Assessing 'alarm-based CN' earthquake predictions in Italy

Matteo Taroni[*], Warner Marzocchi, Pamela Roselli

Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy

ABSTRACT

The quantitative assessment of the performance of earthquake prediction and/or forecast models is essential for evaluating their applicability for risk reduction purposes. Here we assess the earthquake prediction performance of the CN model applied to the Italian territory. This model has been widely publicized in Italian news media, but a careful assessment of its prediction performance is still lacking. In this paper we evaluate the results obtained so far from the CN algorithm applied to the Italian territory, by adopting widely used testing procedures and under development in the Collaboratory for the Study of Earthquake Predictability (CSEP) network. Our results show that the CN prediction performance is comparable to the prediction performance of the stationary Poisson model, that is, CN predictions do not add more to what may be expected from random chance.

1. Introduction

Earthquake prediction is one of the ways in which seismologists make statements about the future seismic activity, usually on the basis of the observation of one or more candidate diagnostic precursors [Sykes and Jaumé 1990, Pulinets and Davidenko 2014]. A prediction consists in casting an alarm, i.e., a deterministic assertion that one target earthquake of a given magnitude will occur in a specified space-time window.

Such predictions appear to be prospective deterministic statements, but do not present the full picture. In fact all prediction schemes proposed to date have an intrinsically probabilistic nature [Jordan et al. 2011]. To quantify this probability it is necessary to consider the possibility of raising false alarms and/or to miss some target earthquakes. When an alarm is cast for a specific space-time window, we have a *hit* (true positive) if a target earthquake occurs, otherwise we have a *false alarm* (false positive). When no alarm is cast for a specific space-time window, if a target earthquake occurs we have a *miss* (false negative), otherwise a *correct* (true) *negative*.

Another way in which seismologists make statement about the future seismic activity is through probabilistic forecasting that consists of the estimation of the probability of one or more events in well-defined magnitude-space-time windows (e.g., Marzocchi et al. [2014], for the Italian region). According to Jordan et al. [2011], probabilistic forecasting provides a more complete description of prospective earthquake information than deterministic prediction, and, more important, it separates hazard estimation made by scientists from the public protection role of civil authorities [Jordan et al. 2014]. Probabilistic forecasts can become predictions once a probability threshold is chosen [Zechar and Jordan 2008]; however, this threshold does not have any specific scientific meaning, but it has to be related to the kind of mitigation actions that are associated with the prediction [Marzocchi 2013]. Regardless of these possible shortcomings, it is undoubted that earthquake predictions have a natural and immediate attractiveness for lay people, probably because a prediction is much easier to understand than a probabilistic forecast [Gigerenzer et al. 2005]. In Italian language, the distinction between the English terms "forecast" and "prediction" cannot be made because only one word "*previsione*" exists and it has a strong deterministic connotation.

The large impact on society of earthquake forecast/prediction is a further motivation for carefully evaluating the performances of each model. This is the main target of an international initiative named Collaboratory for the Study of Earthquake Predictability (CSEP) [Jordan 2006, Zechar et al. 2010, Zechar and Zhuang 2014). In essence, CSEP promotes experiments in which prospective forecasts/predictions are compared to real observations in different testing regions. All experiments are rigidly controlled, i.e., they are replicable by anyone and modelers cannot change their forecasts/predictions in retrospect. Besides the statistical evaluation of each single model, these experiments offer also a unique opportunity to compare the relative forecasting/prediction skill of different models. As a

matter of fact, claiming that one model is able to predict earthquakes may be misleading from a scientific point of view: it is much more interesting to compare the model performances with the skill of other competing models. In order to achieve a meaningful comparison, models have to provide forecasts/prediction in a proper common format. However, not all forecast/prediction models adopt such a format, so they are not all presently ready to be evaluated by CSEP experiment. This is the case of the algorithm under study here.

Since January 1, 1998, a group of researchers began to provide earthquake predictions in Italy through the use of a pattern recognition based on CN (California-Nevada, first regions of application) algorithm [Gabrielov et al. 1986, Keilis-Borok et al. 1990, Peresan, et al. 1999, Peresan et al. 2005, Romashkova and Peresan 2013]. This type of prediction allows a quantitative validation of the prediction ability because the method is rigorously applied forward in time. This is certainly the most crucial aspect for the evaluation of any forecast/prediction model, because it guarantees that the results are not affected by any conscious or unconscious adjustment. In this paper we evaluate the results obtained so far from the CN algorithm applied to the Italian territory, by adopting widely used testing procedures and under development in CSEP.

2. The CN model in Italy and the target earthquakes

The CN algorithm is an earthquake alarm-based model that provides intermediate-term predictions for mainshocks. The algorithm casts an alarm when a time of increase probability (TIP) for the occurrence of target earthquakes in one specific region is identified. The identification of a TIP is based on the observation of a set of candidate seismic precursory patterns that have been identified analyzing the past seismicity and are kept fixed [Keilis-Borok et al. 1990]. Originally, the model was set up for California-Nevada (CN) region, but since then it has been adapted to many other regions of the world. The details of the application of the CN algorithm to the Italian territory can be found in Peresan et al. [1999, 2005], and Romanshkova and Peresan [2013] and are not described here.

The CN predictions in Italy are related to three macro-zones that cover part of the Italian territory and are characterized by a different minimum magnitude M_0 for identification of the target events. These macrozones (Figure 1) are northern Italy (N-Italy; $M_0 \geq 5.4$), central Italy (C-Italy; $M_0 \geq 5.6$), and southern Italy (S-Italy; $M_0 \geq 5.6$). The target earthquakes are only mainshocks, i.e., aftershocks above the magnitude M_0 are not considered target events [Peresan et al. 1999]. The forward predictions have been routinely performed every two months since January 1, 1998. More recently a new macrozone has been added (Adriatic region), but it will be not considered here because of the short time window for testing. The reference magnitude of the target earthquakes is taken from the UCI2001 catalog [Peresan et al. 2002]. We do not perform any quality check to this catalog and we take it for granted. The CN predictions for Italy are available in a public website (http://www.geoscienze.units.it/esperimento-di-previsione-dei-terremoti-mt/algorithm-cn-in-italy/cn-predictions-in-italy.html).
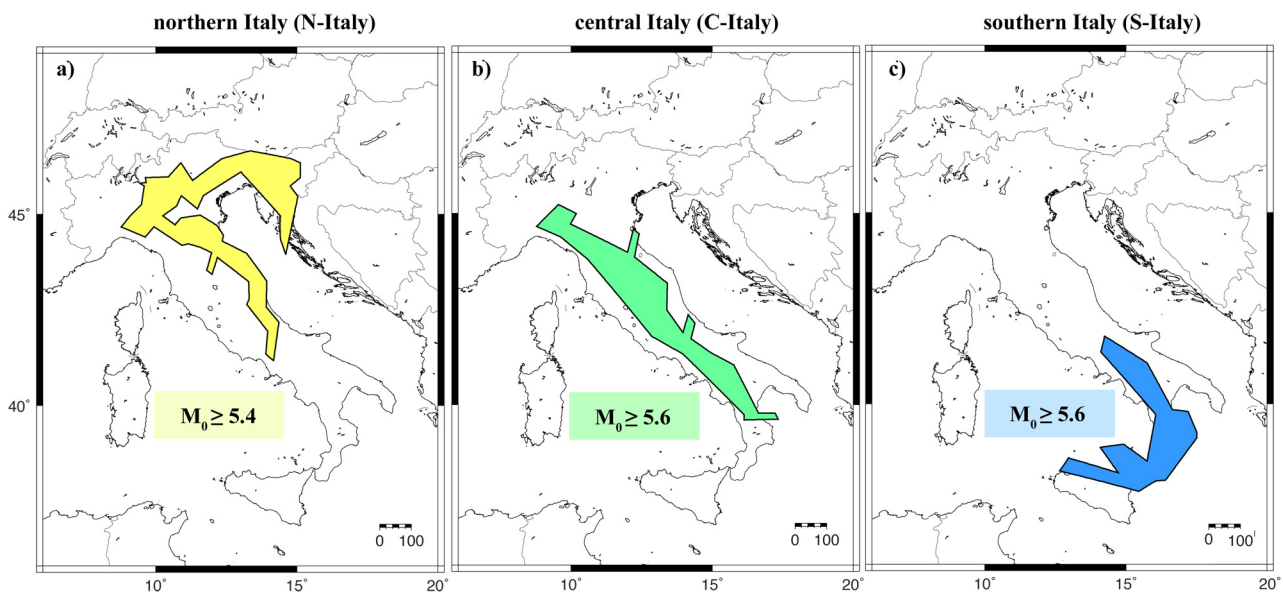

Figure 1. Maps of the three Italian macro-zones related to the regionalization proposed by Peresan et al. [1999]; (a) northern region; (b) central region; (c) southern region. M_0 is the minimum magnitude for the target earthquakes in each macro-zone (for details on magnitudes see Peresan et al. [1999]).

Data (yyyy/mm/dd)	Lat. (°)	Lon. (°)	Depth (km)	M_0	CN-alarm	CN-region
19980412	46.24	13.65	10	6.0	Yes	north
19980909	40.03	15.98	10	5.7	Yes	centre, south
20030914	44.33	11.45	10	5.6	Yes	north
20040712	46.30	13.64	7	5.7	Yes	north
20041124	45.63	10.56	17	5.5	No	north
20090406	42.33	13.33	8	6.3	No	centre
20120520	44.90	11.23	8	6.1	Yes	north
20160824	42.70	13.24	4	6.0	Yes	centre

Table 1. Event list related to CN algorithm website (http://www.geoscienze.units.it/esperimento-di-previsione-dei-terremoti-mt/algorithm-cn-in-italy.html) for earthquakes occurred in Italy since 1/1/1998 to 08/28/2016 and related parameters. Last two columns indicate if there was a TIP (or not) and the macro-zones related, respectively.

Only the current prediction is released by the authors under a password access that has been given to a list of interested scientists.

In Table 1 we show the list of target earthquakes that occurred during a defined testing period (from 01/01/1998 to 08/28/2016, including the very recent Amatrice earthquake of 08/24/2016) reported on CN web-site. A comparison of this catalog with the data extracted by the NEIC catalog (National Earthquake Information Center; http://earthquake.usgs.gov/earthquakes/search/) for the same region, depth and testing period (Table 2), highlights some inconsistencies; for example, some earthquakes above M_0 occurred outside the three macro-zones as shown in Figure 2.

Here, we do not deepen this inconsistency and take for granted the catalog reported in Table 1 that will be used for the following analysis.

Data (yyyy/mm/dd)	Lat. (°)	Lon. (°)	Depth (km)	Mw
19980326	43.26	12.97	10.0	5.4
19980412	46.25	13.65	10.0	5.6
19980909	40.04	15.98	10.0	5.6
20020906	38.38	13.70	5.0	6.0
20021031	41.79	14.87	10.0	5.9
20030329	43.11	15.46	10.0	5.5
20081223	44.56	10.41	28.3	5.4
20090406	42.33	13.33	8.8	6.3
20120520	44.89	11.23	6.3	6.0
20160824	42.71	13.17	10.0	6.2

Table 2. Event list extracted by the NEIC website (National Earthquake Information Center; http://earthquake.usgs.gov/earthquakes/search/) for earthquakes occurred in Italy since 1/1/1998 to 28/8/2016 and related parameters.
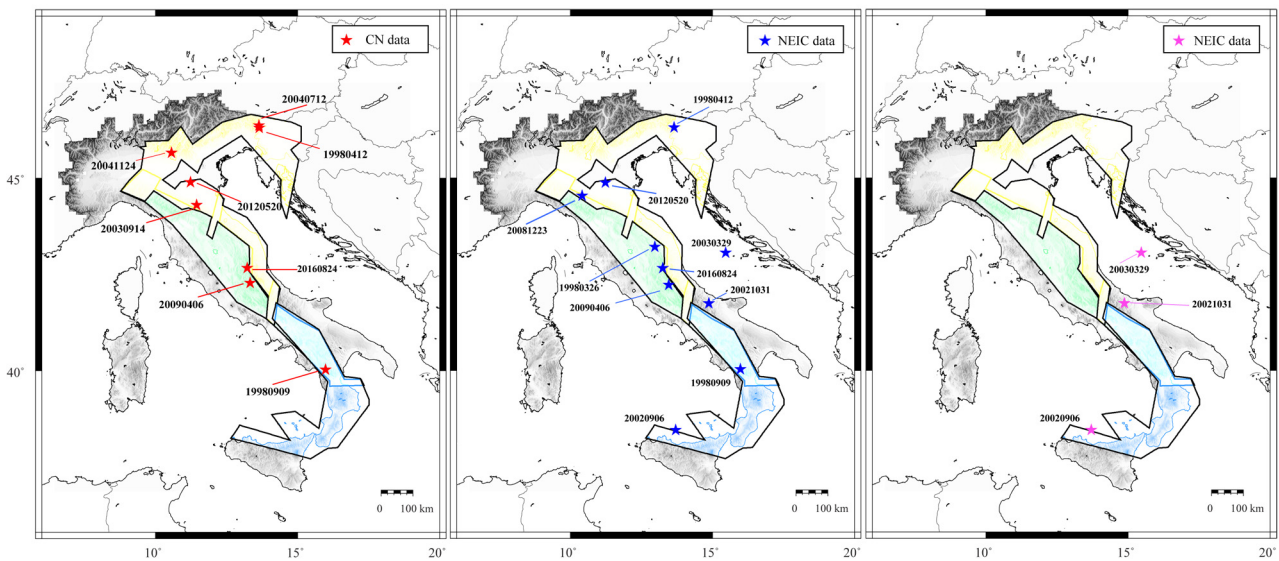

Figure 2. (a) Location map of earthquakes (red stars) listed in Table 1. (b) Location map of earthquakes (blue stars) listed in Table 2. (c) Location map of earthquakes (magenta stars) listed in Table 2 that did not occur inside the macro-zones considered by the CN model.

3. Assessing the CN predictions

We test the CN predictions by using two different statistical methods: the Molchan test (MT) [Molchan 1997, Zechar and Jordan 2008] for the alarm-based model, and the parimutuel gambling score (PGS) [Zechar and Zhuang 2014] for the evaluation of the earthquakes forecasts.

The MT and PGS methods have some remarkable differences. The MT test is based on the definition of a null hypothesis that can be rejected or not according to the observed data. On the contrary, in PGS test the null hypothesis is not requested, because the score is intended to provide a rank of the models based on their relative predictive skill.

Molchan test (MT)

The MT verifies if the number of *hits* for the CN model (in the space-time domain) is consistent with the number of *hits* expected by a reference model (e.g. Poisson model). When the observed number of *hits* is sufficiently bigger than those expected by the Poisson model, we conclude that the model under test is significantly better than the reference model [Zechar and Jordan 2008]. This test is usually represented through a diagram, using different fractions of space-time occupied by alarms. In our case, the CN algorithm sets one specific value for such a variable, so the test collapses into a Bernoulli case. Specifically, h is the number of hits for the CN model, N is the total number of target events, and the parameter of the Bernoulli distribution is given by τ, that is the percentage of the space-time covered by alarms (for instance, if the alarms cover the whole interval of time and half space, then $\tau = 0.5$). This choice of the parameter is appropriate for our specific case, while it may require much more elaborated estimations in testing other earthquake prediction models [e.g., Marzocchi et al. 2003, Molchan and Romashkova 2010].

The null hypothesis H_0 under testing is: the CN model and the Poisson model have the same predictive performances. Under H_0, a binomial distribution describes the probability P of a Poisson model to obtain h or more correct predictions, with N observed target events and τ is the percentage of the space-time occupied by alarm; that is, τ is the probability to observe one target earthquake in a pure random process described by the Poisson distribution. The Poisson model is chosen because it can be considered the simplest random guess forecast strategy within each zone [Kagan 2009].

The binomial distribution reads

$$P = \sum_{i=h}^{N} \binom{N}{i} \tau^i (1-\tau)^{N-i} \qquad (1)$$

In the classical Neyman-Pearson statistical testing framework [Neyman and Pearson 1933] we can 'reject' or 'not reject' the null hypothesis, if P is less or larger that a pre-selected significance level. In practice, we set a significance level of 0.05 (this value is commonly used in science) and we reject the null hypothesis if P is less than 0.05; otherwise, we conclude that there is no empirical evidence supporting that the CN prediction capability is superior to the Poisson process.

Parimutuel gambling score (PGS)

The PGS method is a useful tool to compare the prediction performances of two models: in our case, CN and Poisson models. The PGS method works like each model is a gambler. In the space-time bins there is a gamble between models; each model bets 1 coin on this bin. Then, each model gets a win proportional to the probability assigned to the event (earthquake or not) that has occurred on the bin. In mathematical terms

$$W_j = -1 + k \frac{p_j}{\sum_{m=1}^{k} p_m} \qquad (2)$$

where W_j is the gain that the j-th model obtains in the bin, k is the total number of the models and p_j is the probability of the occurred event (earthquake or not) for the j-th model. The sum of the W_j for each bin represents the skill of the models: the bigger the win, the better the model. The sum of the wins and losses for all the k models is always zero.

In previous works the PGS method has been used to check the performance of the likelihood-based models [Taroni et al. 2014, Zechar and Zhuang 2014]. In this work, we apply the same procedure to the alarm-based models. In particular, for the CN model we define $p_m = 1$ when a target event happens during an alarm or when no target event happens during no-alarm (i.e. *hit* and correct negative, respectively). We define $p_m = 0$ when no target event happens during an alarm or when a target event happens during a period when there is no alarm. (i.e. false alarm and miss, respectively). To compute the probability p_m related to the Poisson model, we use the same macro-zones and the same minimum magnitude used in the CN model. We compute the Poisson rate by using the CPTI11 catalog [Rovida et al. 2011] declustered with the Gardner and Knopoff algorithm [Gardner and Knopoff 1974] until 12/31/1997 (the CN prediction experiment started on 1/1/1998).

4. Results

The Molchan test results are summarized in Table 3, where we shown that P of Equation (1) is never less than the preselected significance level; this means that

Zone	Hit rate (# of hits/ # of EQs)	Percentage of space-time in alarm	P of Equation (1)
north	4/5	35.71%	0.058
center	2/3	35.71%	0.292
south	0/1	25.00%	1.000

Table 3. Molchan test results.

Zone	CN model	Poisson model
north	-32.7	32.7
center	-36.7	36.7
south	-27.7	27.7

Table 4. Parimutuel gambling score results (bold italic for the best model).

we cannot reject the null hypothesis for all three macrozones with the pre-selected significance level of 0.05.

The Parimutuel Gambling Score results are shown in Table 4; the results indicate that the Poisson model is better in all three macro-zones (scoring 32.7 for the north macro-region, 36.7 for center, and 27.7 for south). Of course the scoring of the CN model has opposite values.

5. Conclusions

Considering the data available so far, the Molchan Test does not show that CN prediction performance is significantly better than predictions based on the stationary Poisson model. Moreover, the results of parimutuel gambling score indicate that the Poisson model is even better than CN in predicting earthquakes, as shown by the values associated to each win.

This result is similar to what obtained by the testing of other similar pattern recognition models performed at a global scale once a proper null hypothesis is used [Marzocchi et al. 2003, Zechar and Zhuang 2010].

From a practical perspective, the results show that CN predictions do not add significant information that may be used to enhance societal earthquake preparedness.

References

Gabrielov, A.M., O.E. Dmitrieva, V.I. Keilis-Borok, V.G. Kosobokov, I.V. Kuznetsov, T.A.Levshina, K.M. Mirzoev, G.M. Molchan, S.K. Negmatullaev, V.F.

Pisarenko, A.G. Prozorov, W. Rinehart, I.M. Rotwain, P.N. Shebalin, M.G. Shnirman and S.Y. Schreider (1986). Algorithms of Long-Term Earthquakes' Prediction, International School for Research Oriented to Earthquake Prediction-Algorithms, Software and Data Handling, Lima, Peru.

Gardner, J.K., and L. Knopoff (1974). Is the sequence of earthquakes in southern California, with aftershocks removed, Poissonian?, B. Seismol. Soc. Am., 64 (5), 1363-1367.

Gigerenzer, G., R. Hertwig, E. Van Den Broek, B. Fasolo and K.V. Katsikopoulos (2005). "A 30% chance of rain tomorrow": How does the public understand probabilistic weather forecasts?, Risk Anal., 25 (3), 623-629.

Jordan, T.H. (2006). Earthquake predictability, brick by brick, Seismol. Res. Lett., 77 (1), 3-6.

Jordan, T.H., Y.T. Chen, P. Gasparini, R. Madariaga, I. Main, W. Marzocchi, G. Papadopoulos, G. Sobolev, K. Yamaoka and J. Zschau (2011). Operational earthquake forecasting: state of knowledge and guidelines for utilization, Annals of Geophysics, 54 (4), 315-391.

Jordan, T.H., W. Marzocchi, A. Michael and M. Gerstenberger (2014). Operational earthquake forecasting can enhance earthquake preparedness, Seismol. Res. Lett., 85, 955-959.

Kagan, Y.Y. (2009). Testing long-term earthquake forecasts: likelihood methods and error diagrams, Geophys. J. Int., 177 (2), 532-542.

Keilis-Borok, V.I., I.V. Kuznetsov, G.F. Panza, I.M. Rotwain and G. Costa (1990). On intermediate-term earthquake prediction in Central Italy, Pure Appl. Geophys., 134 (1), 79-92.

Marzocchi, W., L. Sandri and E. Boschi (2003). On the validation of earthquake-forecasting models: the case of pattern recognition algorithms, B. Seismol. Soc. Am., 93 (5), 1994-2004.

Marzocchi, W. (2013). Seismic hazard and public safety, EOS Tran. AGU, 94, 240-241.

Marzocchi, W., A.M. Lombardi and E. Casarotti (2014). The establishment of an operational earthquake forecasting system in Italy, Seismol. Res. Lett., 85 (5), 961-969.

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

Molchan, G.M., and L.Romashkova (2010). Earthquake prediction analysis based on empirical seismic rate: the M8 algorithm, Geophys. J. Int., 183, 1525-1537.

Neyman, J., and E.S. Pearson (1933). On the Problem of the Most Efficient Tests of Statistical Hypotheses, Philos. T. Roy. Soc. A, 231, 694-706.

Peresan, A., G. Costa and G.F. Panza (1999). A proposal of regionalization for the application of the CN earthquake prediction algorithm to the Italian territory, Annali di Geofisica, 42 (5), 883-896.

Peresan, A., I. Rotwain, I. Zaliapin and G.F. Panza (2002). Stability of intermediate-term earthquake predictions with respect to random errors in magnitude: the case of central Italy, Phys. Earth Planet. Inter., 130 (1), 117-127.

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

Pulinets, S., and D. Davidenko (2014). Ionospheric precursors of earthquakes and global electric circuit, Adv. Space Res., 53 (5), 709-723.

Romashkova, L., and A. Peresan (2013). Analysis of Italian earthquake catalogs in the context of intermediate-term prediction problem, Acta Geophys., 61 (3), 583-610.

Rovida, A., R. Camassi, P. Gasperini and M. Stucchi (2011). CPTI11, the 2011 version of the Parametric Catalogue of Italian Earthquakes. Milano, Bologna, http://emidius.mi.ingv.it/CPTI.

Sykes, L.R., and S.C. Jaumé (1990). Seismic activity on neighbouring faults as a long-term precursor to large earthquakes in the San Francisco Bay area, Nature, 348, 595-599.

Taroni, M., J.D. Zechar and W. Marzocchi (2014). Assessing annual global M6+ seismicity forecasts, Geophys. J. Int., 196, 422-431.

Zechar, J.D., and T.H. Jordan (2008). Testing alarm-based earthquake predictions, Geophys. J. Int., 172 (2), 715-724.

Zechar, J.D., and J. Zhuang (2010). Risk and return: evaluating RTP earthquake predictions, Geophys. J. Int., 82, 1319-1326.

Zechar, J.D., M.C. Gerstenberger and D.A. Rhoades (2010). Likelihood-based tests for evaluating space-rate-magnitude earthquake forecasts, B. Seismol. Soc. Am., 100, 1184-1195.

Zechar, J.D., and J. Zhuang (2014). A parimutuel gambling perspective to compare probabilistic seismicity forecasts, Geophys. J. Int., 199 (1), 60-68.

Corresponding author: Matteo Taroni,
Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy;
email: matteo.taroni@ingv.it.