Bollettino di Geofisica Teorica ed Applicata Vol. 56, n. 2, pp. 351-356; June 2015
DOI 10.4430/bgta0142

Purported precursors: poor predictors

F. Mulargia[1], P. Gasperini[1], B. Lolli[2] and P.B. Stark[3]

[1] *Dipartimento di Fisica e Astronomia, Università di Bologna, Italy*
[2] *Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna, Italy*
[3] *Department of Statistics, University of California, Berkeley, CA, U.S.A.*

(Received: April 30, 2014; accepted: December 22, 2014)

ABSTRACT The destructive 2009 L'Aquila and 2012 Emilia-Romagna earthquakes led the Italian Dipartimento della Protezione Civile (DPC) to fund nine research groups to investigate seismic precursors. Three research groups produced testable predictions by the DPC deadline of May 31, 2013, based on: 1) radon in a well in Friuli, 2) temperature, flow, CO_2 flux, and other variables measured in wells in Emilia-Romagna, and 3) an Artificial Neural Network (ANN) algorithm applied to seismicity. We evaluated the geochemical precursors by comparing their predictions to an equal number of predictions at the same locations and with the same individual and total durations as the actual predictions, but at random times. This approach avoids modelling the seismicity, so the accuracy of the predictions is not influenced by the accuracy of any seismicity model. Neither set of geochemical precursors succeeds significantly better than the random predictions. ANN, on the other hand, did not predict any events large enough to affect public safety.

Key words: earthquake prediction, validation, geochemical precursors.

1. Introduction

Predicting earthquakes is the holy grail of seismology: success would save lives, money, and cultural heritage. Claims of precursory phenomena are countless, but there is no quantitative physical theory of seismogenesis. No proposed precursor has been demonstrated to signal impending failure in laboratory experiments (Geller *et al.*, 1997; Mulargia and Geller, 2003) and none has been demonstrated to be reliable in practice. Earthquakes cluster in space and time; that clustering can be used to predict earthquakes with some level of accuracy—but perhaps not a useful level (Luen and Stark, 2007).

After the deadly 2009 L'Aquila and 2012 Emilia-Romagna earthquakes, the DPC began a number of research activities to determine whether purported precursors could provide useful public warning before destructive earthquakes. DPC funded nine research groups, giving a deadline of May 31, 2013 for reporting testable results. Three groups met the deadline: two used chemical and other physical data measured at wells; the third applied an ANN to seismicity. The DPC-funded groups that did not meet the deadline proposed geoelectrical and magnetotelluric signals, variations in the ratio of seismic P-wave and S-wave velocities, and crustal deformation measured by InSAR. No evaluable data were presented for crustal deformation measured by GPS recordings for Pollino. Four "anomalies" were presented for Emilia-Romagna, too few for a meaningful test (Report of the S3 Project, see Data and Resource section).

2. Methods and results

We evaluated the methods of the three teams that produced timely testable results for the Emilia-Romagna and Pollino regions, which are of particular interest to DPC. We tested the first two methods retrospectively using the seismic catalogue of Gasperini *et al*. (2013), a version of the Italian seismic catalogue with homogenized M_W magnitude, starting in 1998 or 2002.

Our statistical test compares the success rate to the probability distribution of the success rate of an equal number of random predictions. The random predictions are at the same locations as the actual predictions, but with starting times distributed independently and uniformly, conditioned to have the same durations as the actual predictions. This condition introduces dependence among the starting times – otherwise, alarms could overlap – so the joint distribution of starting times is not independent and uniform. Comparing the predictions to these random predictions, rather than modelling seismicity as a stochastic process, avoids the possibility of concluding erroneously that the predictions are reliable simply because the model of seismicity is inaccurate. Moreover, it addresses the fact that even rudimentary predictions that exploit clustering can look wildly successful when tested against random catalogs (Stark, 1997; Luen and Stark, 2007).

We estimated the success rate of random predictions by simulation. Let [0, T] be the test interval; T the total duration of the test; T_A the total duration of all the alarms; and A the total number of alarms. The following algorithm generates alarms with independent, uniformly distributed starting times, conditioned to have the same individual durations as the actual alarms:

1. Permute the A alarm durations into random order. Let a_i denote the ith permuted duration. Set $a_0 = 0$.

2. Generate A+1 independent, identically distributed uniform random variables on the interval $[0, T - T_A]$. Let t_i denote the ith smallest of those A+1 variables, so $0 \le t_1 \le t_2 \le \ldots \le t_{A+1} \le T - T_A$.

3. For $1 \le i \le A$, alarm i runs from time $t_i + \sum_{j=1}^{i-1} a_j$ to time $t_i + \sum_{j=1}^{i} a_j$.

The first purported precursor is radon measured at a well located in Cazzaso, north-eastern Italy (lat. = 46.430669° N; lon. = 12.995248° E). This well is not within Emilia-Romagna or Pollino. The data consist of 17,783 measurements of natural radioactivity (in Bq/m^3), made nominally every three hours between November 12, 2002 and March 27, 2013, but with gaps. Measurements cover about 59% of the interval. We consider a measurement to be anomalous if it is two or more standard deviations above 495.3 Bq/m^3, the mean of the measurements for the interval 2002–2013. There were 807 anomalies, according to this definition. The proponent quoted the threshold for an anomaly to be 496 Bq/m^3 between 2002 and 2009 and 148.7286

Bq/m^3 between 2010 and 2013, but gave no justification for this choice. We inferred that the change resulted from *a posteriori* selection designed to maximize the predictive accuracy retrospectively, and hence we used a threshold calculated from the data—although because this threshold is calculated from all the data, it is known only retrospectively. The spatio-temporal alarm windows corresponding to anomalies was not specified by the proponents, and there is no obvious way to tie the anomalies to a particular seismogenic region. To use the anomalies to make predictions, we consider each anomaly to be the start of an alarm of length 7 days, 15 days, 30 days, or 100 days. When alarms overlapped, we considered an alarm to be the union of all overlapping alarms (a single extended alarm), resulting in 68 disjoint alarms, listed in Table S1 in the electronic supplement. Radon anomalies were considered to be successful predictions if they were followed by an event of magnitude 3.0 or greater, within all combinations of the four temporal windows and within 15 km, 30 km, 50 km, or 100 km of the well. These 16 spatio-temporal windows span a range of reasonable options.

The performance of these random predictions is summarized in Table 1, along with the "backward" success rate for the radon anomalies. For a 7-day window, random predictions outperform the actual predictions 22%–91% of the time. For a 15-day window, random predictions were more successful 12%–30% of the time. For a 30-day window, random predictions were more successful 17%–93% of the time. For a 100-day window, random predictions were better 8%–41% of the time. Compared with random predictions, the actual predictions performed best for a 100-day temporal window and a 30-km radius spatial window. Given that 16 spatio-temporal windows were tested, statistical multiplicity is a serious issue: the 8% P-value should not be considered statistically significant.

Table 1 - Predictive performance of the radon anomalies in forward and reverse time, compared to the performance of random predictions.

Temporal Window	Spatial window radius	alarms	Total alarm duration	Forward success rate	**Forward P** (random predictions succeed at least as often as claimed precursors)	Reverse success rate	**Reverse P** (random predictions succeed at least as often as claimed precursors)
7d	15 km	56	17%	40%	0.22	0%	1.0
	30 km	56	17%	18%	0.41	0%	1.0
	50 km	56	17%	13%	0.65	13%	0.0
	100 km	56	17%	8%	0.91	10%	0.0
15d	15 km	41	27%	40%	0.30	20%	0.0
	30 km	41	27%	36%	0.16	27%	0.0
	50 km	41	27%	42%	0.12	23%	0.0
	100 km	41	27%	30%	0.28	16%	0.0
30d	15 km	27	39%	40%	0.93	20%	0.58
	30 km	27	39%	36%	0.87	36%	0.29
	50 km	27	39%	61%	0.17	36%	0.36
	100 km	27	39%	49%	0.19	35%	0.38
100d	15 km	7	62%	60%	0.41	20%	0.75
	30 km	7	62%	73%	0.08	58%	0.23
	50 km	7	62%	81%	0.10	67%	0.29
	100 km	7	62%	62%	0.34	64%	0.30

There are a surprising number of combinations for which the radon anomalies perform better than random predictions, both in forward and in reverse time; the association in reverse time is marginally higher. This is likely because earthquakes, and the strain readjustment following earthquakes, crush rock. Crushing carbonatic rocks, typical of north-eastern Italy where the Cazzaso well is located, releases interstitial radon. The radon anomalies are presumably a coseismic/postseismic effect. In a temporal cluster of earthquakes, the anomalies appear to be both a precursor of the next event and a consequence of the previous one. In summary, there is no evidence that the radon anomalies have any value as precursors.

The second purported precursor consists of physical and geochemical measurements from wells in Emilia-Romagna in 1998–2000 and 2004–2012. There were 129 anomalies, defined as measurements of temperature, flow, electrical conductivity, spontaneous potential, and CO_2 flow more than two or three standard deviations from the mean. The spatio-temporal windows of association are given in Table S2 in the electronic supplement; the proponents did not provide a quantitative basis for these windows. Anomalies in the Emilia-Romagna wells were considered to be successful predictions if they were followed by events within the spatio-temporal and magnitude windows specified by the proponent. A number of anomalies occurred within the previous windows. We concatenated overlapping windows, which reduced the number of predictions from 129 to 73, of which 20 successfully predicted an event. Random predictions had a success rate at least that high in 80% of 10,000 simulations.

The ANN predictions, developed by the private company Semeion, involved training an ANN on data provided by the Istituto Nazionale di Geofisica e Vulcanologia (INGV) starting in January 1, 1981: the CSI catalogue (Castello *et al.*, 2007), the BSI online bulletin, and the ISIDe online bulletin (see Data and Resource section). The ANN was trained on Italian seismicity aggregated into 2054 0.1 by 0.1 degree cells daily. The ANN forecasts next-day maximum magnitude in each daily cell from prior seismicity in that and surrounding cells.

We evaluated the ANN predictions prospectively over the 9-month period July 1, 2012 to March 31, 2013 against the seismic catalogue that was used to train the ANN. Following the proponents of the ANN method, we considered a prediction to be successful if an event occurred within a daily cell that was predicted to have an event with a magnitude within 0.5 of the predicted magnitude. We conducted tests using magnitude thresholds of M=2.0, 2.5, 3.0, 3.5, and 4.0. Because the magnitude window is ±0.5, there can be more successful predictions than actual earthquakes of a given magnitude. For example, a M1.5 earthquake could make a M2 prediction successful. The associations for magnitude $M \geq 3.0$ are given in Tables S3 and S4 in the electronic supplement. The prediction performance is summarized in Table 2. For each of the 5 magnitude thresholds, the sum of the rate of false alarms and the rate of unpredicted events ranges from 94% (for magnitude 2.5 and up) to 200% (for magnitude 4.0 and higher). The ANN method is useless for events large enough to be of public concern (it did not successfully predict any event of magnitude 4.0 higher, but did give many false alarms).

Table 2 - Rates of earthquakes predicted by Semeion ANN and of predictions that succeeded.

Magnitude	predicted earthquakes	correct predictions
M ≥ 4.0	0 /25	0 /6
M ≥ 3.5	5/57	5/47
M ≥ 3.0	38/176	29/128
M ≥ 2.5	308/566	337/652
M ≥ 2.0	1385/1908	1513/2202

3. Data and resources

Report of the S3 Project, "Short-term earthquake prediction and preparation", Agreement DPC-INGV 2012-2021. Available at https://9e03c889-a-62cb3a1a-s-sites.googlegroups.com/site/ingvdpc2012progettos3/documents/Final%20Report%20S3.pdf?attachauth=ANoY7cq0jirDjS3lpQPak6s1WX1t1hFoxnIS8SON4D1JwiOzcDLu65F3O_RUa-4q2IW-f6mFTveS4eQSHinMwDEbFzKsLtLpWssopl3SIRyBN3ZU_TpfHX311WIpX4qWcn0XAo0QftnHCs1FJVpecc-EwYL0yoPUF_xvQH2CueIq-RSwyIO_cGQIrXpJViHSPrrOCN5zL6gMv1-_Gtlj5kPCbHesGJRr3xPpefCebHP3OuW7Q3kUgerxiBCb76vTbilKnO_60hzi&attredirects=1 (last accessed September 2014).

CSI catalogue, available at http://csi.rm.ingv.it/ (last accessed September 2014).

BSI online bulletin, available at http://bollettinosismico.rm.ingv.it/ (last accessed September 2014).

ISIDe online bulletin, available at http://iside.rm.ingv.it/iside/standard/index.jsp (last accessed September 2014).

Acknowledgments. Online material (lists of radon and other geochemical/geophysical anomalies and of M≥3 predictions and associated events for the Artificial Neural Network predictions) is available at the BGTA website (www2.inogs.it/bgta). FM, PG, and BL gratefully acknowledge funding provided by the Italian Presidenza del Consiglio dei Ministri – Dipartimento della Protezione Civile (DPC). Scientific papers funded by DPC do not represent its official opinion and policies.

REFERENCES

Castello B., Olivieri M. and Selvaggi G.; 2007: *Local and duration magnitude determination for the Italian earthquake catalog, 1981-2002*. Bull. Seismol. Soc. Am., **97**, 128-139.

Gasperini P., Lolli B. and Vannucci G.F.; 2013: *Empirical calibration of local magnitude data sets versus moment magnitude in Italy*. Bull. Seismol. Soc. Am., **103**, 2227-2246, doi: 10.1785/0120120356.

Geller R.J., Jackson D.D., Kagan Y.Y. and Mulargia F.; 1997: *Cannot earthquakes be predicted?* Sci., **278**, 488-490.

Luen B. and Stark P.B.; 2007: *Testing earthquake predictions*. In: Essays in Honor of David A. Freedman, IMS Lecture Notes Monograph Series, Probability and Statistics, Beachwood, OH, USA, pp. 302-315.

Mulargia F. and Geller R.J. (eds); 2003: *Book review: Earthquake science and seismic risk reduction*. Survey Geophys., 37, 383-386, doi 10.1007/s10712-005-5417-9.

Stark P.B.; 1997: *Earthquake prediction: the null hypothesis*. Geophys. J. Int., **131**, 495-499, doi: 10.1111/j.1365-246x.1997.tb06593.x.

Corresponding author: Philip B. Stark
Department of Statistics, University of California, Berkeley, CA, 94720-3860, U.S.A.
Phone: +1 510 394 5077 fax: +1 510 642 7892; e-mail: stark@stat.berkeley.edu