On the validation of the CN algorithm applied to Italian territory

(Warner Marzocchi, June 2004)

1. Premise

Since January 1, 1998, a group of researchers began to provide "prediction" of earthquakes in Italy through the use of a pattern recognition based algorithm called CN. The procedure allows a quantitative validation of the forecasting ability because the method is rigorously applied forward in time. This brief report is an attempt to provide a first evaluation of the results reported so far (more than 6 years of forward testing). In order to do that, we **assume** (without proving) that all the data used by the authors were assembled homogeneously (homogeneous seismic catalog), and that the patterns used for "prediction" were obtained through a rigorous scientific methodology. The scientific validation of such hypotheses is beyond the scope of the present report.

2. Some preliminary considerations on the CN "predictions"

A detailed description of the CN code can be found in the web page and the references reported therein. Here, we focus our attention only on some specific issues that, in my opinion, deserve to be clearly stated.

The Italian territory is divided in three tectonic macrozones partially overlapped, called NORD, CENTRO and SUD. The area of the zones is ~ 83,000 Km^2 for zone NORD (more than twice the Switzerland), ~ 80,000 Km^2 for zone CENTRO (about twice the Switzerland), and ~ 62,000 Km^2 for zone SUD (about 1.5 times the Switzerland). The target earthquakes have M=5.4+ for NORD, and M=5.6+ for CENTRO and SUD.

At first, we need to clarify the meaning of "prediction" and the derived practical implications. Usually, the scientific literature defines earthquake prediction as the specification of the time interval, geographical area, and magnitude range of a future earthquake within stated limits of uncertainty. This definition, however, unavoidably raises a scientific paradox. In fact, we can state that it is very easy to predict (successfully) earthquakes, depending on the spatio-temporal window used. For instance, I "predict" the occurrence of an earthquake of magnitude in the range 3 to 4, in the next week, in the northern hemisphere. In this case, the time interval, the magnitude range, and the geographical area are clearly defined, and it is undoubted that the "prediction" will be successful.

Even though the semantic nature of this issue might appear to be only of philosophical concern, we remark that the abuse, or the incorrect use, of the term "prediction" can create unrealistic expectations in the nonscientific community. In practice, for nonscientific community the term "prediction" is almost always associated to rational strategies for earthquake risk mitigation, such as evacuations. It is obvious that this could be possible only for small spatio-temporal window. On the contrary, if the spatio-temporal window is large, even in case of successful "prediction" we can hardly suggest practical measures different from the ones

commonly suggested by more classical seismic hazard maps.

In a recent paper some authors (Panza, Peresan, and Gorshkov; *Mitigazione della pericolosità sismica - Scenari deterministici del moto del suolo*. 21 SECOLO Scienza e Tecnologia n.4-2002[a]) suggest that the "prediction" made by CN can be useful also to guide verifications of the preparedness of rescue measures and the efficiency of communication roads, as well as to suggest structural reinforcement of specific buildings. Taking into account the dimension of the spatio-temporal window considered by CN, I think that the actual practical measures that can be taken by Civil Protection are not different from what usually done to mitigate long-term seismic risk. For example, I do not understand what kind of specific practical measures could be taken if the NORD zone (ranging from Lazio region, to Slovenia, up to Liguria) is in "alarm" for years (like the present case).

3. Evaluating the forecasting ability

In order to evaluate the capability of any kind of prediction model, we need to check if the prediction model is able to make a "better job" in forecasting earthquakes compared to models based on past seismicity of magnitude comparable to the target earthquakes (see, Marzocchi et al., 2003[b]).

In the table, I list the target earthquakes occurred since 1998 reported by the NEIC catalog, and by the authors in their web page (marked with asterisks). Note that a significant difference exists between the two lists. Part of these differences might be due to my scarce understanding of the magnitude assigned to the earthquakes (therefore of the selection of target earthquakes). In any case, I note that the authors do not report at least three earthquakes (Val Venosta and two Molise earthquakes) occurred in Italy since 1998 that certainly have a magnitude larger than the thresholds considered in the closest macrozones.

Date	***Magn.***	***Lat. & Long.***	***Region***	***Comment***
Mar. 26, 1998	5.6	43.3 13.0	Umbria-Marche	Inside zone CENTRO **Predicted**
Apr. 12, 1998*	6.0	46.2 13.6	Slovenia	Inside zone NORD **Predicted**
Sept. 9, 1998*	5.9	40.0 16.0	Pollino	Inside zone CENTRO **Predicted** Inside zone SUD **Not predicted**

[a] Translation: "Mitigation of seismic hazard: Deterministic scenarios of the ground motion". Personal comment: it is impossible to mitigate the seismic hazard! We can mitigate seismic risk, while the hazard can only be estimated

[b] Marzocchi et al., 2003; BSSA vol. 93, 1994-2004.

Date	*Magn.*	*Lat. & Long.*	*Region*	*Comment*
Aug. 21, 2000	5.4	44.9 8.5	Alessandria	Outside macrozones **Not predicted**
Jul. 17, 2001	5.6	46.7 11.2	Val Venosta	Outside macrozones **Not predicted**
Sept. 6, 2002*	6.0	38.4 13.7	Off coast of Palermo	Outside macrozones **Not predicted**
Oct. 31, 2002	5.9	41.8 14.9	Molise	Outside macrozones **Not predicted**
Nov. 1, 2002	5.8	41.7 14.9	Molise	Outside macrozones **Not predicted**
Sept. 14, 2003*	5.6	44.3 11.4	Monghidoro (BO)	Inside zone NORD **Predicted**

From the table, we can see that 5 out of 9 events occurred outside the three macrozones, therefore "not predicted". The other 4 earthquakes were "predicted".
A fundamental issue to evaluate the efficiency of the prediction is the temporal coverage of the alarms. Until the end of April 2004 (76 months of forward testing), the NORD zone has been set in alarm for 46 months (about 61% of time coverage); the CENTRO zone has been set in alarm for 34 months (about 45% of time coverage); for SUD zone there were no alarms.

In summary, since January 1998, **9** target earthquakes occurred in Italy; **5** of them occurred outside the macrozones considered, therefore "**not predicted"**. The other **4** target earthquakes were **"predicted"**, keeping in alarm 46 out of 76 months an area (NORD) more than twice of Switzerland, and 34 out of 76 an area (CENTRO) twice of Switzerland.

At this stage of the forward testing, I think that a rigorous and formal test is not yet necessary, because the results reported above clearly indicate that the method proposed by the authors does not make a "better job" (from both scientific and practical point of view) of what we can do by modeling statistically the past earthquakes with a magnitude similar to the target earthquakes (see Marzocchi et al., 2003; Faenza et al., 2003[c]).

[c]Faenza et al., 2003; Geophys. J. Int., vol. 155, pp. 521-531.