Evaluation of Earthquake Predictions

Recommendations to the USGS Earthquake Hazards Program from the
National Earthquake Prediction Evaluation Council (NEPEC), September 2016

Background

Many people within and outside the Earth science community are keenly interested in earthquake prediction. It remains a grand scientific challenge. The National Earthquake Prediction Evaluation Council (NEPEC) was created in 1980 to provide authoritative evaluation of earthquake predictions, in cases where there has been sufficient research and testing of a prediction method to warrant consideration. In the past, the NEPEC has also been consulted in situations in which a prediction based on untested or spurious methods caused serious public policy concern. The NEPEC meets as needed at the request of the USGS, and reports to the USGS Director and head of the Earthquake Hazards Program (EHP). The goals of this document are, first, to explain the roles and responsibilities of the USGS and NEPEC, and, second, to offer guidelines for effectively composing a prediction and testing an earthquake prediction method.

In this document we will use the term prediction in a general way as is typically done by the public. In everyday English the words prediction and forecast have the same meaning, but in earthquake science the terms are used in slightly different ways as we define below. Nevertheless, for our purposes of discussing the formulation and evaluation of statements about the future occurrence of earthquakes, distinguishing between predictions and forecasts is important.

Despite widespread optimism in the 1970s and 1980s, reliable short-term (over days to a year) prediction of large earthquakes has proven an elusive goal. Scientists have made substantial progress with the forecasting of long-term average earthquake rates in certain areas, as well as of rates of aftershocks following large earthquakes. However, decades of rigorous research have failed to produce a reliable short-term prediction method. This lack of progress has led many researchers to conclude that short-term prediction may be impossible.

If a viable earthquake prediction method can be developed, it will need to be tested rigorously to prove that it works. The method must be clearly explained using scientifically credible ideas, and tested using available data. The NEPEC cannot provide a scientific evaluation of predictions based on methods that are not clearly and rigorously stated or that have not been tested. In turn, the USGS relies on the NEPEC for evaluation of proposed prediction methods, and will not consider or endorse any prediction method that has not been first tested, and then vetted by the NEPEC. Public broadcasting of predictions before expert evaluation is strongly discouraged, as is basing actions on untested methods.

Definitions

To explain in more detail how predictions need to be formulated and tested, we define what we mean by key terms in this document, and then detail the criteria needed for effective development and testing of earthquake predictions.

- An ***earthquake prediction*** is a statement that one or more earthquakes of a clearly stated magnitude range will occur within a clearly specified time interval and a clearly specified geographic region.
- An ***earthquake forecast*** is a statement of probabilities that one or more earthquakes of a clearly specified magnitude range may occur within a clearly specified time interval and

a clearly specified geographic region. Thus, the statement “a magnitude 7 or greater earthquake will occur in California this year” is a prediction; the statement "there is a 70% chance of a magnitude 7 or greater earthquake in California in the next year" is a forecast.

- A special type of forecast is a ***time of increased probability (TIP),*** which is a statement that one or more earthquakes of a clearly stated magnitude range are more likely than usual to occur within a clearly specified time interval and a clearly specified geographic region. A TIP is a type of forecast but without necessarily specifying a probability. Without a statement of probability it is more difficult to rigorously test the approach.
- An ***earthquake prediction method*** is a recipe for issuing an earthquake prediction whenever certain specified criteria are met. Similarly an ***earthquake forecasting method*** is a recipe for estimating earthquake probabilities.
- The phrase ***“earthquake rates based on normal conditions”*** refers to probabilities of earthquake occurrence that are based on the established average long-term rate of earthquakes in a given area plus the rate of aftershocks that typically follow earthquakes. To demonstrate that a proposed prediction or forecasting method is successful, it must be shown to perform significantly better than just predicting the expected earthquake rates based on normal conditions.
- An ***anti-prediction*** is a statement that *no* specified earthquakes above a certain magnitude will occur within a clearly specified time interval and region. Similarly, a forecast or TIP could be for a lower than usual earthquake rate. This information, which some methods produce, is useful in the rigorous testing of the method as outlined below.
- Predictions and forecasts can cover a wide range of time periods from short-term (days to a year), intermediate-term (a year to a few years) or long-term (several years to decades).

The requirements for testing predictions, forecasts, and TIPs, regardless of the length of time, are the same and so we discuss them together, below. For convenience, we will refer to all three types as predictions, except in situations where using the term forecast is important.

Evaluation of a Prediction Method

A. Formulation of the Prediction

To evaluate a proposed earthquake prediction, it must first be stated clearly. Any earthquake prediction must include an unambiguous definition of the time, location, and magnitude windows of the predicted event. The beginning and end time of the prediction window must be unambiguously specified. For example, “between January 18, 00:00 GMT and January 24, 23:59 GMT, 2019” is a clear time specification, but “around the time of the full moon on 21 January 2019” is ambiguous.

The location of an earthquake must also be clearly specified. For example, a predicted hypocenter should be described in terms of a closed region. Examples of well-defined locations include “within 100 km of Pasadena, California (latitude 34.15, longitude -118.14),” or “between latitude 32 and 33, and between longitude -117 and -119." A prediction stated as “near Pasadena, California,” or “on the San Andreas Fault near Bakersfield” is not sufficiently precise because the word “near” is open to interpretation.

The size of a predicted earthquake must also be clearly specified, including the magnitude range and magnitude scale. It is further necessary to specify which authoritative earthquake catalog will be used to evaluate whether or not a prediction is successful. Terms like “a strong earthquake" or "a damaging earthquake" are not meaningful. If rounding off to the nearest integer or tenth of an integer in magnitude is meant, it must be explicitly stated.

Other criteria defining a predicted earthquake should be clearly specified. If the prediction does not include aftershocks, a specific way of identifying aftershocks must be provided. If only a "strike-slip" or "shallow" earthquake is to be counted, then those terms must be defined in a precise way. If an earthquake is predicted to break a specified extended segment of a large fault, the predicted segment needs to be clearly stated. For forecasts the probability must also be clearly stated, such as the "the probability of one or more events is 40%," or "the Poisson rate of these events, in the specified time window, is 0.5." For a TIP, one must state that it is a time of increased probability rather than a prediction.

B. Significance and Probability

The scientific community has already determined reliable long-term average rates of earthquakes in many areas, and can estimate the probability that any given earthquake prediction will be successful based on random chance, given normal conditions. The demonstrated success of any earthquake prediction method therefore depends on improving the estimation of the probability that it will occur, compared to the current estimate of the probability that the earthquake would have occurred under normal conditions. For example, a successful prediction of an earthquake $M>6$ worldwide in the coming week would not be impressive because such events are very likely to occur by random chance.

C. Methods and data

Earthquake prediction methods may be based on entirely empirical geophysical or environmental observations, or on a physical model of the earthquake generation process. In either case, the methods must be specified clearly enough that they can be reproduced by others. Peer-reviewed publications are not strictly necessary, but are very helpful. An evaluation of the method requires that the data used must be publicly available, archived, and adequately described.

A common flaw in prediction research has been the after-the-fact identification of an apparently significant precursor (i.e., a signal that occurred prior to, and is physically linked to, a large earthquake) solely because it is an outlier or otherwise unusual among observations from a short period of time. To demonstrate that a proposed precursor is significant, it is necessary to analyze data from a suitably long period of time to understand the natural variability of the data. If the statistical variability of the data is described, that provides a rigorous basis for identifying when an observation is anomalous. A reproducible prediction method should include a means by which to objectively identify proposed precursors in the data without subjective human intervention.

D. Testing criteria and results

Claims of discovery of a prediction method are sometimes based on what we call *retrospective tests.* A retrospective test uses earthquake-related data collected before the method was developed. Retrospective testing is often done to develop a method and give assurance that the method is worth pursuing. Demonstration of the success of a method, however, requires testing the method on data that were not used in its development. The details of the method must be established clearly, and other scientists must be able to reproduce its results completely independently, for example by coding the same algorithm in a different programming language and running it on a separate computer.

The most rigorous form of testing is what we call *prospective testing*. Prospective testing means that the success of the method is evaluated according to whether or not it successfully predicts future earthquakes (i.e., those that had not yet occurred when the method was developed), more successfully than expectation of normal conditions, and without undergoing modification during the testing period. While a prediction method is under evaluation, predictions from prospective tests might be circulated only amongst a trusted group of referees, rather than made publicly available. However, prediction researchers are encouraged to build trust within the scientific community by openly publishing methods and tests, and where possible by submitting methods for independent prospective testing by established scientific groups who will verify that the method has been documented and remains unchanged during the test.

The duration of an effective prospective test depends on several factors; in general testing will be challenging because earthquakes of societal concern occur infrequently. In general, the duration of the needed test can usually be estimated based on prior observations, but we are still improving our methods for that process. Developers of earthquake prediction methods need to be aware that testing periods will typically be measured in years. In the following section we describe the testing period for a project undertaken recently by the research community.

Results from retrospective and prospective tests should include:

- All times and areas considered, and all times when the method was not employed (i.e., gaps in the tested interval).
- All times, locations, and magnitudes for predictions issued. For retrospective tests, this includes all events that met the target criterion for predicted events. For prospective tests, the complete prediction statement should be given for each instance, when the predictions were made and distributed, as well as the distribution list for that prediction.
- Depending on the nature of the method, there may also be a list of times for which there was a prediction that no earthquake would occur (i.e., an anti-prediction).
- Documentation of any changes over time in the quality or availability of input data.
- From the above lists, classification of (1) predictions that were successful (an earthquake occurred that matched the prediction criteria), (2) false alarms (no earthquakes fulfilled the prediction), and (3) missed predictions (an earthquake occurred when no prediction was in force).

Example of effective evaluation

Under guidelines established by the Regional Earthquake Likelihood Methods (RELM) project <www.relm.org> of the USGS and the Southern California Earthquake Center, various researchers registered competing techniques for forecasting $M \geq 4.95$ earthquakes in California. The methods evaluated were not short-term prediction methods per se, but rather forecasts of rates of moderate-to-large earthquakes that the community felt could be tested within a reasonable time frame. Project leaders determined that, based on earthquakes observed in California during the past century, a five-year test would be sufficient to accumulate a sufficient number of qualifying earthquakes to usefully compare the performance of the different methods.

During the five-year testing period for the RELM project, enough moderate-to-large earthquakes occurred in California to evaluate which of the proposed methods were most successful in predicting the numbers and locations of the earthquakes that occurred. Indeed, there were clear contrasts in performance between competing models. The best-performing methods do provide a basis for improving forecasts of expected earthquake rates under normal conditions. However, since the formal conclusion of the RELM project in 2010, concerns have been raised about the stability of the results and it is now recognized that an even longer test is needed to get definitive answers to some questions. In particular, it has become apparent that had testing continued, the ranking of methods might have changed and/or the improvements to forecasts could have appeared less significant.

A testing facility for proposed prediction methods

While none of the methods provide a basis for reliable short-term earthquake prediction according to the criteria explained in this document, the best-performing methods do provide a basis for improving forecasts of expected earthquake rates under normal conditions. The Collaboratory for the Study of Earthquake Predictability (CSEP) www.cseptesting.org has built upon RELM to develop procedures for registering and rigorously and independently testing a broad range of prediction algorithms. CSEP offers an open and well-defined test bed for predictions. At the time of this writing, there are hundreds of methods undergoing prospective testing at CSEP centers in a number of regions worldwide.

Roles and responsibilities of USGS and the NEPEC

The USGS serves as the federal government's principal source of expertise on earthquake science, including earthquake prediction. Its Earthquake Hazards Program supports research on earthquake physics, occurrence and likelihood, both within its own scientific ranks and via research grants and cooperative agreements with experts in academia, state agencies, and the private sector. The USGS partners with the National Science Foundation on additional research and the construction and operation of seismic and geodetic networks to support earthquake research, and with other federal agencies including NASA and the Department of Energy.

An important focus of USGS work is the forecasting of earthquake occurrence and effects, and conducting targeted research to develop and improve such forecasts. USGS products such as the National Seismic Hazard Maps and the Uniform California Earthquake Rupture Forecasts (UCERF) forecast where earthquakes are likely to strike, the magnitudes and frequencies of such earthquakes, and the probability of strong ground motion and ground failure. USGS science, in turn, also supports forecasts of risk to structures and communities, and potential economic losses and casualties. Following significant earthquakes in the United States, the USGS provides forecasts of aftershock probability.

Although there is no current ability to predict the time of occurrence of major earthquakes, under the National Earthquake Hazards Reduction Program the USGS has lead federal responsibility for earthquake prediction and the evaluation of proposed prediction methods. To support this mandate, the NEPEC provides advice and recommendations to the Director of the USGS on earthquake prediction and related scientific research. The NEPEC is a Federal Advisory Committee originally created by Congressional legislation in 1980. The Council includes members appointed by the Director of USGS from academia, the USGS, and the private sector who are experts on a range of relevant topics, and consults as needed with other experts.

The NEPEC does not have the responsibility or the resources to study all proposed earthquake prediction or forecasting methods. Nor is it feasible for the NEPEC or the USGS to endorse any prediction method that has not been thoroughly tested and vetted as described in this document. NEPEC aspires to evaluate methods that have demonstrated their potential value through rigorous scientific studies and have a sound statistical and/or physical basis, but may limit evaluations to methods with public policy implications. We encourage people studying prediction methods to carry out rigorous studies as described in this document.

Further reading

Earthquake Predictability, Brick by Brick (T. Jordan, Seis. Res. Lett., V77/1, 2006)

Operational Earthquake Forecasting: Some Thoughts on Why and How (T. Jordan and L. Jones, Seis. Res. Lett., V81/4, 2011).

Predicting the Unpredictable: The Tumultuous Science of Earthquake Prediction, S. Hough, 272 pp, Princeton University Press, Princeton (2009).

The Italian government convened an *International Commission on Earthquake Forecasting for Civil Protection* following the deadly L'Aquila earthquake of 2010. Their report provides a thorough review of the relevant scientific literature, summarizes the state of practice in several countries, and outlines procedures for developing, testing and implementing an operational forecasting capability. (ICEFCP, 2011, reference below)

Technical Articles

International Commission on Earthquake Forecasting for Civil Protection (ICEFCP). "Operational earthquake forecasting, state of knowledge and guidelines for utilization." , *Annals of Geophysics*, vol. 54, no. 4, 315-391, 2011; doi: 10.4401/ag-5350 (free download http://www.annalsofgeophysics.eu/index.php/annals/article/view/5350/5371).

M. K. Sachs, Y. Lee, D. L. Turcotte, J. R. Holliday, and J. B. Rundle, "Evaluating the RELM Test Results," International Journal of Geophysics, vol. 2012, Article ID 543482, 8 pages, 2012. doi:10.1155/2012/543482.

D. Schorlemmer, J. D. Zechar, M. J. Werner, E. H. Field, D. D. Jackson, and T. H. Jordan, "First results of the regional earthquake likelihood models experiment," Pure and Applied Geophysics, vol. 167, no. 8-9, pp. 859–876, 2010.