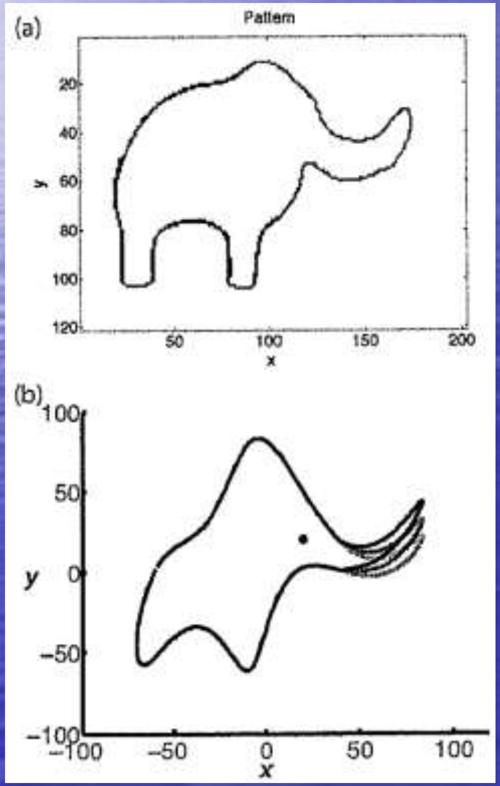
EARTHQUAKE HAZARD PREDICTION IN A BIG DATA WORLD

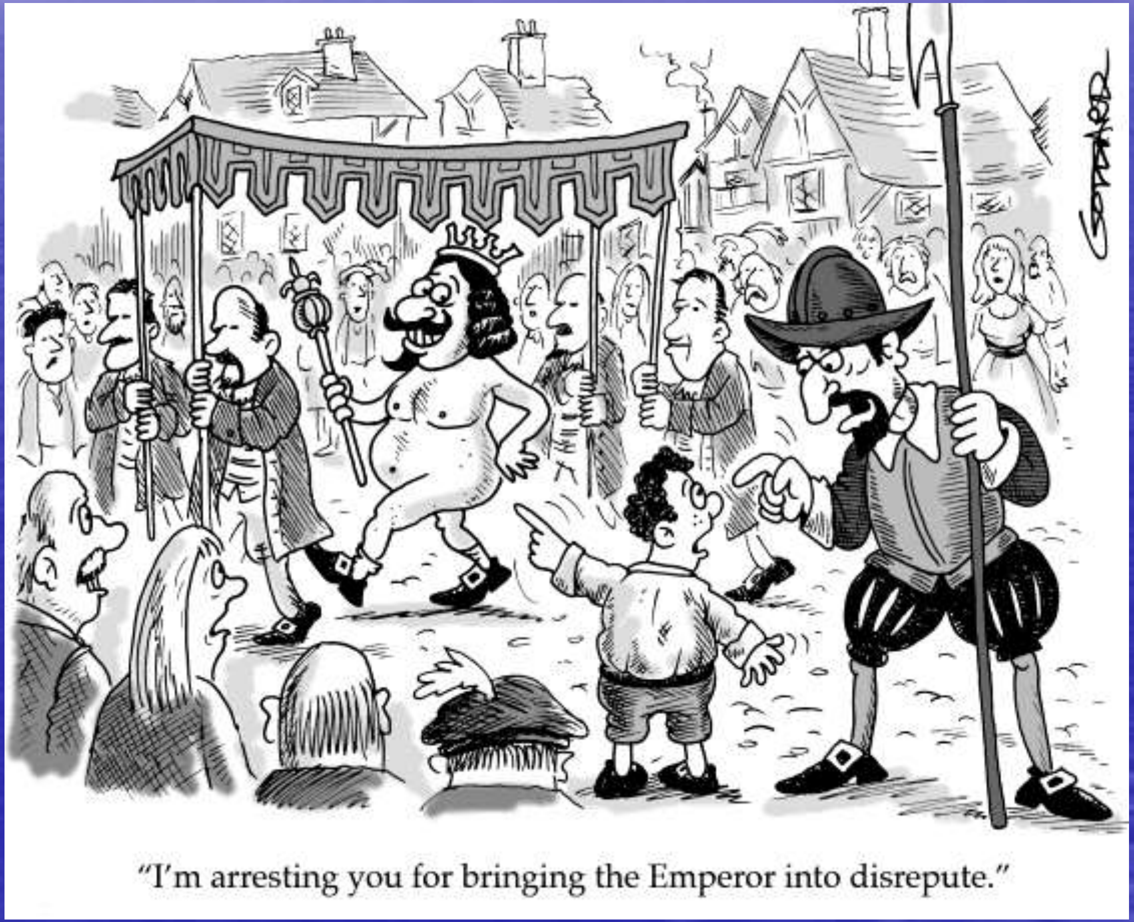
VLADIMIR G. KOSSOBOKOV

Institute of Earthquake Prediction Theory & Mathematical Geophysics,
Russian Academy of Sciences, Moscow, Russian Federation
Institut de Physique du Globe de Paris, Paris, France
International Seismic Safety Organization (ISSO)
E-mail: volodya@mitp.ru

The digital revolution started just about 15 years ago has already surpassed the global information storage capacity of more than 5000 Exabytes (in optimally compressed bytes) per year. Open data in a Big Data World provides unprecedented opportunities for enhancing studies of the Earth System. However, it also opens wide avenues for finding deceptive associations in inter- and transdisciplinary data and for inflicted misleading predictions.

Mayer et al. (2010) Drawing an elephant with four complex parameters. *Am. J. Phys*. 78(6): 648-649; doi: 10.1119/1.3254017





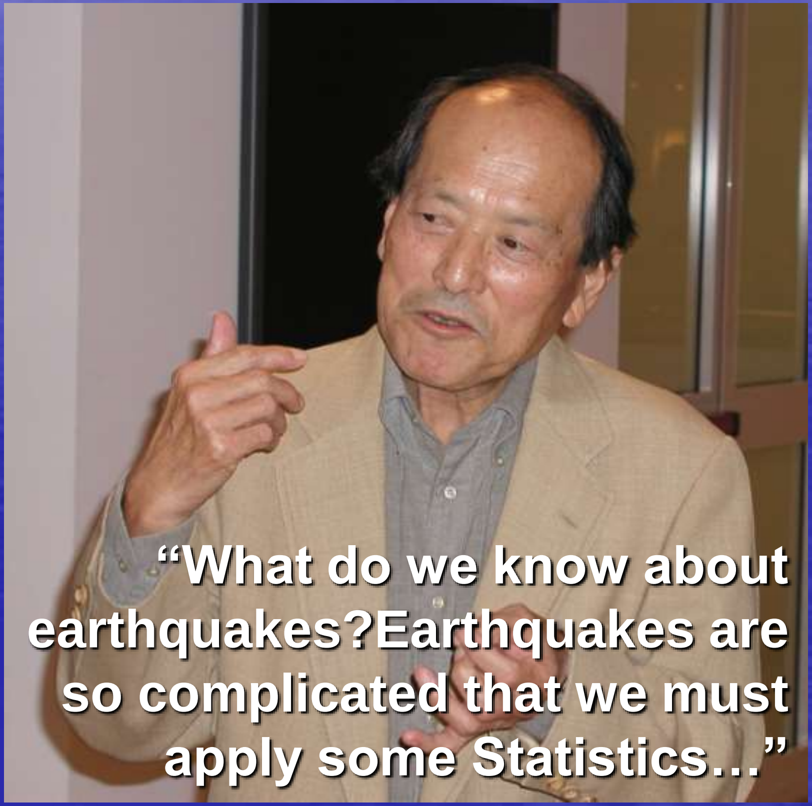
‘ SCIENCE SHOULD be able to warn people of looming disaster, Vladimir Keilis-Borok believes.
“My main trouble,” he says, “is feeling of responsibility.” ‘

(Los Angeles Times, 9 July 2012)

Vladimir Isaacovich Keilis-Borok (1921-2013)

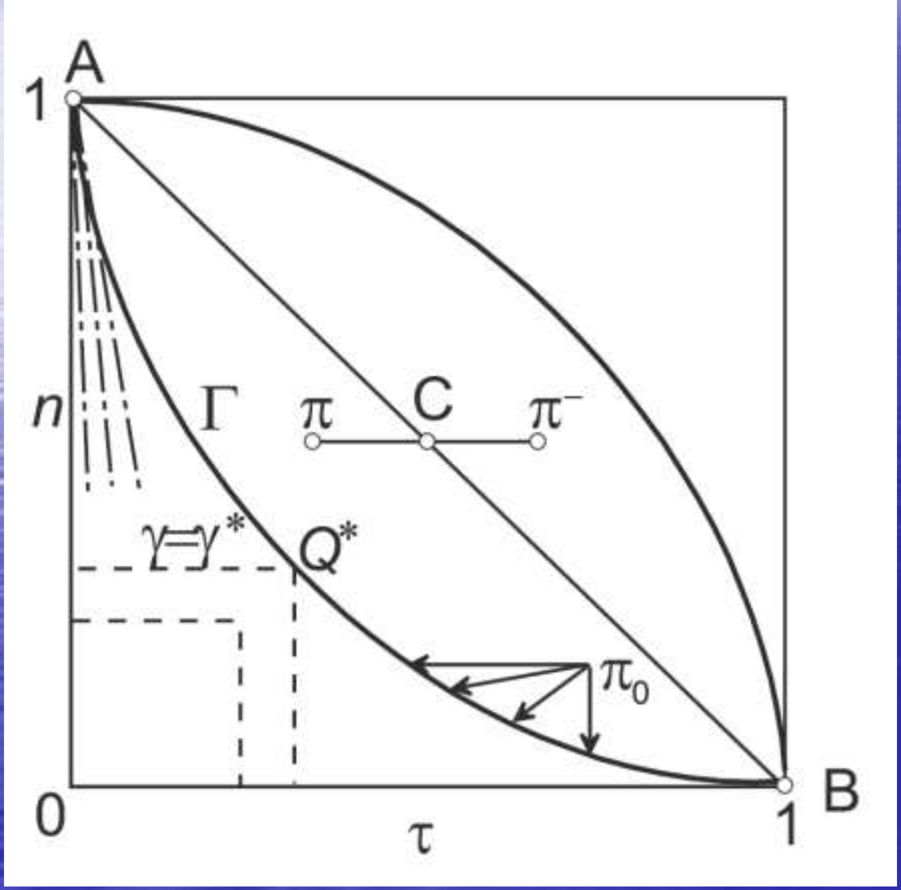
Earthquake prediction is not an easy task that implies a delicate application of statistics.

So far, none of the proposed short-term precursory signals showed sufficient evidence to be used as a reliable precursor of catastrophic earthquakes. Regretfully, in many cases of seismic hazard assessment (SHA), from term-less to time-dependent (probabilistic PSHA or deterministic DSHA), and short-term earthquake forecasting (StEF), the claims of a high potential of the method are based on a flawed application of statistics and, therefore, are hardly suitable for communication to decision makers.



Keiiti **Aki** (1930-2005)

Self-testing must be done in advance claiming prediction of hazardous areas and/or times. The necessity and possibility of applying simple tools of Earthquake Prediction Strategies, in particular, Error Diagram and Seismic Roulette null-hypothesis as a metric of the alerted space, is evident.



ERROR DIAGRAM

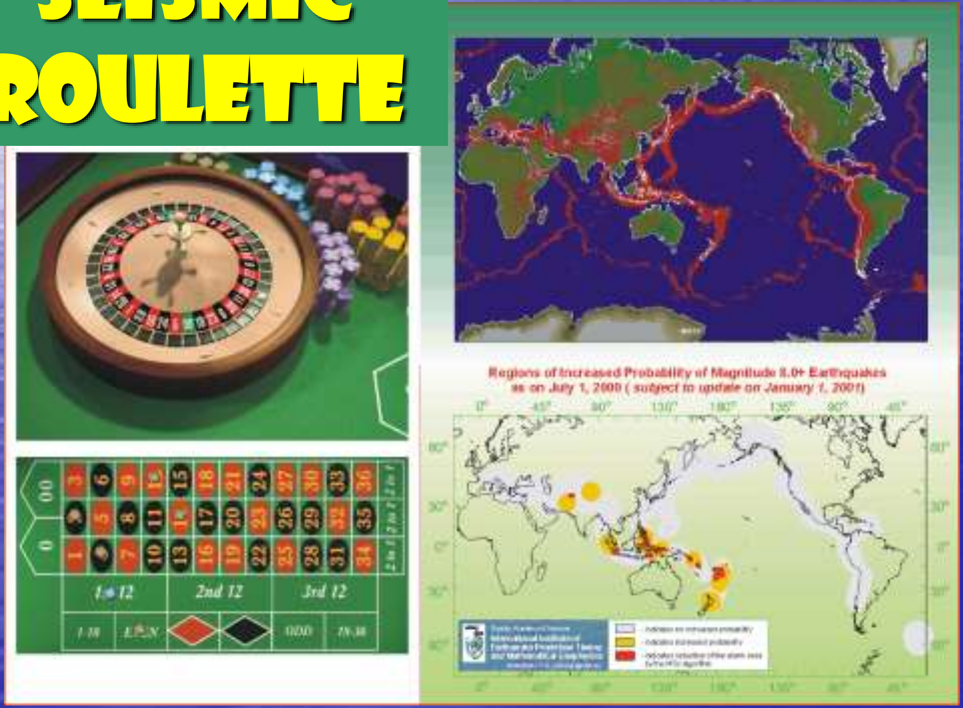
Molchan, G.M. Earthquake Prediction as Decision-making Problem. *Pure Appl. Geoph*, **149**, 233-247, 1997.

Molchan, G.M. Chapter 5. Earthquake Prediction Strategies: a theoretical analysis. In: Keilis-Borok, V.I., and A.A. Soloviev, (Editors). *Nonlinear Dynamics of the Lithosphere and Earthquake Prediction*. Springer, Heidelberg, 208-237, 2003.

Self-testing must be done in advance claiming prediction of hazardous areas and/or times. The necessity and possibility of applying simple tools of Earthquake Prediction Strategies, in particular, Error Diagram and Seismic Roulette null-hypothesis as a metric of the alerted space, is evident.

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

SEISMIC ROULETTE



- 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! ☺

or lose … ☹

If you are smart enough to know "antipodal strategy" (Molchan, 1994; 2003), make the predictions efficient -- ----

and your wins will outscore the losses! ☺

☺ ☹ ☺ ☺ ☺ ☹ ☺ ☺ ☺

Note that statistics can never prove things, but disprove them.

The set of errors, i.e. the rates of failure and of the alerted space-time volume, can be easily compared to random guessing, which comparison permits evaluating the SHA method effectiveness and determining the optimal choice of parameters in regard to a given cost-benefit function. These and other information obtained in such a simple testing may supply us with a realistic estimates of confidence and accuracy of SHA predictions and, if reliable but not necessarily perfect, with related recommendations on the level of risks for decision making in regard to engineering design, insurance, and emergency management.

The examples of independent expertize of “seismic hazard maps”, “precursors”, and “forecast/prediction methods” are provided.

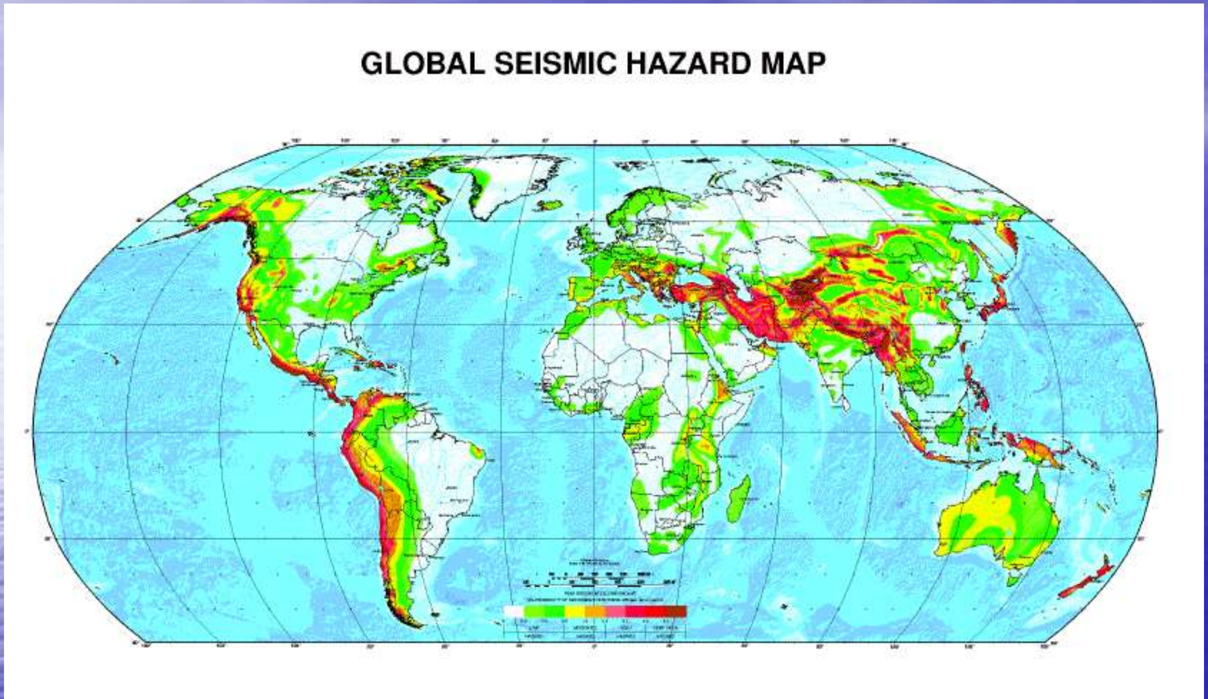
Note that earthquake related observations are limited to the recent most decades or centuries in just a few rare cases.

Getting, experimentally, reasonable confidence limits on an objective estimate of recurrence rate of an earthquake requires a geologic span of time which is unreachable for instrumental, or even historical, seismology (see, e.g., *Beauval et al., 2008***). That is why**

PROBABILITY ESTIMATES BY PROBABILISTIC SEISMIC HAZARD ANALYSIS REMAIN SUBJECTIVE VALUES RANGING FROM 0 TO 1, derived from analytically tractable hypothetical models of seismicity.

Making SHA claims, either termless or time dependent (t-DASH), quantitatively probabilistic in the frames of the most popular objectivists' viewpoint on probability requires a long series of "yes/no" trials, which cannot be obtained without an extended rigorous testing of the method predictions against real observations.

"One is well advised, when traveling to a new territory, to take a good map and then to check the map with the actual territory during the journey" [Wasserburg, 2010].



The Global Seismic Hazard Assessment Program (GSHAP) was launched in 1992 by the International Lithosphere Program (ILP) with the support of the International Council of Scientific Unions (ICSU), and endorsed as a demonstration program in the framework of the United Nations International Decade for Natural Disaster Reduction (UN/IDNDR). The GSHAP project terminated in 1999 .

A systematic comparison of the GSHAP peak ground acceleration estimates with those related to actual strong earthquakes discloses gross inadequacy of this "probabilistic" product, which appears **unacceptable for any kind of responsible seismic risk evaluation and knowledgeable disaster prevention.**

- Kossobokov, V.G., 2010. Scaling Laws and Earthquake Predictability in Assessment of Seismic Risk. Advanced Conference on Seismic Risk Mitigation and Sustainable Development. The Abdus Salam International Centre for Theoretical Physics (Trieste - Italy, 10 - 14 May 2010). http://cdsagenda5.ictp.trieste.it/full_display.php?ida=a09145)
- Kossobokov, V. G. ; A. K. Nekrasova, 2010. Global Seismic Hazard Assessment Program Maps Are Misleading. Eos Trans. AGU, 91(52), Fall Meet. Suppl., Abstract U13A-0020.
- Kossobokov, V., Nekrasova, A., 2011. Global Seismic Hazard Assessment Program (GSHAP) Maps Are Misleading. *Problems of Engineering Seismology*, 38 (1), p. 65-76 (in Russian).

Each of 1181 strong crustal earthquakes in 2000-2009 has from 6 to 58 values of GSHAP PGA in the ¼° × (¼ cos ϕ)° cell centered at its epicenter (ϕ, λ).

We count a "surprise" when the observed value, $I_0(M)$, is larger than the GSHAP expected maximum, $I_0(mPGA)$, $\Delta I_0 = I_0(M) - I_0(mPGA) > 0$

We found (i) about 50% of strong earthquakes surprised the GSHAP map (ii) each of the 59 magnitude 7.5 or larger earthquakes in 2000-2009 was a "surprise" for GSHAP Seismic Hazard Map; the minimum of the 59 values of ΔI_0 is 0.6. The average and the median of ΔI_0 are about 2.

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	V. Heavy

Kossobokov V., Peresan A., Panza G.F. (2015)
REALITY CHECK: SEISMIC HAZARD MODELS YOU CAN TRUST. EOS 96(13): 9-11

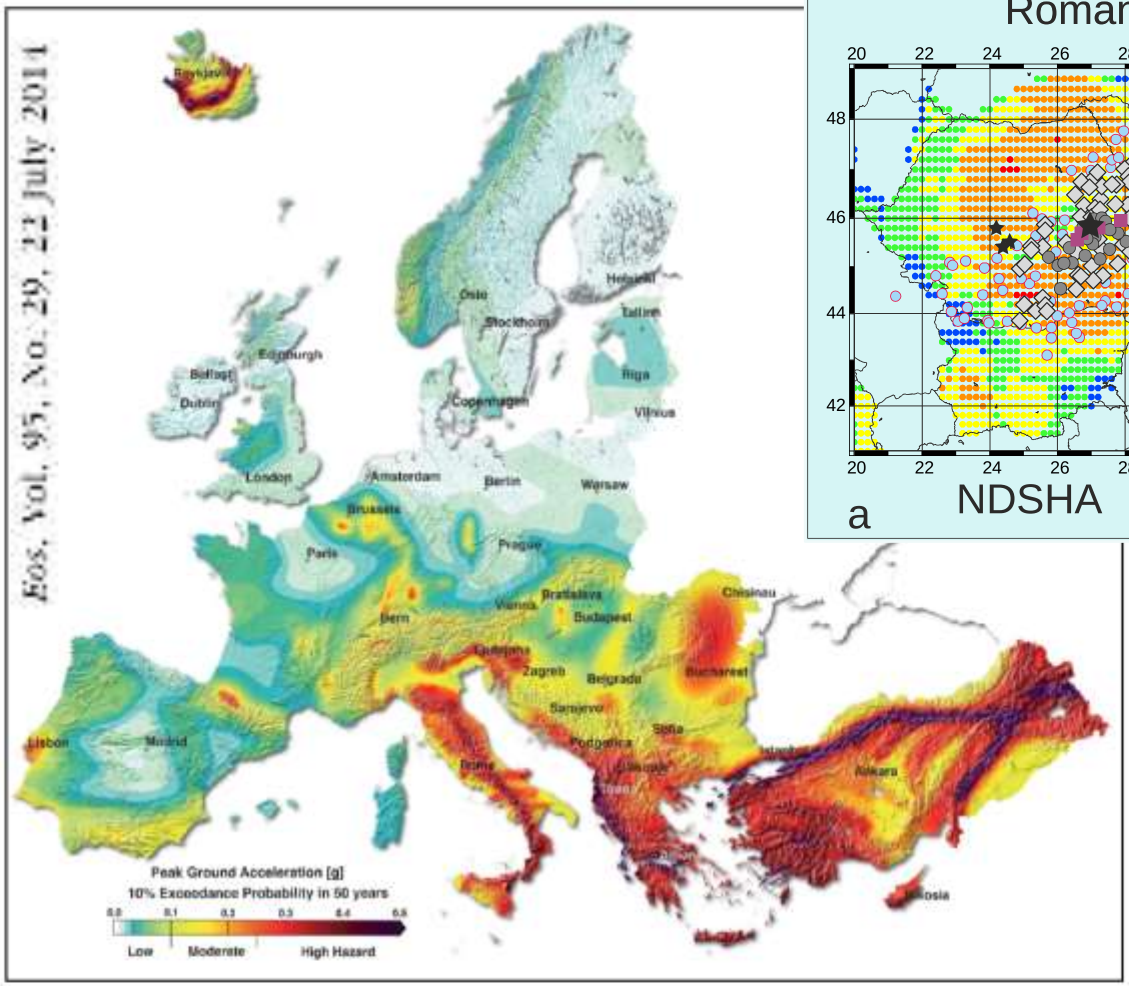
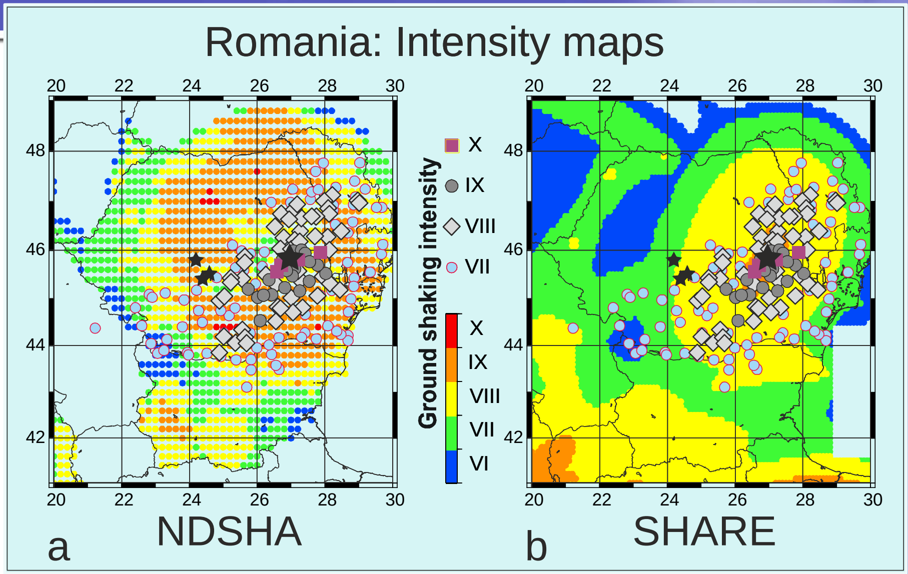
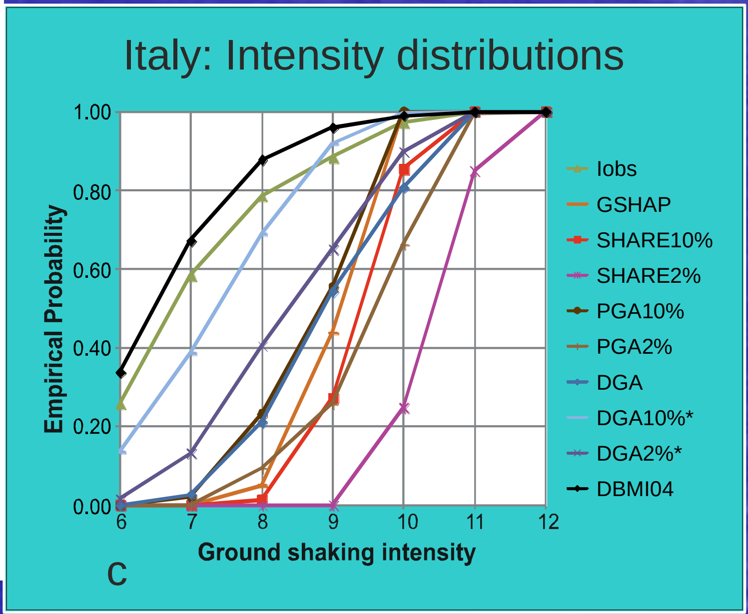


Fig. 1. European Seismic Hazard Map (ESHM13) displaying the 10% exceedance probability in 50 years for peak ground acceleration (PGA) in units of gravity (g). Cold colors indicate comparatively low hazard areas (PGA ≤ 0.1g), yellow and orange indicate moderate-hazard values (0.1g < PGA ≤ 0.25g), and red colors indicate high-hazard areas (PGA ≥ 0.25g).





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)

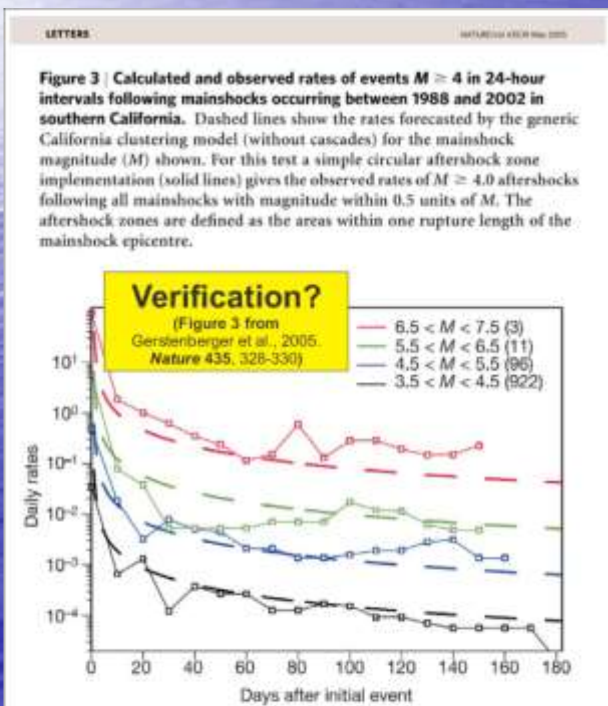

LETTERS

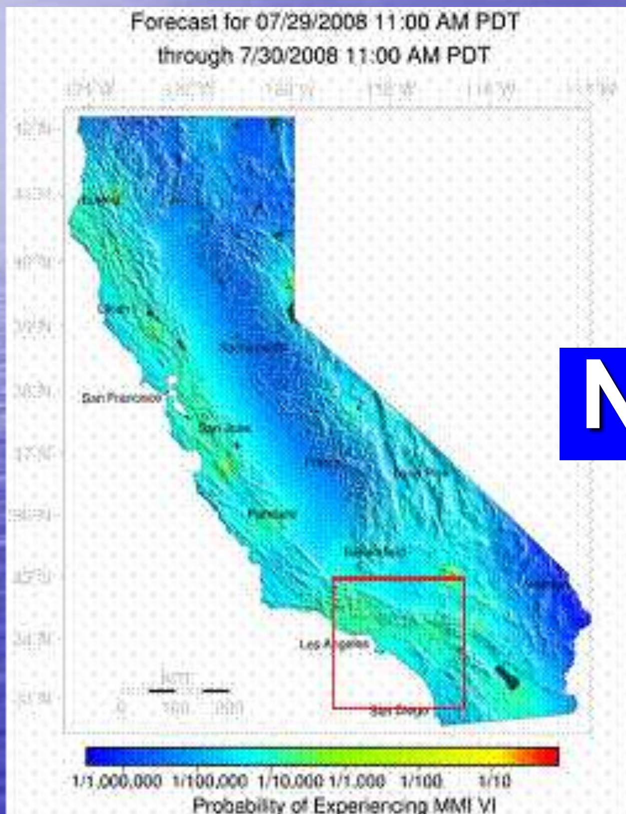
Figure 3 | Calculated and observed rates of events $M \geq 4$ in 24-hour intervals following mainshocks occurring between 1988 and 2002 in southern California. Dashed lines show the rates forecasted by the generic California clustering model (without cascades) for the mainshock magnitude (M) shown. For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

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.

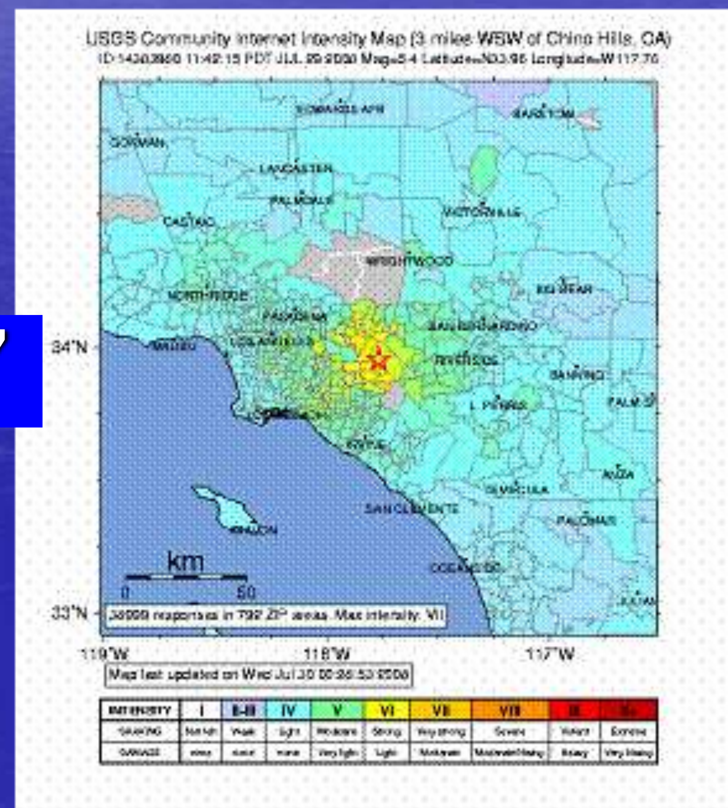
The probability of a smaller value of the Kolmogoroff-Smirnoff statistic D than that for the two samples used to plot the daily rates after 5.5 < M < 6.5 (green plot in Figure 3) event and after 3.5 < M < 4.5 (black plot) event (which D accounts to the value
$D = \max | F_{green}(t) - F_{red}(t) | \cdot (N_1 N_2/(N_1+N_2))^{1/2} \geq 2.12$)
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. ■

In 1200 days since publication all the seven earthquakes of MMI = VI or larger in California occurred in the areas of the lowest risk (p<1/10000), while the extent of the observed areas of intensity VI or larger is by far less than the one expected from the calculations (a crude low bound estimate of the ratio was above a factor of 8.5)...

№ 7 (29 Jul 2008, M5.4 WSW of Chino Hills)



№ 7



NATURAL ACCURACY IN EARTHQUAKE FORECASTING

Prediction of time and location of an earthquake of a certain magnitude range can be classified as follows -

- Term-less prediction of areas prone to earthquakes of certain magnitude
- 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

- The Gutenberg-Richter law suggests limiting magnitude range of prediction to about one unit of magnitude.

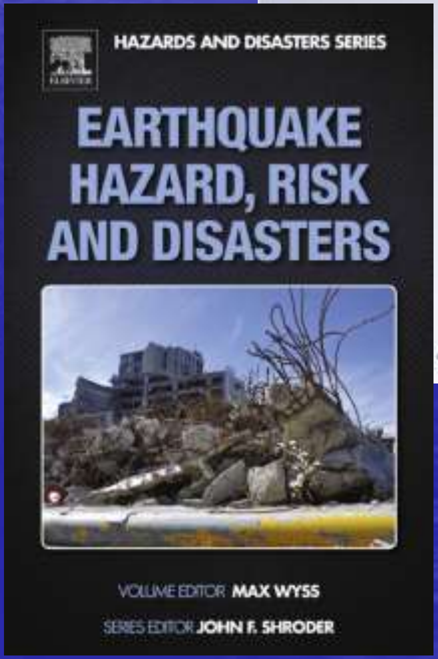
Otherwise, the statistics would be essentially related to dominating smallest earthquakes.

M8-MSc Algorithms

Healy, J. H., V. G. Kossobokov, and J. W. Dewey (1992) *A test to evaluate the earthquake prediction algorithm, M8,* U.S. Geol. Surv. Open-File Report 92-401, 23 p. with 6 Appendices

The results of truly global 25-year old experiment are indirect confirmations of the existing common features of both the predictability and the diverse behavior of the Earth's naturally fractal lithosphere.

The statistics achieved to date prove (with confidence above 99%) rather high efficiency of the M8 and M8-MSc predictions limited to intermediate-term middle- and narrow-range accuracy.



Kossobokov V (2014) Chapter 18. Times of Increased probabilities for occurrence of catastrophic earthquakes: 25 years of hypothesis testing in real time. In: Wyss M, Shroder J (eds) *Earthquake Hazard, Risk, and Disasters*. Elsevier, London, 477-504.

Kossobokov VG (2013) Earthquake prediction: 20 years of global experiment. *Natural Hazards* **69**(2):1155–1177; doi: 10.1007/s11069-012-0198-1

TABLE 18.2 Worldwide Performance of Earthquake Prediction Algorithms M8 and M8-MSc

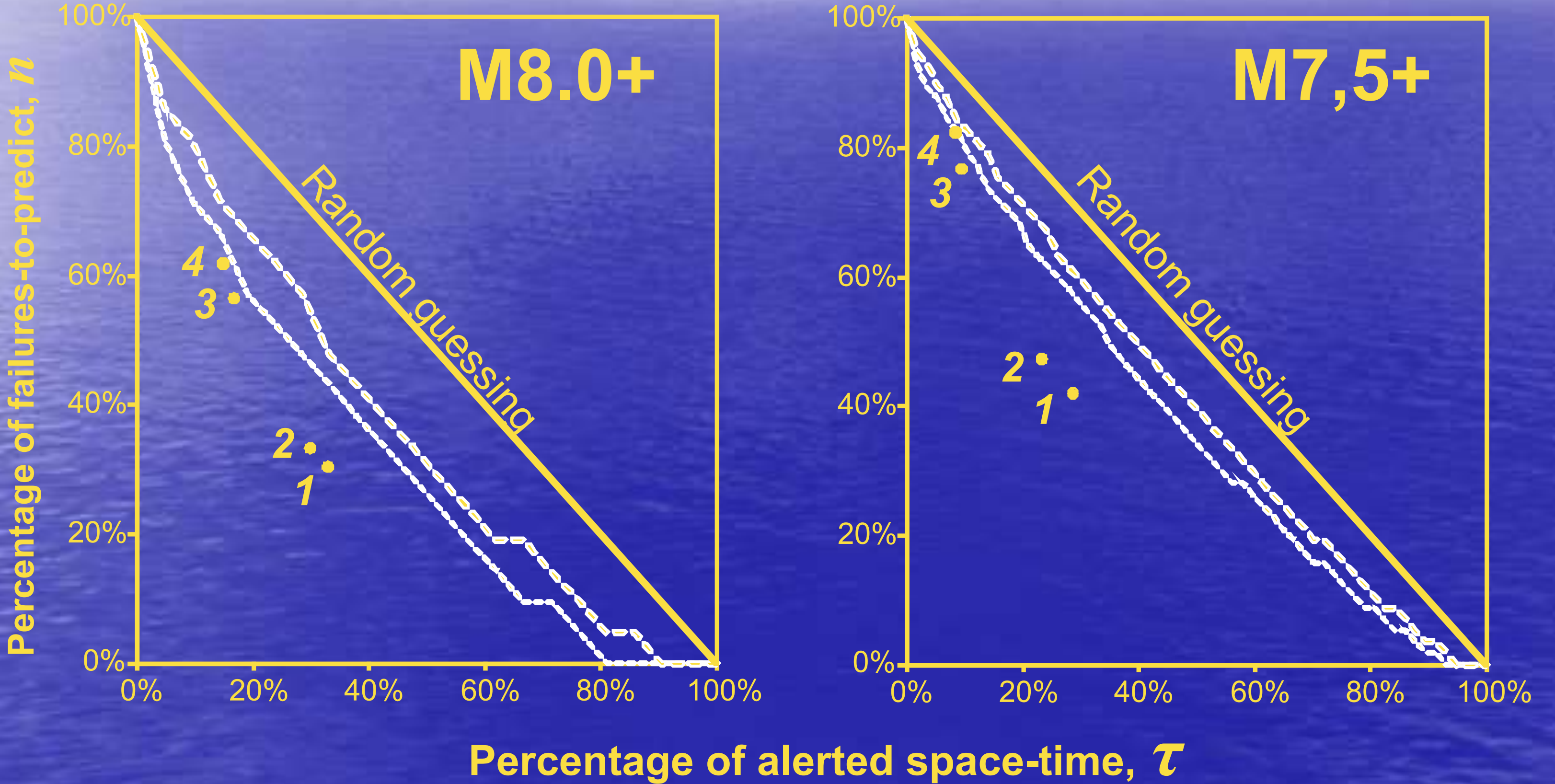
	Large Earthquakes						
		Predicted by		Measure of Alarms, %		Confidence Level, %	
Test Period	Total	M8	M8-MSc	M8	M8-MSc	M8	M8-MSc
Magnitude 8.0+							
1985–present	21	16	10	32.84	16.62	99.99	99.90
1992–present	19	14	8	29.80	14.78	99.99	99.63
Magnitude 7.5+							
1985–present	68	40	16	28.73	9.32	99.99	99.96
1992–present	56	30	10	23.14	8.31	99.99	98.36

Note: Confidence level tells how sure one can be that the achieved performance is not arisen by chance.

To drive any of the achieved confidence levels below 95%, the Test should encounter ten failures-to-predict in a row.

Error Diagrams for the results of the Global Test of the M8-MSc predictions of the great (M8.0+) and significant (M7.5+):

M8, 1985–2013 (1); 1992–2013 (2); M8–MSc, 1985–2013 (3), and 1992–2013 (4). The "random guessing" is outlined with the 95 and 99% confidence level curves (for 21 and 57 independent tests on the left and right).

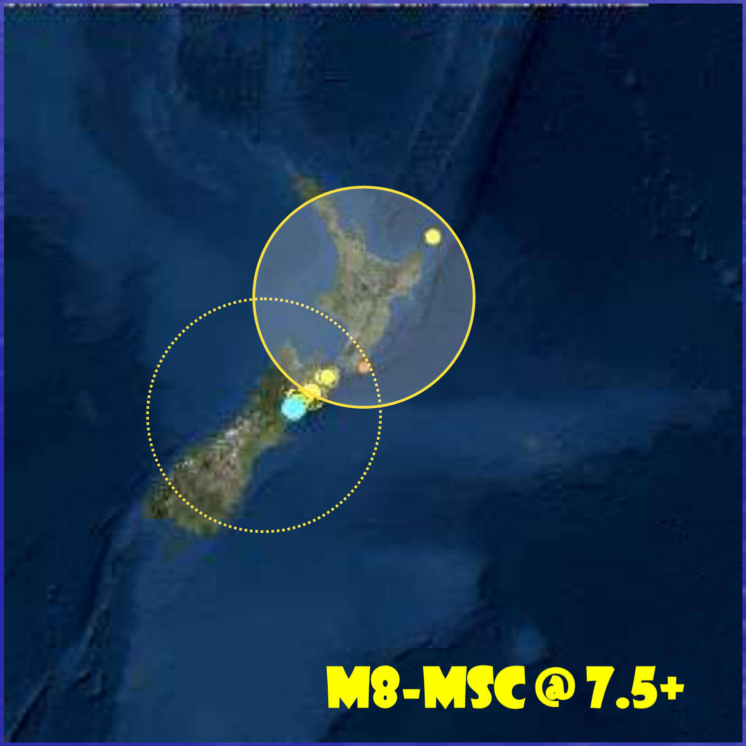


Kossobokov V, Soloviev A (2015). Evaluating the Results of Testing Algorithms for Prediction of Earthquakes. *Doklady Earth Sciences, 2015, Vol. 460, Part 2, pp. 192–194*

The recent examples of the M8-MSc confirmed predictions for the second half of 2016 suggest intersections of morphostructural lineaments as centers of CI in Italy and expansion of the Global Test to the areas where seismic catalog data was insufficient in 1992 but is enough complete nowadays (like in New Zeeland).

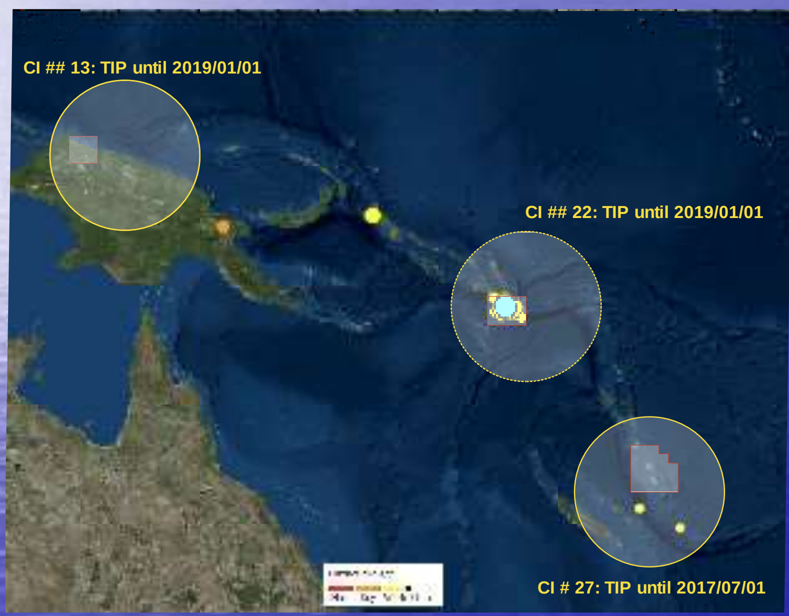
Amatrice, 24th August 2016 (M6.2), Visso, 26th October 2016 (M6.1), and Norcia, 30th October 2016 (M6.6)

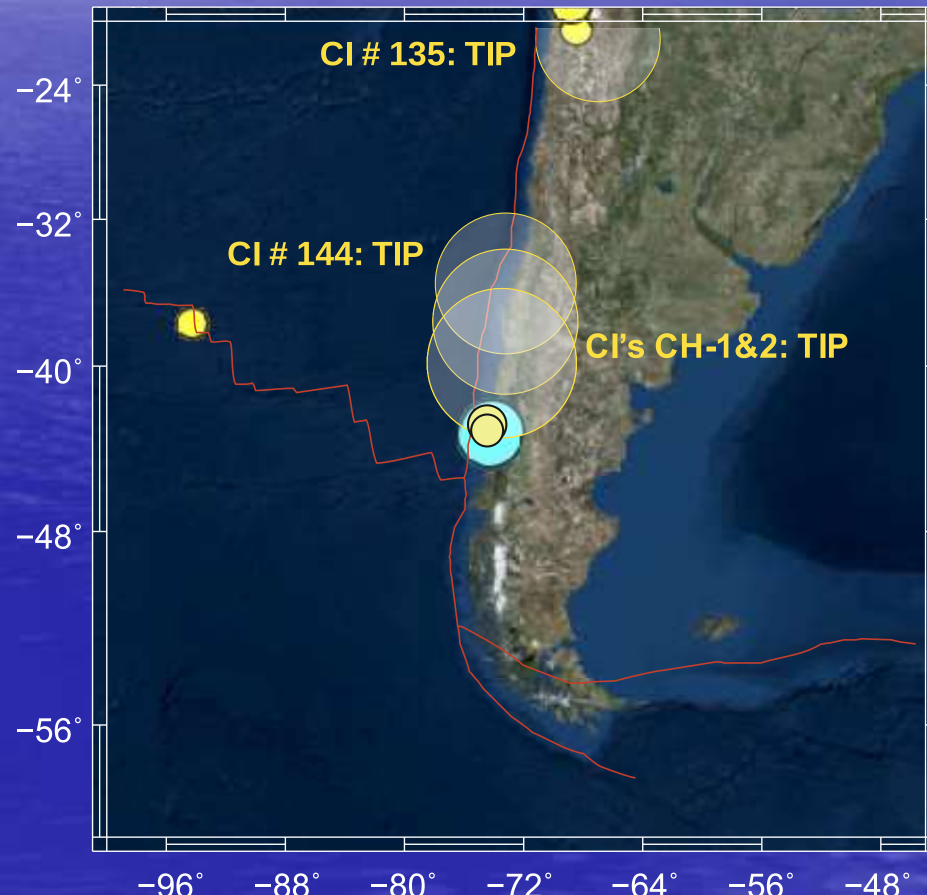
53 km NNE of Amberley, New Zeeland, 13th November 2016 (M7.8)



- 8 December 2016, M7.8, 69km WSW of Kirakira, Solomon Islands

25 December 2016, M7.6, 42 km SW of Puerto Quellon, Chile





M8-MSC @ 7.5+

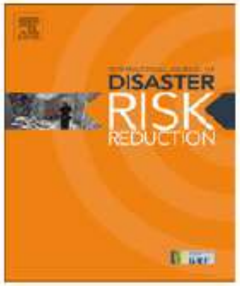
The earthquake detection could have been utilized to implement measures and improve earthquake preparedness in advance; unfortunately this was not done, in part due to the predictions' limited distribution and the lack of applying existing methods for using intermediate-term predictions to make decisions for taking action.

ELSEVIER

Contents lists available at SciVerse ScienceDirect

International Journal of Disaster Risk Reduction

journal homepage: www.elsevier.com/locate/ijdrr



Advance prediction of the March 11, 2011 Great East Japan Earthquake: A missed opportunity for disaster preparedness

C. Davis[a,*], V. Keilis-Borok[b,c], V. Kossobokov[c,d], A. Soloviev[c]

[a] Geotechnical Engineering Group, Los Angeles Department of Water and Power, 111 North Hope Street, Room 1368, Los Angeles, CA 90012, USA
[b] Department of Earth and Planetary Sciences and Institute of Geophysics and Planetary Physics, University of California, Los Angeles, CA 90095-1567, USA
[c] Institute of Earthquake Prediction Theory and Mathematical Geophysics, Russian Academy of Sciences, 84/32 Profsouznaya, Moscow 117997, Russia
[d] Institut de Physique du Globe de Paris, 1 Rue Jussieu, 75238 Paris Cedex 05, France

Possible actions in response to an intermediate-term prediction

Possible scenario actions to reduce seismic risks in response to alarm. Actions are described in the text.

Item	Action	C_a ($1000)	P ($1000)	Gain, G ($1000)	
				$1-f=20\%$	$1-f=50\%$
Nuclear power plants					
1a	Raise tsunami wall	10,000	500,000	90,000	240,000
1b	Protective generator housing	1000	500,000	99,000	249,000
1c	Raise wall + protective housing	11,000	500,000	89,000	239,000
Home, office, maintenance, industrial buildings					
2	Anchor furniture, cabinets, computers, equipment, etc.	11	101	9.2	39.5
3a	Relocating out of tsunami inundation area, or	500	1,500	−200	250
3b	Retrofitting structure for tsunami	280	1,500	20	470
Lifeline systems					
4	Railway bridge and track	400	700	−260	−50
5	Water pipe replacement	420	2100	0	630
6	Highway tunnel landslide repair	2000	2000	−1600	−1000
7	Power transformers	1500	10,000	500	3500
8	Roadway bridge	500	1500	−200	250
9	Liquid Fuel Tank	30	2000	370	970
Cultural					
10	Nikko temples	50	100	−30	0
Disregard/Unaware of Prediction					
11	Do nothing	0	−518,401	−103,680	−259,200

Using equation - $G = P(1-f) - Ca$ - to estimate f at the breakeven point when G=0 identifies that it was cost effective to take action for the Fukushima nuclear power plant with a 99.99% probability of false alarm.

CONCLUSION

The confirmed reliability of pattern recognition results, along with realistic and exhaustive scenario modeling and testing against Reality, allow concluding –

Science can disclose Natural Hazards, assess Risks, and deliver the state-of-the-art knowledge of looming disaster in advance catastrophes along with useful recommendations on the level of risks for decision making in regard to engineering design, insurance, and emergency management.

Policy may wish to stop wearing the exposed “emperor’s new clothes” that do not protect from Natural Hazards and avoid buying such in the Future.

Thank you!

“When sorrows come, they come not single spies, but in battalions”
(William Shakespeare, 1564-1616)