Comment on "Assessing CN earthquake predictions in Italy" by M. Taroni, W. Marzocchi, P. Roselli

George Molchan[1], Antonella Peresan[2,3], Giuliano Francesco Panza[3,4,5], Leontina Romashkova[1], Vladimir Kossobokov[1,3]

[1] *Institute of Earthquake Prediction Theory and Mathematical Geophysics, Russian Academy of Sciences, Moscow, Russian Federation*

[2] *National Institute of Oceanography and Experimental Geophysics. CRS-OGS, Udine, Italy*

[3] *International Seismic Safety Organization, Arsita, Italy*

[4] *Institute of Geophysics, China Earthquake Administration, Beijing, People's Republic of China*

[5] *Accademia Nazionale dei Lincei, Roma. Italy*

Adequate assessment of prediction results is a fundamental step in earthquake prediction research and requires a correct application of appropriate statistical tools, respectful of their basic assumptions and data quantity/quality. For this reason we wish to draw the attention of the readers of *Annals of Geophysics* to some basic pitfalls of the paper by Taroni et al. [2016], which may become a source of misleading interpretations and follow-up erroneous conclusions. In a nutshell, Taroni et al. [2016] failed to assess correctly the results of the on-going CN prediction experiment in Italy started in 1998 [Peresan et al. 2005]. The main reasons for this are the following:

(1) the CN forecast is implemented in three overlapping regions of Italy, where the statistics of target events is small, and the analysis of the prediction statistical significance is forcedly applied to each region separately, where the statistic of target events is insufficient;

(2) there are methodological errors in Taroni et al. [2016], which are analyzed in detail in Molchan et al. [2017].

Taroni et al. [2016] declared intent to give "*a careful assessment of CN prediction performances … using standard testing procedures.*" This is an unlikely feasible goal when splitting the entire, yet small, sample of target earthquakes into smaller parts related to sub-regions of Italy used in the CN algorithm application: the number of target events within each of the three sub-regions is 5, 3, and 1.

A priori the *standard* statistical methods may not be effective in any of the CN sub-regions. Molchan et al. [2017] show that such a small number of binomial trials implies, with necessity, low resolution of testing, and may lead to erroneous interpretation of the entire statistics in total. Taroni et al. [2016] made their choice of splitting the total into parts to conclude that the model CN and the Poisson model have comparable predictive performances. Note another, although typical text-book, methodological error: "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!" [Press et al. 1992]. Still, a single flip of a coin does not allow reliably assessing whether the coin is fair or not.

As an example, let us illustrate the uncertainty of binomial testing with the best case statistics of 4 successful predictions out of 5 target events in the CN North sub-region [Molchan et al. 2017]. The Poisson model with the same rate of alarm, i.e., 35.7%, provides such or a better hit score in 5.8% cases. However, due to stability of the rate of alarm, an additional single one prediction success will change the hit score to

5 out of 6, which corresponds to 2.4% cases of the same or better score by random guessing and rejection of the Poisson model at significance level of 5%. On the other hand, an additional single one failure-to-predict will change the hit score to 4 out of 6 which corresponds to significance level of 12.5% that does not allow excluding the random guessing alternative to the CN predictions in North sub-region.

The considerations made by Molchan et al. [2017] do not contradict statements by Peresan et al. [2011 and 2012] about the significance of CN results, since their conclusions are based on a significant number of target earthquakes, as provided by aggregate analysis of the three regions [following rules well specified in Peresan et al. 2005]. The analysis and considerations by Taroni et al. [2017] persistently overlook the basic definitions and rules of CN application, including declustering procedure and definition of target events [Peresan et al. 2005], therefore their conclusions are fundamentally flawed.

To strengthen the negative conclusion, Taroni et al. [2016] apply, a modification of so-called Pari-mutuel Gambling Score [Zhuang 2010, Zechar and Zhuang 2014]. It is the zero-sum game approach whose results, in general, depend strongly on the "wagers and commission rate". In the considered case, the applied methodology leads to practically complete loss of information on successful prediction. Specifically [Molchan et al., 2017), when, on average, the interevent time between target earthquakes is much larger than the time step of prediction updates (as in the case of CN algorithm application in Italy), a negative verdict about significance of any prediction algorithm is predetermined *a priori*.

The methodological error of Taroni et al. [2016] is that the authors compare the alarm based method with the random forecast (RF), while the actual problem consists in the comparison of two alarm based predictions, namely CN versus random guessing (RG).

Let us recall the difference between RF and RG: RF admits the target event in each time bin with probability p; whereas the RG use this rule to generate alarm zone, suggesting that only within such zone the target event may happen. The comparison of two similar in nature, but different in content, prediction methods, namely PF vs RG, shows that in the framework of the PMGS approach and Poisson seismicity model RF almost always wins against RG [Molchan et al. 2017].

On the other side, the comparison of two alarm based methods, for example CN vs RG, in the framework of the game approach can be quite informative (see GS approach by Zechar and Zhuang [2010]).

Molchan and Romashkova [2011] successfully applied GS in modified form to the analysis of the alarm-based forecasts produced by the M8 algorithm in non-homogeneous time-space bins.

It is of common knowledge that a very limited amount of data is a serious obstacle for a reliable statistical analysis, in particular, when quantifying the performance of an earthquake forecast/prediction method at a regional level. An in-depth comprehensive discussion of methodologies for assessing earthquake prediction results can be found in Molchan et al. [2017] and, earlier, in [Molchan 1997, Molchan and Romashkova 2011]. We recommend the interested readers to consider carefully and critically the paper by Molchan et al. [2017], which provides the basic elements for an objective independent assessment of Taroni et al. [2016], who persistently bypass these elements.

References

Molchan G., Romashkova L., Peresan A. (2017). On some methods for assessing earthquake predictions. Geophys J. Int., 210, 1474-1480, https://doi.org/10.1093/gji/ggx239.

Molchan, G. and Romashkova, L., (2011). Gambling Score in Earthquake Prediction Analysis. Geophys. J. Int., 184, 1445-1454.

Molchan, G. and Romashkova, L., (2010). Earthquake Prediction Analysis. Based on empirical seismic rate: the M8 algorithm, Geophys.J.Int., 183(30,1525-1537.

Molchan, G.M. (1997). Earthquake prediction as a decisionmaking problem, Pure Appl. Geophys, 147(1), 1-15.

Peresan, A., Kossobokov, V., Romashkova, L., Panza, G.F. (2005). Intermediate-term middle-range earthquake predictions in Italy: a review, Earth Sci. Rev., 69, 97-132; doi:10.1016/j.earscirev.2004.07.005

Peresan, A., E. Zuccolo, F. Vaccari, A. Gorshov, and G.F. Panza (2011). Neo-Deterministic Seismic Hazard and Pattern Recognition Techniques: Time-Dependent Scenarios for North-Eastern Italy, Pure and Appl. Geophys. 168, 583-607.

Peresan, A., V.G. Kossobokov, G.F. Panza (2012). Operational earthquake forecast/prediction, Rend. Fis. Acc. Lincei, 23, 131-138. DOI 10.1007/s12210-012- 0171-7.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1992). Numerical recipes in C: the art of scientific computing 2nd ed.. New York: Cambridge University Press, 994 p.

Taroni, M., Marzocchi, W., Roselli, P. (2016). Assessing

'alarm-based CN' earthquake predictions in Italy Annals of Geophysics, 59, 6, S0648; doi:10.4401/ag-6889S0648.

Zechar, J.D., Zhuang, J., (2014). A pari-mutuel gambling perspective to compare probabilistic seismicity forecasts, Geophys. J. Int. 199, 60-68.

Zhuang, J. (2010). Gambling scores for earthquake forecasts and predictions, Geophys. J. Int., 181, 382–390, doi:10.1111/j.1365-246X.2010.04496.x

---

[*]Corresponding author: Antonella Peresan,

Istituto Nazionale di Oceanografia e di Geofisica Sperimentale.

CRS, Udine. Italy

email: aperesan@inogs.it