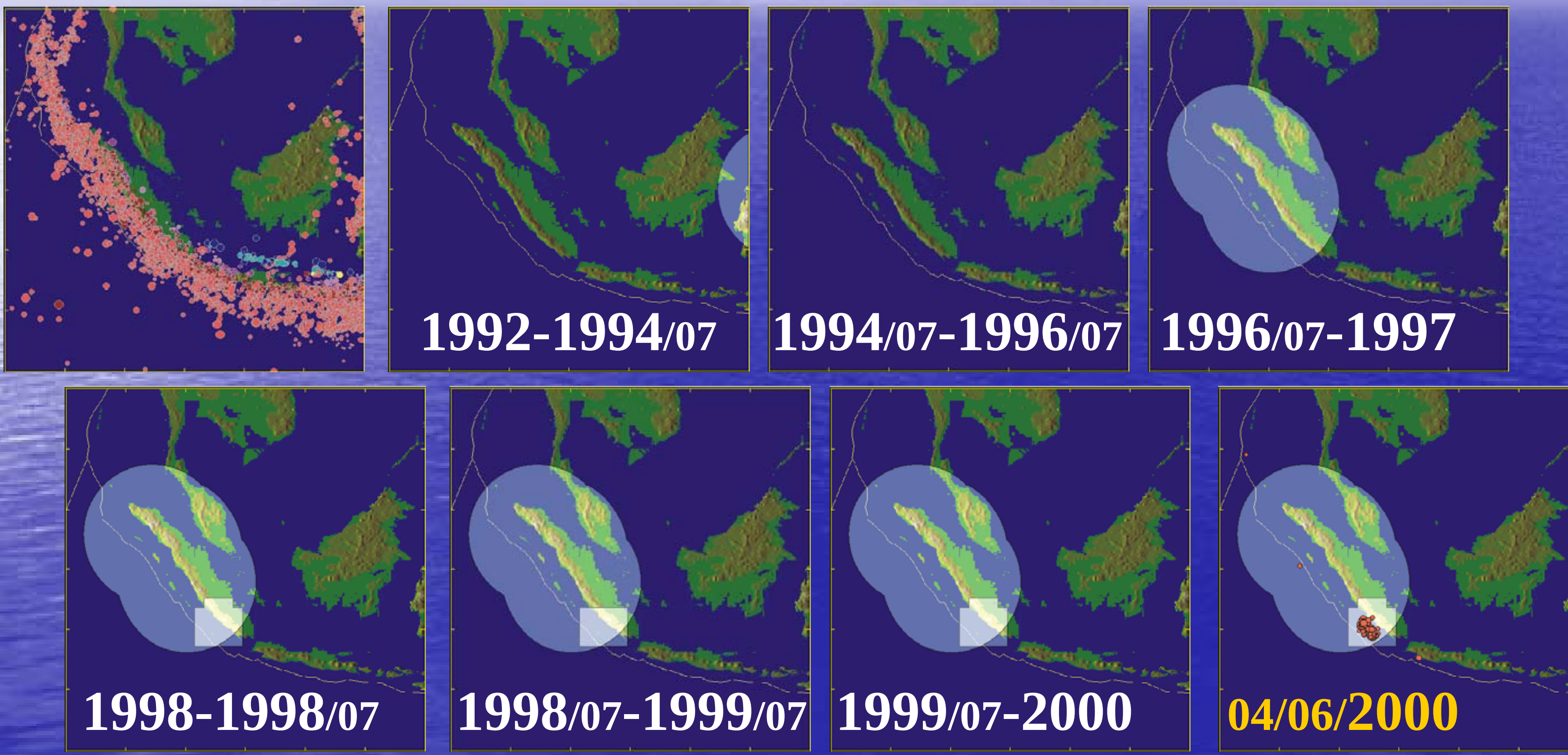


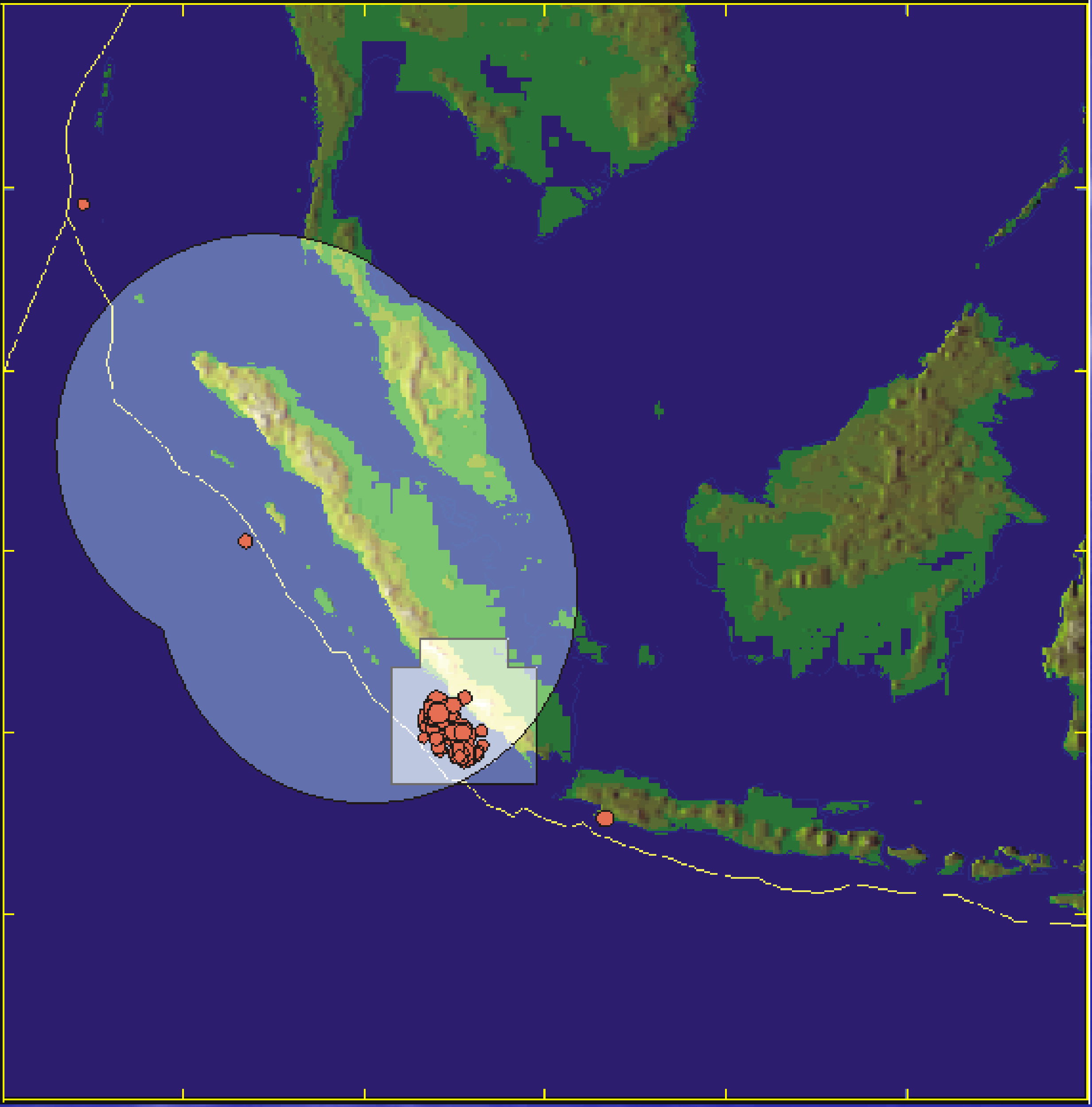
04/06/2000 South Sumatera Earthquake





Case history of the South Sumatera Earthquake



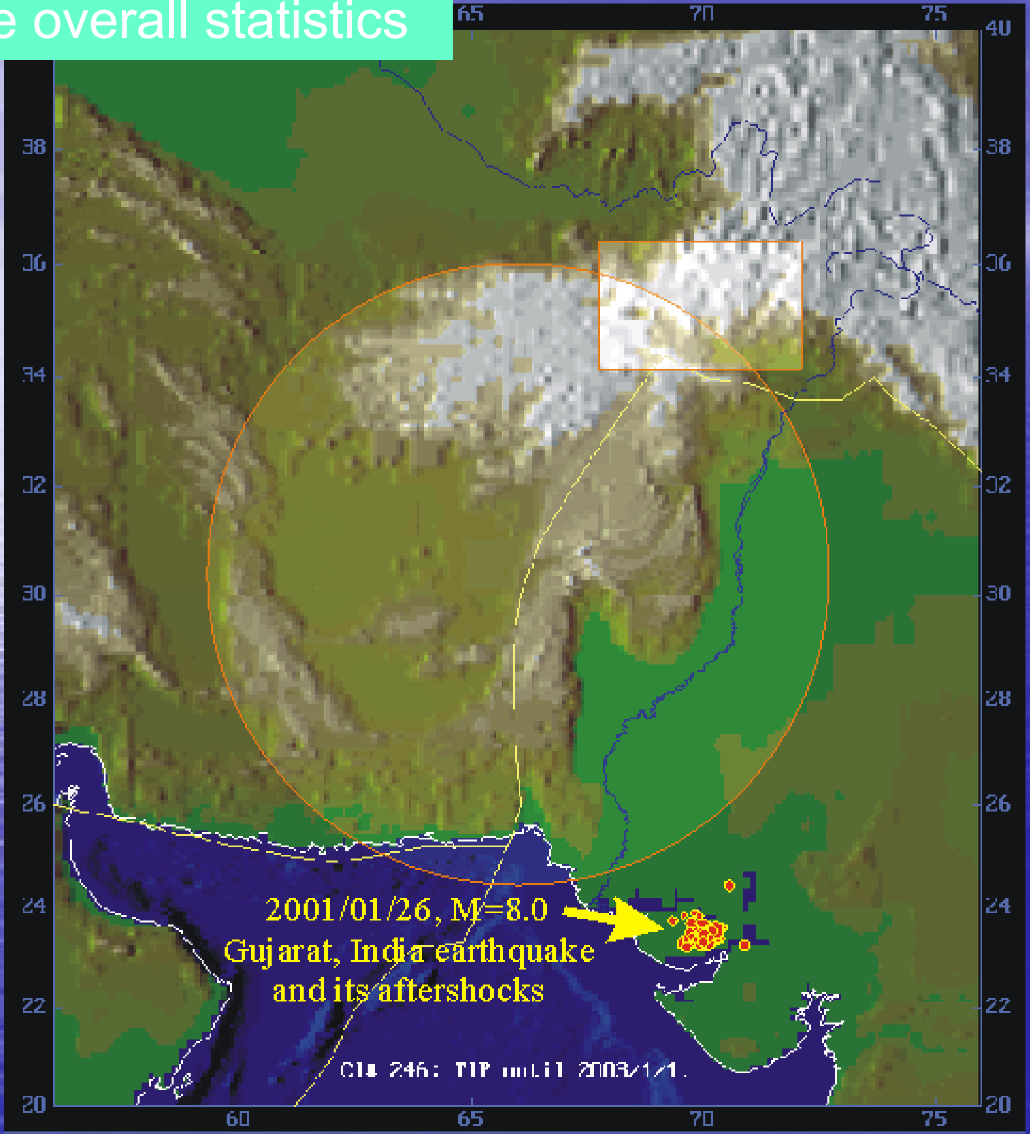


Seismic events that big were reported in the Indian Ocean subduction zones only twice in the 20th century:
These are
the 1941 Andaman, Ms8.1 and
the 1977 Sumbawa, Ms8.0 earthquakes.

This implies local probability gain of more than 20

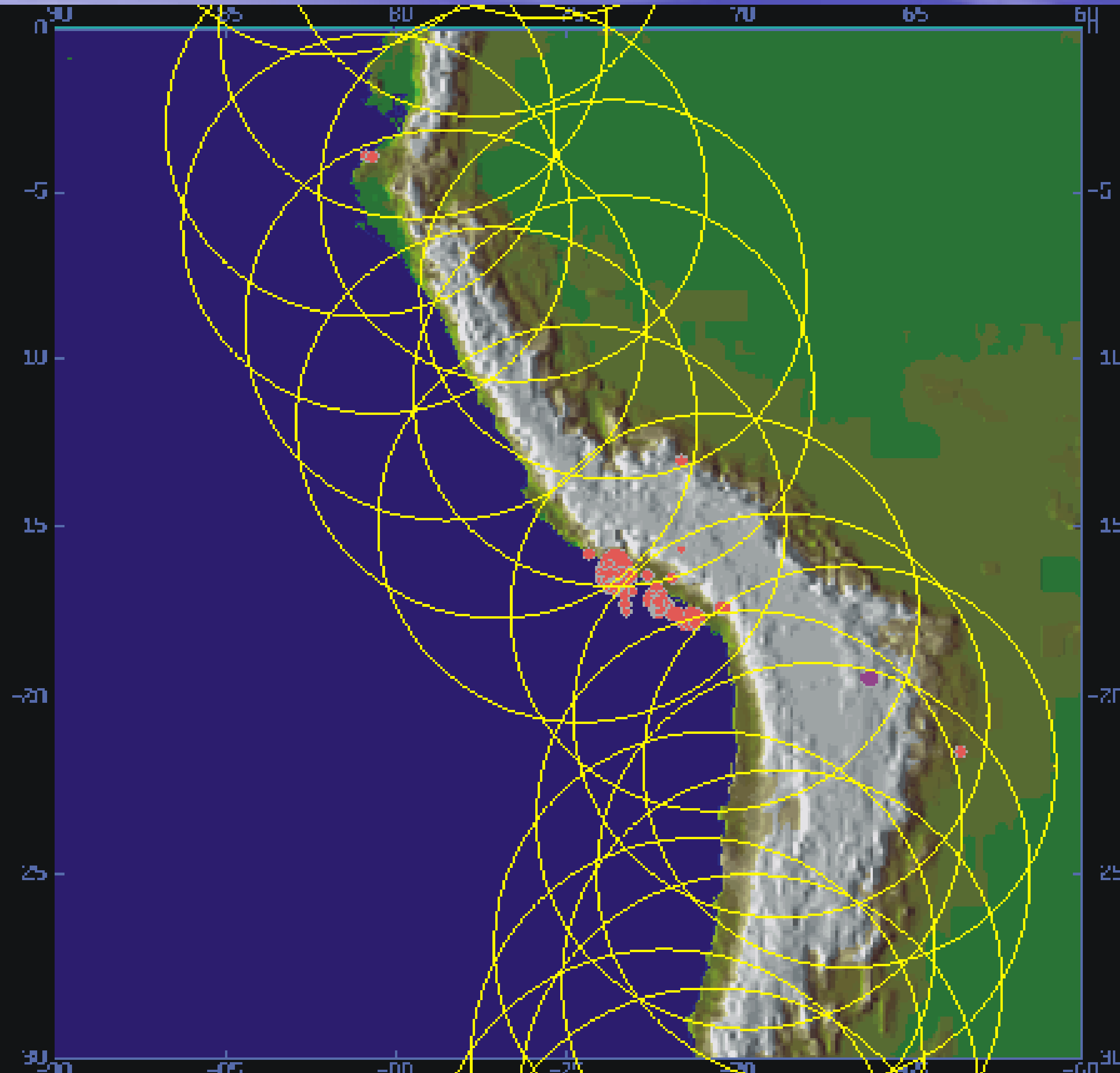
Outside Test Area, NOT COUNTED in the overall statistics

26/01/2001 Gujarat, India earthquake



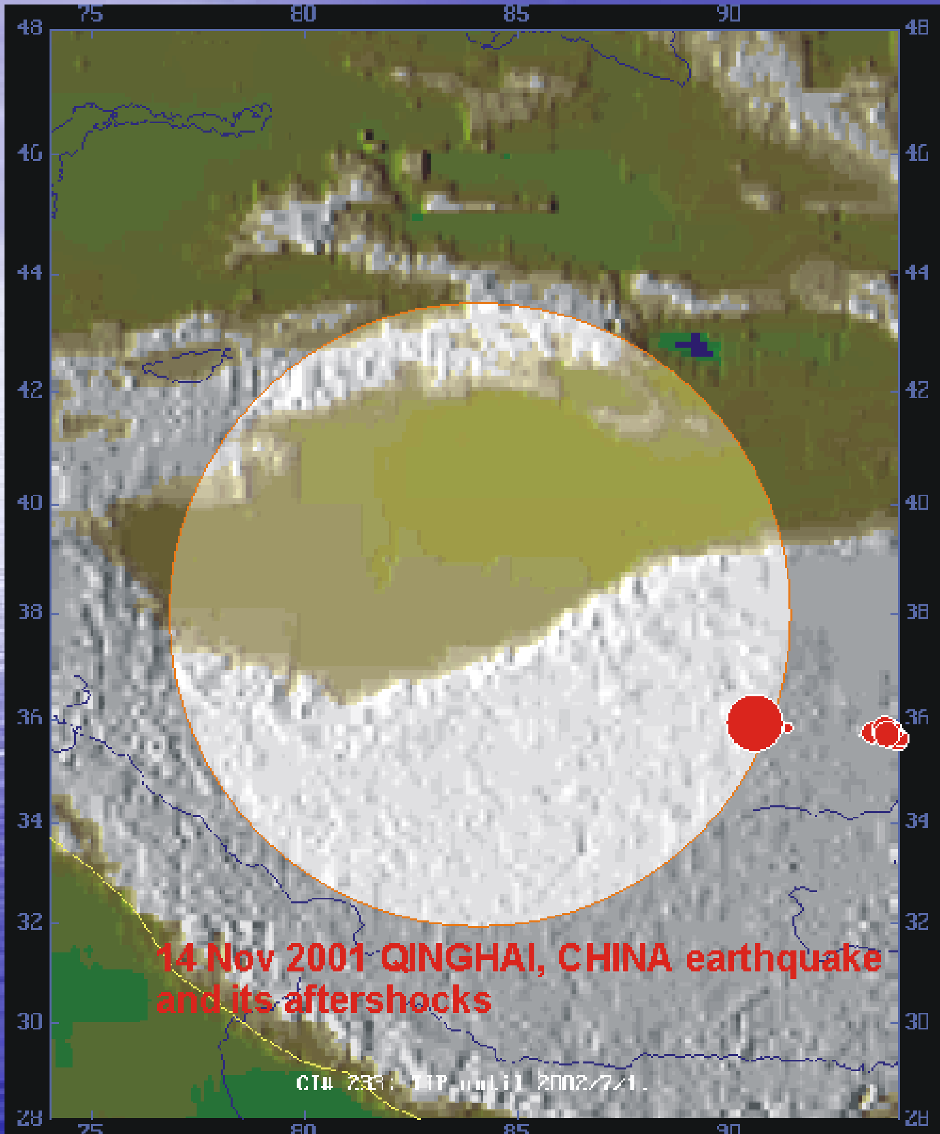
The 26 Jan 2001 Gujarat, India earthquake is just outside the area, where the NEIC data permits to run the original version of the M8 algorithm. Note that one of the circles, nearest to the epicenter of the 2001 Gujarat earthquake was in state of alarm, although the MSc predicts an opposite side of it as the most dangerous area.

23/06/2001 earthquake NEAR COAST OF PERU



This earthquake is the first failure-to-predict in M8-MSc testing aimed at magnitude 8.0+.

14/11/2001 QINGHAI, CHINA earthquake



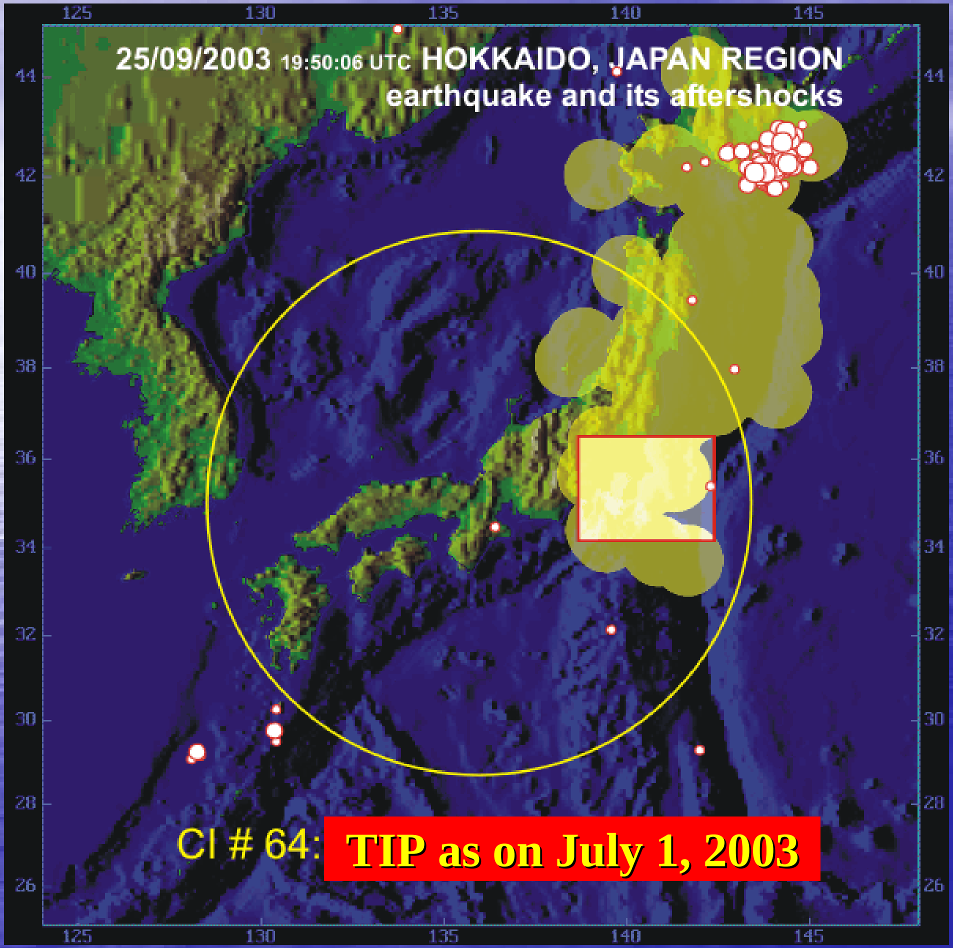
No earthquake of such magnitude had been ever reported inside CI#233 before the 2001 Qinghai earthquake.

The largest one in the 20th century has magnitude MS= 7.9 and happened on November 08, 1997 four months after declaration of the M8 alarm in our Test. (The next largest magnitude is 7.3.)

A conservative estimation of probability gain is about 20, so that the prediction is not trivial indeed.

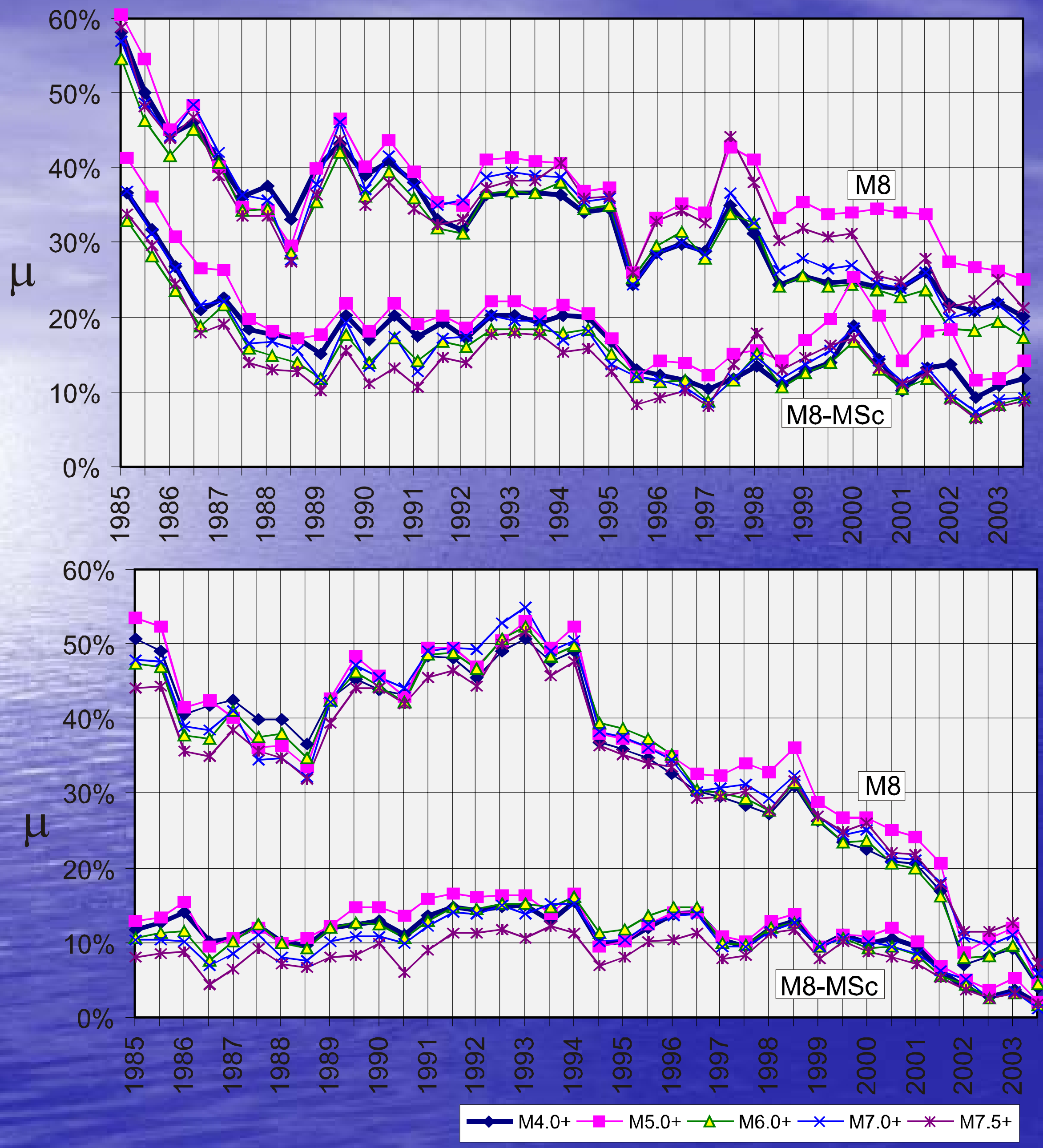
The nearest magnitude 8.0+ earthquake happened on November 18, 1951 near Lhasa, Xizang (Tibet) 375 miles (600 km) south of the November 14, 2001 epicenter.

25/09/2003 19:50:06 UTC HOKKAIDO, JAPAN REGION earthquake



This is the second failure-to-predict the world largest earthquakes in course the Global real-time prediction experiment aimed at M8.0+ events.

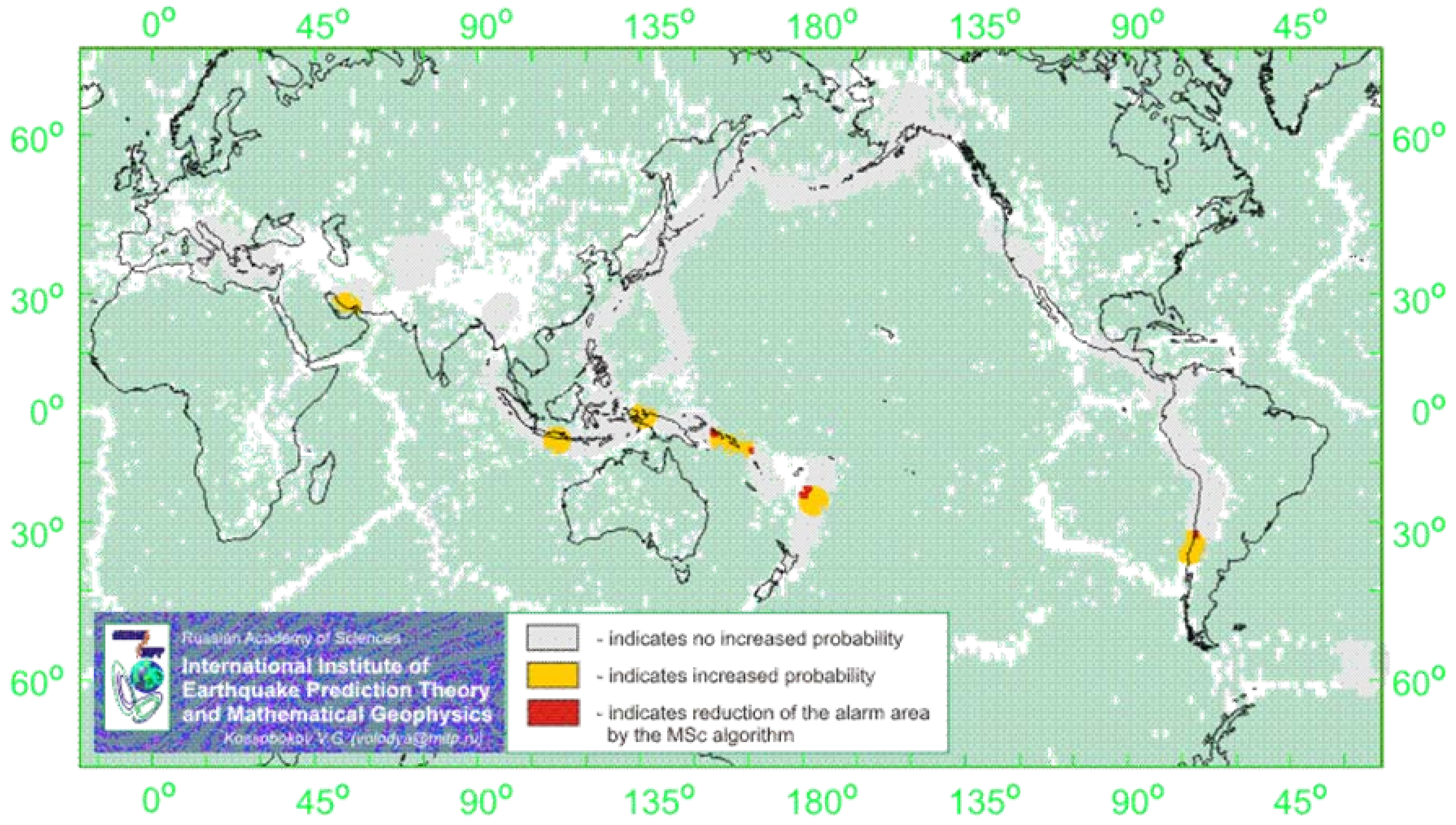
Can we exclude a possibility that the *Time of Increases Probability*, TIP, in CI#64 is related to the occurrence of 25 September 2003 great quake? The analysis at a shorter-term lower-magnitude scales [*Shebalin, Keilis-Borok, Zaliapin, Uyeda, Nagao, Tsybin, 2003. Short-term Premonitory Rise of the Earthquake Correlation Range. In IUGG2003, June 30 – July 11, 2003* **] suggests that, perhaps, we can not.**



The percentage of alerted area as a function of time for M8.0+ (above) and M7.5+ (below).

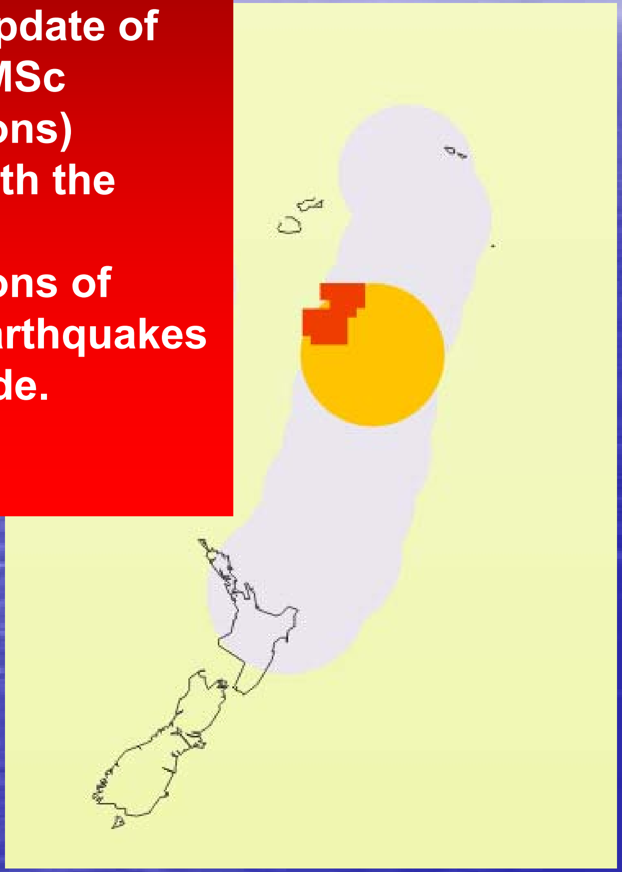
The obtained estimates are based on the counts of magnitude 4 or more and 5 or more earthquakes in the period from 1964 through 1984, while the counts of magnitude above 6.0, 7.0, and 7.5 in 1900-1984

Regions of Increased Probability of Magnitude 7.5+ Earthquakes as on July 1, 2002 (*subject to update on January 1, 2003*)



What was predicted…

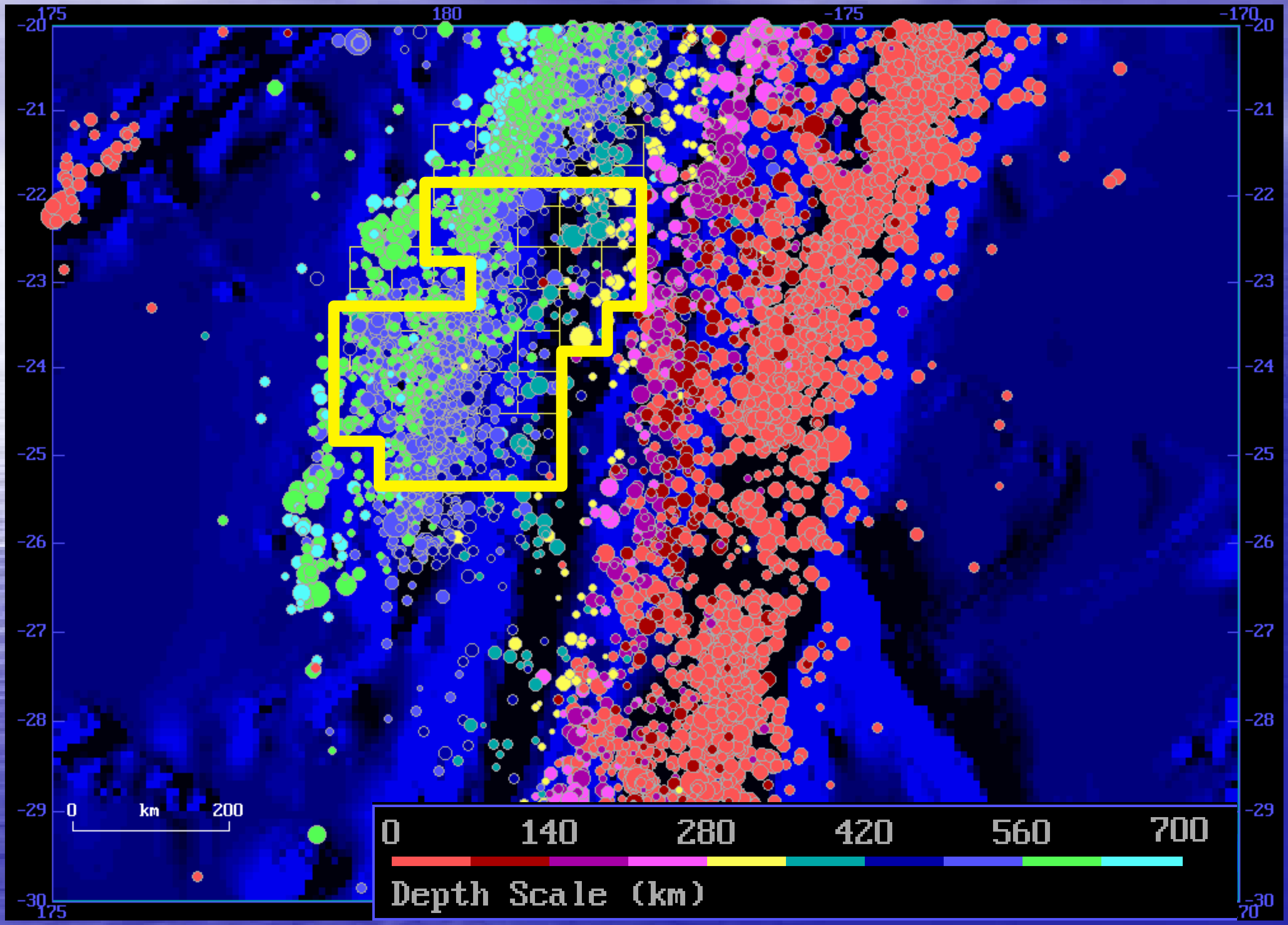
Sent on Monday, July 15, 2002 (Subject: The 2002b Update of the M8-MSc predictions) along with the updated predictions of major earthquakes worldwide.



- Earthquake(s) with magnitude 7.5 or more will occur in CI #5 (yellow) during the time period from July 2002 through July 2003.
- In the second approximation the MSc algorithm has identified the area (red) that stretch between

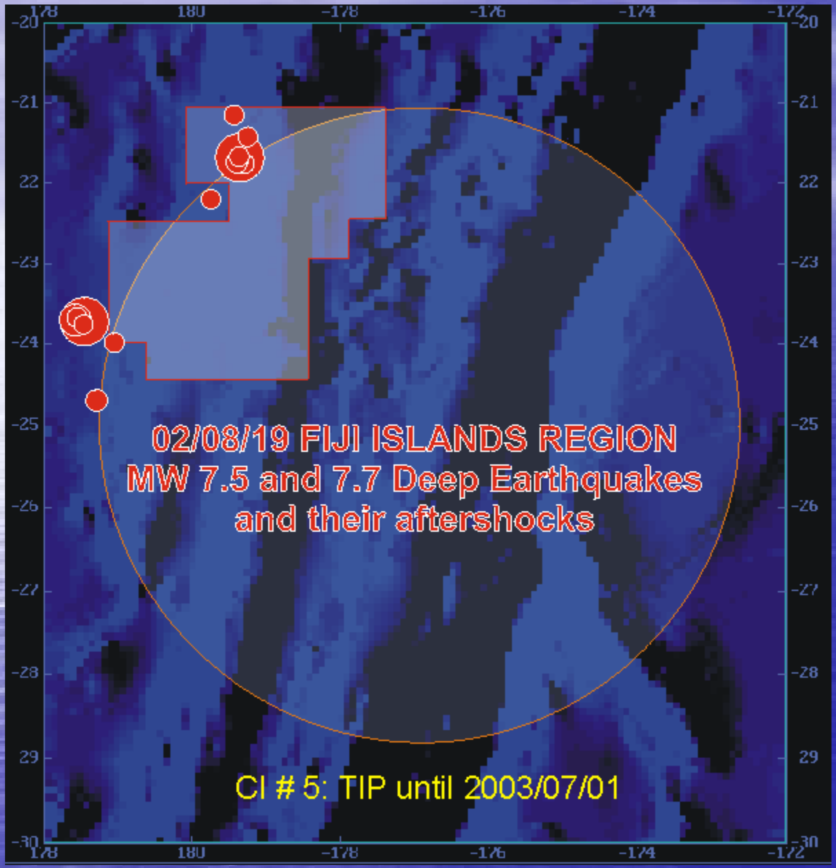
  24.52S - 21.16S and 178.76E - 177.53W.

What was predicted...



- The position of the M8-MSc alarm that narrow down substantially the prediction area suggested the occurrence of the great deep earthquakes **(depth of about 240-700 km).**

What happened...



- EARTHQUAKES:

  Origin times -

  2002/08/19 11:01:01
  2002/08/19 11:08:25 ;

  Coordinates –

  21.80S 179.49W
  23.85S 178.41E;

  Depths - 586.8 and 693.7 km;

  Magnitudes –

  MwGS (MeGS)
  7.5 and 7.7 (7.7 and 7.4);

  F-E Regions –

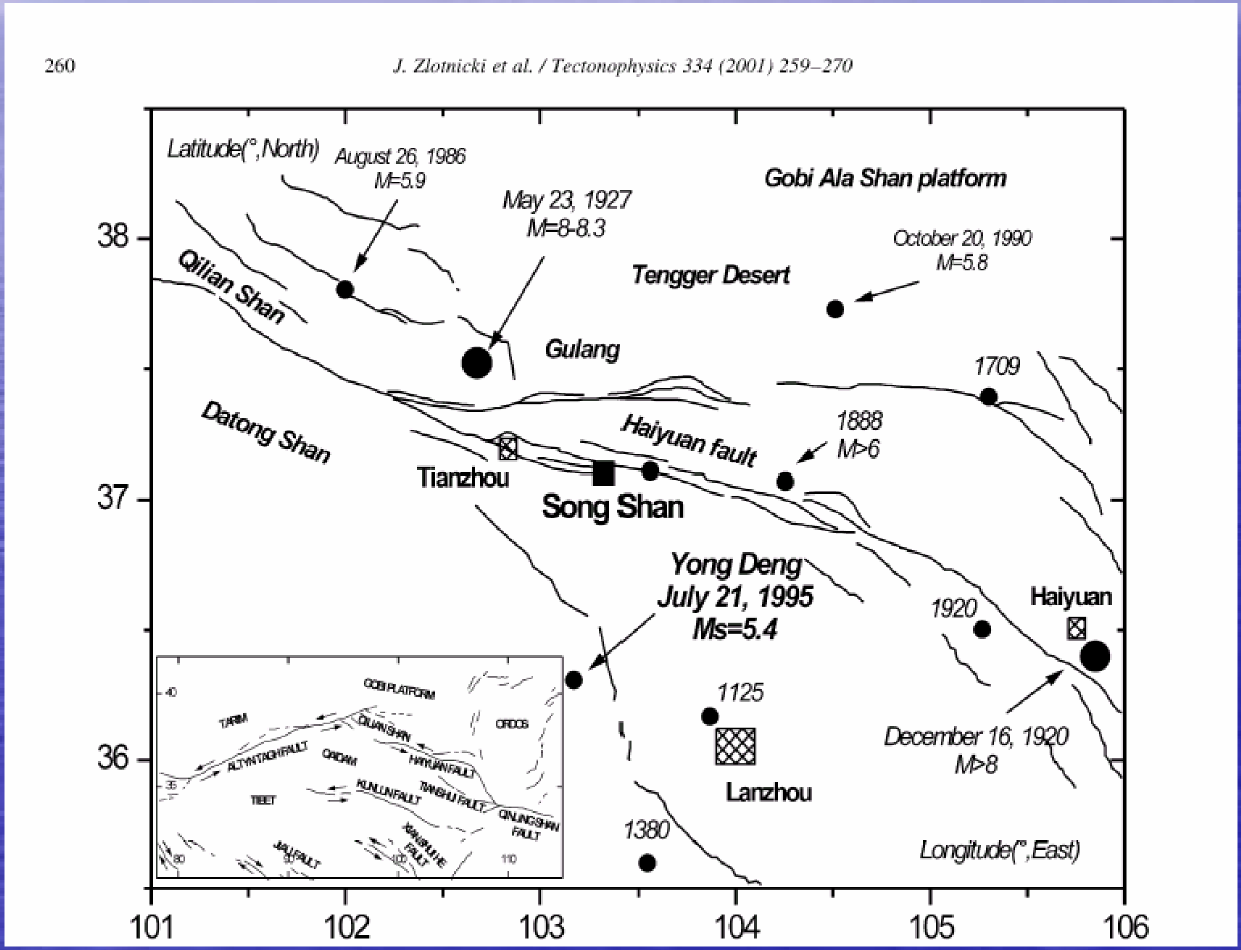
  FIJI ISLANDS REGION and
  SOUTH OF FIJI ISLANDS.

The two August 19 main shocks mark both northern and southern edges of the prediction area. Does it mean that sometimes exact prediction is not possible? This reduction of the uncertainty provides probability gain of more than 25.

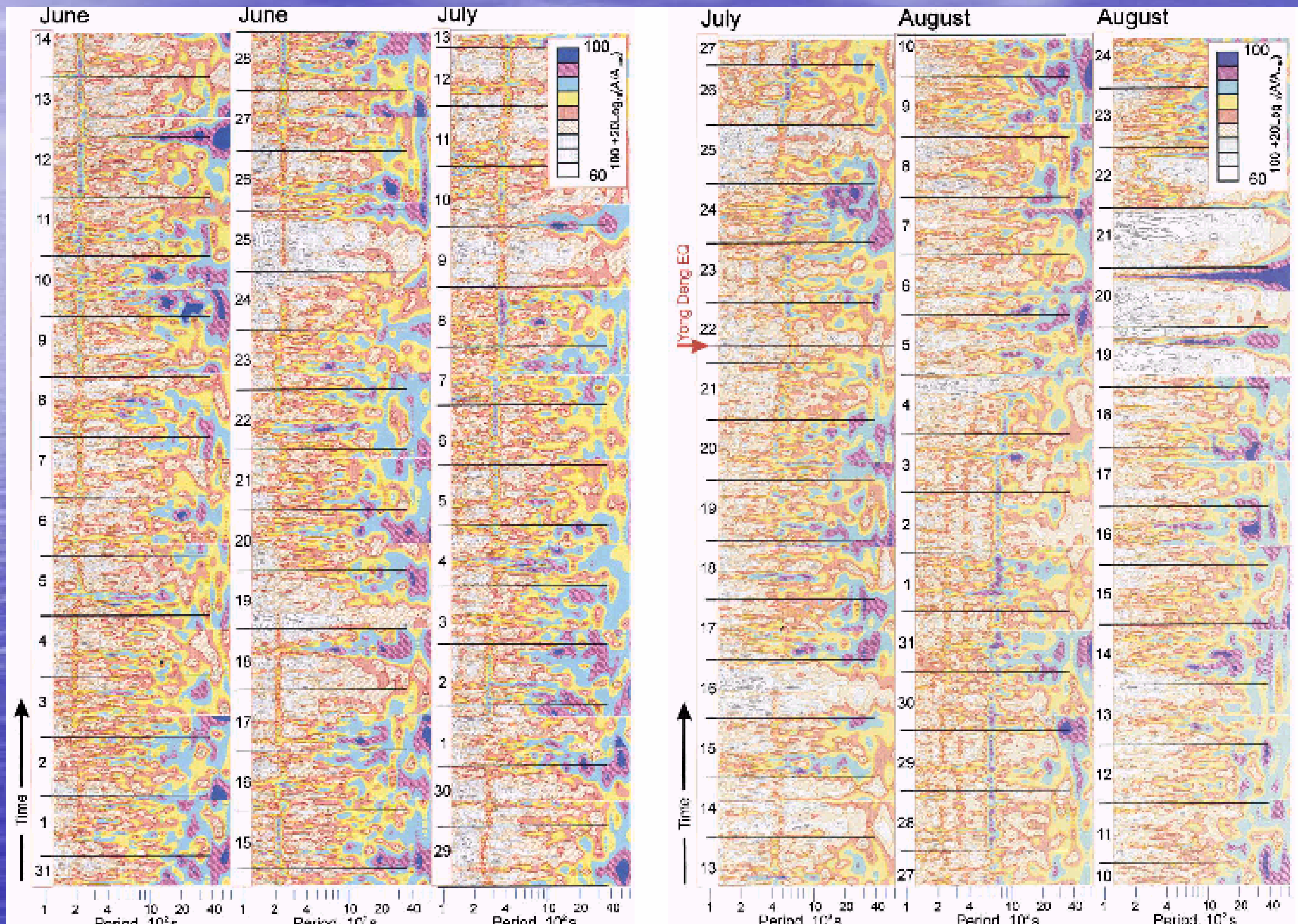
Thus, the accuracy achieved by M8 and MSc algorithms in the on-going Global testing is intermediate in time domain and varies from middle to exact in space domain.

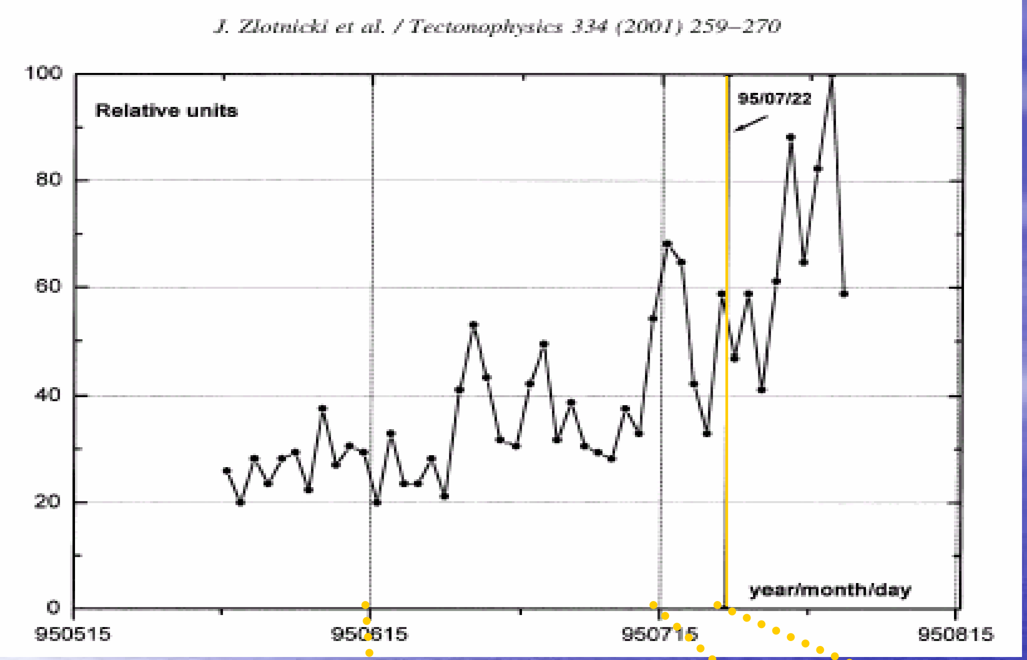
In some cases, the accuracy could be improved by making use of additional short-term monitoring of seismic activity and, perhaps, other geophysical fields in the alerted area of investigation.

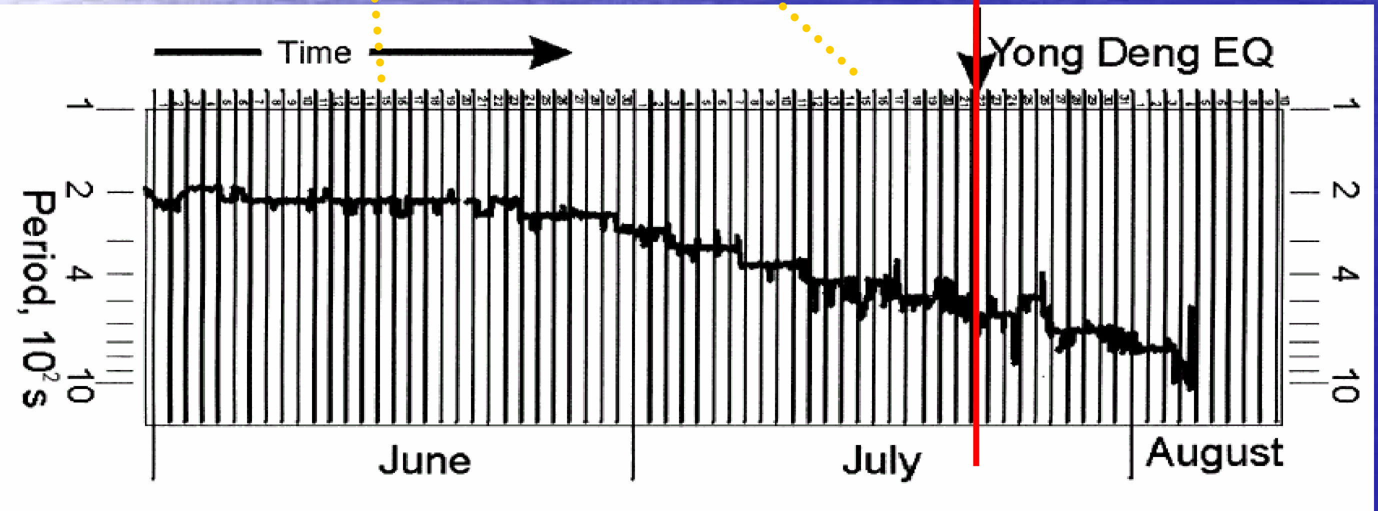
One case-study of electromagnetic record about the site of 21 July 1995, M5.7 Yong Deng, China, earthquake in Tibet



FTAN diagram of the resistance observed on NS 250-m line



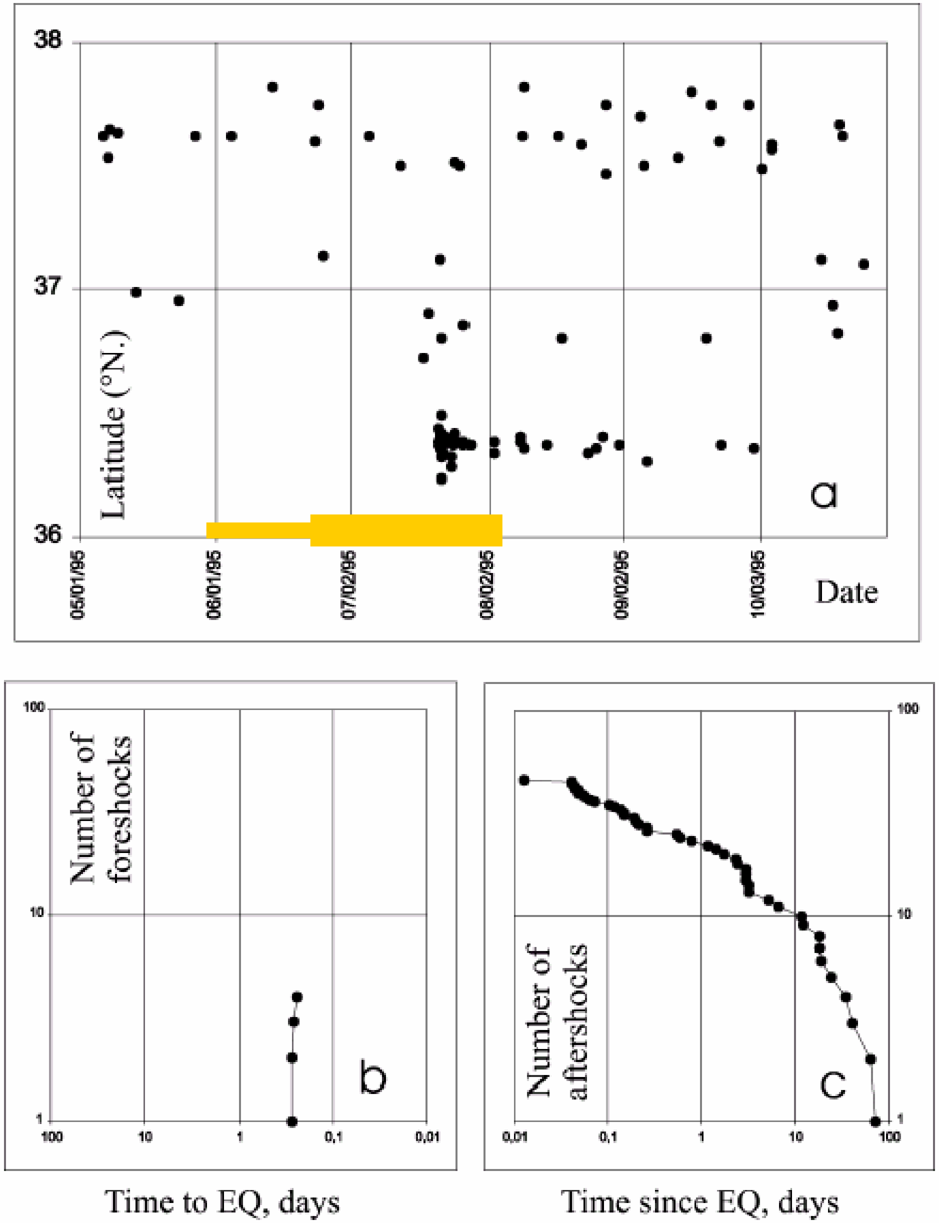




Evolution of the ULF signal

- Intensity and
- Period

GROW



Evolution of the ULF signal

The 1995 Yong Deng earthquake occurred in less than 100 km from the instrument at the time of characteristic ULF and/or its power decay on component directed at the epicenter.

The appearance of the ULF signal accompanied with a rise of seismic activity on adjusting segment of Haiyuan fault system.

The characteristic ULF collapsed just before aftershocks fast disappeared (exponentially).

Thank you