

## Earthquake Forecasting and Earthquake Prediction: Different Approaches for Obtaining the Best Model

Warner Marzocchi<sup>1</sup> and J. Douglas Zechar<sup>2,3</sup>

### INTRODUCTION

We consider the general problem of constructing or selecting the “best” earthquake forecast/prediction model. While many researchers have presented technical methods for solving this problem, the practical and philosophical dimensions are scarcely treated in the scientific literature, and we wish to emphasize these aspects here. Of particular interest are the marked differences between approaches used to build long-term earthquake rupture forecasts and those used to conduct systematic earthquake predictability experiments. Our aim is to clarify the different approaches, and we suggest that these differences, while perhaps not intuitive, are understandable and appropriate for their specific goals. We note that what constitutes the “best” model is not uniquely defined, and the definition often depends on the needs and goals of the model’s consumer.

Words with the roots “forecast” and “predict” are used nearly synonymously. Indeed, in many languages, these two words translate to the same term, and in everyday English usage, these words are essentially interchangeable. In discussions of meteorological modeling, “forecast” is commonly used to describe a specific quantitative statement, and in a general scientific context one might say that a model “predicts” a phenomenon or behavior. These usages imply that a prediction is less definite than a forecast. However, in the context of earthquake science, this interpretation was turned on its ear by Jackson (1996), who suggested that an earthquake prediction implies a (temporarily) higher probability than an earthquake forecast; in other words, a prediction is more definite than a forecast. As a result, “earthquake forecast” now usually describes a statement or set of statements characterized by relatively low probabilities. Further muddying the waters is the term “assessment” as in “probabilistic seismic hazard assessment”—this term is even vaguer. To be clear: in this article, we take earthquake forecasts to be quantitative statements expressed in probabilistic terms, and because such statements are our emphasis, we hereafter

only use the term “forecast.” Nevertheless, we note that most statements related to future earthquakes can be reduced to (or interpreted in) a common deterministic (alarm-based) framework, and we refer the reader to Zechar and Jordan (2008) for details (see also Marzocchi *et al.* 2003).

Successful forecasting of earthquakes is of primary importance for two complementary reasons. The first is practical: a reliable and skillful forecast is a fundamental component required to mitigate seismic risk. The second is more philosophical: forecasting is a cornerstone of scientific knowledge. Indeed, the American Association for the Advancement of Science, the world’s largest general scientific society, states that “the growing ability of scientists to make accurate predictions about natural phenomena provides convincing evidence that we are really gaining in our understanding of how the world works” (AAAS, 1989, 26). Of course, although an accurate forecast suggests understanding, it might be accurate by chance, or for the wrong reason. Moreover, several thorny questions are inspired by this AAAS statement: Is any forecast of natural phenomena inherently more useful or more important than another? More generally, is all scientific understanding equally important? If not, how should the importance of understanding be measured: by its application to other understanding? By its connection to the problems of society? By the number of times it is cited? We do not presume to answer these questions here, but they are certainly relevant to the earthquake forecast problem.

The basic problem in earthquake forecasting is to obtain the best model; to date, there is not a unique way to achieve this goal. One possible strategy is to apply a particular metric to a set of candidate models and select the optimal model. From a purely scientific point of view, this corresponds to evaluating the reliability and skill of each forecast (International Commission on Earthquake Forecasting for Civil Protection, ICEF 2010). In general, this is fairly straightforward if many data are available. For instance, in seismology, we can usually check the reliability and skill of forecast models applied to small and moderate earthquakes, because these occur frequently. Indeed this is the current practice within the Collaboratory for the Study of Earthquake Predictability (CSEP) testing centers (Jordan 2006; Zechar, Schorlemmer *et al.* 2010). Nevertheless, the largest earthquakes (*i.e.*, **M** 7.0 or larger)—those having the highest impact on society—are rare, and it is therefore difficult to conduct satisfying prospective experiments regarding these events in a short period. Clearly this problem is not unique to seismology; among others, volcanologists face the same difficulty.

In earthquake forecasting, the problem that large events happen so infrequently is managed in two different ways. One

1. Istituto Nazionale di Geofisica e Vulcanologia, Roma, Italy  
2. Swiss Seismological Service, ETH Zurich, Zurich, Switzerland  
3. Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York, U.S.A.

strategy is to assume that the largest events sample the statistical distribution of small-to-moderate events. In this approach, the empirical distribution of small-to-moderate earthquakes is extrapolated to large-magnitude events. The other strategy is to assume that the largest events have some peculiarities that make them distinct from smaller events (*e.g.*, a different distribution and/or different epistemic uncertainty). In this case, extrapolation is not useful and a specific statistical model should be constructed. Unfortunately, we do not have enough data to build a unique, robust model and, even if we did, we do not have enough data to check its reliability or skill in forecasting. To address this problem, some earthquake scientists abandon the common scientific practice of hypothesis testing and instead build consensus models based on expert opinion, or the so-called “best available science.”

These two opposing strategies are now in common usage; the former motivates systematic prospective earthquake forecast experiments, and the latter informs many large-scale, long-term seismic hazard projects. In this article, we consider specific instances from these two fields of interest: the Regional Earthquake Likelihood Models (RELM) experiment and the Uniform California Earthquake Rupture Forecast (UCERF) project. These two initiatives have different goals. While RELM is essentially a scientific experiment to evaluate reliability and skill of a set of forecasting models, UCERF aims to build the “best” rupture forecast model. These differences notwithstanding, we suggest that the different working philosophy of each group can explain the different approaches each group takes to obtain the “best” model. Intuitively, it might seem that one of these two strategies should be preferred for all problems and that the other should be abandoned. However, we suggest some reasons to reconsider this notion in the context of the RELM and UCERF efforts. We suggest that the dissimilarities between the RELM and UCERF approaches can be explained by the fact that the intrinsic purpose of the corresponding forecasts is different, and by the different kind of probability used in UCERF and RELM. Below we discuss some basic principles of earthquake forecasting applied to seismic risk reduction, with an emphasis on the practical and philosophical implications; we briefly report the relevant details of RELM and UCERF; and we explain why these two strategies are appropriate in their current applications.

## PRACTICAL RELEVANCE OF EARTHQUAKE FORECASTING

An earthquake forecast is a basic prerequisite for planning rational risk mitigation actions. Recently, the term “operational earthquake forecast” has become popular as it emphasizes the primary goal of such research: “to provide communities with information about seismic hazards that can be used to make decisions in advance of potentially destructive earthquakes” (Jordan *et al.* 2009; see also Jordan and Jones 2010). In other words, the word “operational” highlights the need to have a reliable and skillful prospective forecast in a format that can be used for risk mitigation. This is not to say that a specific

forecast horizon is optimal, but rather that the horizon dictates which risk mitigation steps may be taken. From the other point of view, if a particular mitigation action is targeted, it defines the horizon of a useful operational forecast.

A short-term forecast model may be used to reduce risk during a sequence of small to moderate earthquakes, which might be an aftershock sequence, a foreshock sequence, or a swarm that is not punctuated by a large earthquake. “Regardless of the sequence details, the appropriate mitigation actions in this example would emphasize a reduction of exposure—the value at risk expressed, *e.g.*, in terms of human beings or economic costs—rather than a reduction of vulnerability—the fraction of the value that could be lost after an earthquake. In other words, a short-term forecast model does practically nothing to inform development of building codes, but it might be useful for deciding whether or not to call for an evacuation. The planning and implementation of such actions are the responsibility of decision-makers, be they from a civil protection agency or otherwise; in this context, as scientists we can contribute by explaining how the forecast can be used effectively (Marzocchi and Woo 2007, 2009; van Stiphout *et al.* 2010). This topic is under active investigation in several regions, but so far a generally accepted procedure has not been formalized.

Regardless of the forecast horizon, the nature of the forecast model, and the specific mitigation actions, a probabilistic model allows scientists and decision-makers to clearly distinguish their roles in risk assessment and mitigation: scientists provide probabilities to decision-makers (*e.g.*, Marzocchi and Lombardi 2009), who then compare them with pre-identified thresholds to choose the appropriate actions (see also Aki 1989; Kantorovich and Keilis-Borok 1991). This procedure is followed by many researchers dealing with different natural hazards, and it has the advantages that it makes the decision-making process transparent and it justifies the selected mitigation actions. In developing thresholds, seismologists should conduct a probability gain analysis, comparing the forecast probability to “background” probabilities based on simple assumptions. Threshold development has been discussed extensively elsewhere (see, *e.g.*, Molchan 1997; Marzocchi and Woo 2007, 2009; Woo 2008), and we emphasize only that the output of any earthquake forecast model should be probabilistic.

Whether a probability is “high” or “low” is always a matter of context, and in the context of decision-making, it is irrelevant. The only question that matters is whether the probability exceeds the threshold for acceptable risk. Moreover, describing a probability as high or low can be misleading when communicating risk to the public.

## A PHILOSOPHICAL VIEW OF EARTHQUAKE FORECASTING

Some seismologists consider forecasting nothing more than a technical application of developed theories, and they therefore think earthquake forecasting is of little scientific interest. We strongly disagree with this dismissive, reductive point of view. Rather, we see earthquake forecasting as a mighty opportu-

nity to check the reliability of theories and models and to gain insight into the issue of which parameters are most important for modeling earthquake occurrence.

Many works on the philosophy of science have noted that forecasting is a fundamental task of the scientific enterprise. The nature and definition of capital-S Science has been debated since Aristotle's time, and the last century has hosted some of the strongest philosophical attacks on the role and intrinsic meaning of Science. Indeed, the four pillars of Science (*e.g.*, Gauch 2003)—rationality, objectivity, realism, and truth—have been questioned by philosophers such as Karl Popper (indirectly through the work of Lakatos 1970; see, *e.g.*, Theocharis and Psimopoulos 1987) and Kuhn (1970), among others. In this article, we do not directly contribute to this interesting and important debate, and we refer interested readers to the excellent review of Gauch (2003). Rather, we echo the AAAS view that an increased ability to forecast natural phenomena is the best argument to justify the role of Science as a valid guide toward objective reality. Moreover, we assert that the ability to provide reliable forecasts is what distinguishes the scientific enterprise from other activities.

Another key philosophical issue is the intrinsic meaning of a forecast. A forecast is always some form of extrapolation and therefore suffers from all of the problems associated with the logic of induction (Gauch 2003). We never have full access to the complete details of the Earth system, and therefore, while a sound forecast model should include all *known unknowns*, it is not possible to account for *unknown unknowns*. In practice this means that any forecast model that worked satisfactorily in a given period may fail in future experiments. This idea has inspired some bold statements, such as the claim that the verification of earth science models is impossible (Oreskes *et al.* 1994). We do not agree with this extreme position. We concede that some of these terms—*e.g.*, reliability and verification—are used too loosely in the earth sciences. Perhaps *accuracy* is more appropriate than *reliability* because it emphasizes the ability to fit a series of independent data rather than implying something about the “truth” of the model. (Nevertheless, we persist in using the term “reliability” because its meaning in the context of earthquake forecasting seems to be well understood.) On the other hand, the paramount importance of inductive logic to Science is unquestionable. For risk mitigation in particular, it is certainly more reasonable to rely on a model that worked well in the past than to take action on the basis of a randomly selected forecast model. The thrust of this argument is that any forecast model may fail in the future despite good performance in the past. This is an unavoidable limitation strictly linked to the nature of extrapolation.

Other (more philosophical) critiques regarding the meaning of forecasts have addressed the nature of Science's implicit presuppositions. A thorough discussion of this debate is beyond the scope of this article, but we report the assumptions that stand behind any reliable forecast. According to AAAS (1989, 25), Science “presumes that the things and events in the Universe occur in consistent patterns that are comprehensible through careful and systematic studies...Science also assumes

that the Universe is a vast single system in which the basic rules are always the same.” These assertions are particularly relevant to earthquake forecasting because some of the critiques reported in high-profile scientific journals (Broad 1979; Theocharis and Psimopoulos 1987) are not addressed to the science of forecasts but rather to the intrinsic meaning of Science.

## THE BEST MODEL: THE RELM APPROACH

The Regional Earthquake Likelihood Models (RELM) working group was supported by the United States Geological Survey (USGS) and the Southern California Earthquake Center (SCEC). Here, we use the term “RELM” to refer to the five-year scientific experiment that is comparing several time-invariant earthquake-rate forecast models in California (Field 2007; Schorlemmer and Gerstenberger 2007; Schorlemmer *et al.* 2010). We note that this is a scientific experiment in the traditional sense: each participant built a model that formulates hypotheses in terms of a forecast, and these forecasts are being tested against an independent set of earthquakes. In other words, the primary goal of the RELM experiment is to evaluate the comparative skill and reliability of each forecast model. Rather than taking place within a traditional brick-and-mortar laboratory, the RELM experiment is being conducted in the natural laboratory of California and ultimately within a CSEP testing center, a computational laboratory (Zechar, Schorlemmer *et al.* 2010).

The RELM experiment embodies the spirit of this proclamation by the AAAS (1989, 27): “A hypothesis that cannot in principle be put to test of evidence may be interesting but it is not scientifically useful.” In the RELM setup, selecting the best model is relatively straightforward once the preferred metric is chosen; forecasts can be ranked by the metric of choice, and the best model will be the one that obtained the best score (*e.g.*, Zechar, Gerstenberger *et al.* 2010). We take RELM to be representative of systematic earthquake predictability experiments driven by hypothesis testing and rigorous statistical evaluation.

Although the RELM experiment and the CSEP initiative that it spurred have undoubted merits, it is important to note some features that may limit the extent to which the experiment results may be applied. For example, due to the exponential distribution of earthquake magnitudes, the smallest qualifying earthquakes will constitute the majority of target earthquakes, and therefore the RELM results will be dominated by the smallest events. May we expect good forecasting performances obtained from the RELM experiment to apply to the largest events, which are frankly more interesting, at least from a practical point of view? Preliminary results indicate that the forecast of **M** 5+ earthquakes is improved if we consider the distribution of smaller, **M** 2+ to **M** 4+, shocks (Helmstetter *et al.* 2007; Schorlemmer *et al.* 2010). This result lends credence to the idea that such conceptual “extrapolation” is reliable, at least for this magnitude range.

But is it reasonable to expect that we can extrapolate RELM results and say something meaningful about **M** 6+, **M** 7+, **M** 8+, or larger temblors? The answer to this question



would receive different answers from members of the seismological community. Generally speaking, there are two end-member positions. Some believe that every earthquake is the same regardless of its magnitude. This belief justifies the application of RELM results for some (most) of the models to the largest earthquakes; even though in principle a RELM model may assume different distributions for small and large earthquakes, in practice most models do not. On the other hand, if the model assumes different distributions for small-to-moderate and large events, the RELM experiment mostly tests the former distribution rather than the latter.

Others believe that system-size events are fundamentally and substantially different from their smaller counterparts, or, more generally, we have additional information about such events that should be taken into account for forecasting. This belief has important implications that appear to move this topic away from the traditional domain of Science. For example, this belief opposes the parsimony that has typically led the evolution of Science—in other words, it defies Occam's Razor. Some scientists and philosophers maintain that the practice of preferring the least-complicated model that can explain the observations is essential for scientific enterprise (*e.g.*, Gauch 2003, and references therein). A second implication of believing that large earthquakes are unique is that it is practically impossible to calibrate and/or to rigorously verify forecast models. Because the largest events may happen only once in a lifetime, achieving robust results would require an impractically long experiment. In this view, testing experiments at a worldwide scale would not substantially improve the situation, because it is practically impossible to have, for example, the same conditions and/or knowledge of the San Andreas fault elsewhere. Both of these implications—moving against parsimony and toward untestable models—seem quite negative, at least from a scientific point of view, and it is therefore remarkable and counterintuitive that UCERF, one of the most well-known forecast initiatives, takes the approach that large earthquakes differ from small earthquakes.

## THE BEST MODEL: THE UCERF APPROACH

The Uniform California Earthquake Rupture Forecast (UCERF) is a project of the Working Group on California Earthquake Probabilities, which is a multidisciplinary collaboration of scientists and engineers from universities, SCEC, USGS, and private companies (Field *et al.* 2009). The primary task of UCERF is to provide a comprehensive framework for computing a rupture forecast for California. By linking the rupture forecast to the ground shaking, UCERF yields important information for improving seismic safety engineering, revising building codes, setting insurance rates, and helping communities prepare for inevitable future earthquakes. Despite recent UCERF efforts that emphasize short-term rupture forecasts, one of the most important targets is the long-term rupture forecast. Hereinafter, the term UCERF refers to the long-term forecast.

Beyond the ambition and sophistication of such a model, perhaps the most striking feature of UCERF is that it is based

on an expert opinion procedure. Conceptually, the relevant pieces of scientific information are treated as modules that are ultimately merged according to an expert opinion evaluation. The final model is a complex convolution of scientific data and expert opinion, where we can think of Science as the component pieces and expert opinion as the glue that binds them.

The UCERF process is based on the implicit assumption that large earthquakes have some peculiarities that make them different from smaller ones. This is a reasonable hypothesis, although the scarcity of large earthquakes makes it difficult to demonstrate convincingly. In this context, the problem of interest is how to *build* the best model when we lack sufficient data for calibrating, testing, and comparing the available models. The UCERF solution is to use the *best available scientific information*. It is worth noting that this phrase comes not from scientists but from the model consumer: insurance companies. The California Insurance Code section 10089.40 (a) states: "Rates shall be based on the best available scientific information for assessing the risk of earthquake frequency, severity and loss...Scientific information from geologists, seismologists, or similar experts shall not be conclusive to support the establishment of different rates...unless that information, as analyzed by...experts in the scientific or academic community, clearly shows a higher risk of earthquake frequency, severity, or loss between those most populous rating territories to support those differences."

From an objective point of view, the concept *best available science* is rather dangerous. It is not clearly defined and creates the possibility of irresolvable controversy. Moreover, the search for the best available science might degenerate into so-called "trifle-worship," the belief that "the more detailed the model, the better." Salt (2008) considered trifle-worship the first habit of highly defective projects. Essentially, those who make this mistake confound details with realism (or precision with accuracy), and this often results in models that are exceedingly complex and, by Popper's (1968) criterion that a scientific model must be testable, less scientific.

UCERF uses an informal expert elicitation procedure to identify the best available science and thereby build the best model. To do this, UCERF has established a community where expert principals discuss the scientific components that should constitute the final model and how these components should interact. Conceptually, the final model is a weighted combination of these components. Ultimately, the scientific components are selected and coordinated subjectively, with component weights decided by the opinions of the relevant experts. For example, expert geologists were polled to determine the most likely slip rate or recurrence interval of particular faults (N. Field, written communication).

## DISCUSSION AND CONCLUSIONS

The RELM procedure for selecting the best forecast model adheres closely to the scientific principle of objectivity, while UCERF does not do the same in building the best model. From this point of view, it might seem that the RELM approach is

good and the UCERF concept is bad. But such a conclusion is simplistic when we consider two issues: the meaning of probability in this context and the meaning of “best model.” The meaning of probability has been hotly debated since the time of Laplace. Is it an objective physical property (like mass or distance), or does it merely express a degree of belief, which is intrinsically subjective? The recent trend is to relax the monist view, where probability has only one specific interpretation, and to embrace a pluralist view, where probability may have different meanings. The pluralist interpretation also accommodates the existence of two types of uncertainty: *aleatory* uncertainty that is the intrinsic (and unavoidable) random variability due to the complexity of the process, and the *epistemic* uncertainty that is due to our limited knowledge of the process. The former cannot be reduced, while the latter can be reduced by increasing our knowledge of the seismogenic process.

The discussion is certainly more complex than this, particularly because these two interpretations are not mutually exclusive; for an excellent review of this topic, see Gillies (2000). In this article, we merely report one aspect that is particularly relevant. Specifically, it has been suggested that the objective interpretation (under which probability is also called *propensity*) is more appropriate for the probability assessment of a sequence of events, while the subjective interpretation is more suitable for the probability assessment of one event. For example, Howson and Urbach (1989) stated that objective probability can be used for outcomes having a set of repeatable conditions. For a single event, it is very difficult to establish a set of repeatable conditions, and this difficulty results in subjectivity that may be faced only through a subjective probability. This distinction is not only semantic or philosophical because subjective probabilities are intrinsically untestable (De Finetti 1931; see below for more details). The same is true of the UCERF model. To be clear, by “untestable” we do not mean that it literally cannot be subjected to a test but rather that the results of any test are not meaningful.

To illustrate the differences between these subjective and objective interpretations of probability, consider the following example paraphrased from Gillies (2000). Chiara is a sixteen-year-old girl living in Rome, and her uncle David is a statistician. He tries to dissuade her from buying a motorbike by citing the frequency of fatal teenage motorbike accidents in Rome, *i.e.*, by telling her that the propensity for a teenager to die in a motorbike accident in Rome is very high. Chiara replies that such a propensity doesn't apply to her because, as he knows, she is much more careful than the average Italian teenager. Chiara's precocious understanding corresponds to the idea that propensity may not apply to a given individual if you know something more about that person's behavior. This example would deserve a long and careful discussion that is beyond the scope of this paper. The point here is that long-term propensity and the probability of the “next” event are often different. With respect to seismology, we have statistics that can describe “propensity” for small to moderate earthquakes. Some researchers believe that we know something more (or different) about the occurrence of large earthquakes—for instance,

that large earthquakes occur only on specific faults, or that the time since the last large earthquake is an important property. Therefore, they too would argue that the propensity derived from a population of small to moderate earthquakes is not applicable to a system-size event like that which is of interest to UCERF.

These two interpretations of probability are directly applicable to RELM and UCERF. While the RELM earthquake forecast models are probabilistic and may be interpreted as a long-term propensity for an earthquake to occur, the probabilistic assessment of UCERF represents an expert group's “degree of belief” applied to the next large events. Objective probability is focused on a sequence of events, and therefore it is assumed that each event has the same probability; subjective probability implies that each event has some peculiarities that make it different from the others. In UCERF, this degree of belief comes from a large community instead of a single person. The subjective probability derived from a group of experts is called inter-subjective (see, *e.g.*, Gillies 2000), and it has the distinct advantage that it results in a set of coherent probability values (*i.e.*, a set of probabilities that satisfy the Kolmogorov axioms of probability), as opposed to the subjective probabilities given by a single person.

Subjective probability is by definition untestable, because it is mostly associated with the epistemic uncertainty that changes with our knowledge of the system (see De Finetti 1931). Let  $A$  be an event in which we are interested, and  $P(A)$  its probability of occurrence. Subjective probability implies that only conditional probabilities exist, and  $P(A)$  should be more precisely written as  $P(A | S, I)$ , where  $S$  is the set of repeatable conditions (that defines a conceptual experiment of which  $A$  is one possible outcome), and  $I$  is the information level. If  $I$  is complete (no epistemic uncertainty exists) or our lack of knowledge is negligible (we define this state with  $IC$ ), this means that we know exactly the set of repeatable conditions  $S$ , and the process (*i.e.*, the appropriate parameterization and parameter values). Therefore  $P(A | S, IC)$  represents the frequentist or propensity interpretation of probability since only aleatory uncertainty is present. When  $I$  is incomplete we don't know the full set of  $S$ -conditions and/or the models/parameters describing the process. In this case,  $P(A | S, I)$  has to account for both aleatory and epistemic uncertainty and it cannot be interpreted in a frequentist/propensity way, because each event will modify the variable  $I$  even in the presence of the same set of conditions  $S$ , and consequently  $P(A | S, I)$  will also change. The fact that  $P(A | S, I)$  cannot be interpreted in a frequentist way—but as a degree of belief—poses many problems to its verification with data, since the verification of a probability may be done only assuming that the expected frequency of events tends to the probability.

It might seem that, given enough time, any probability can be tested, but this is true only assuming that the dominant uncertainty is aleatory and represents an intrinsic feature of the random process. This is not true for subjective probabilities in general and UCERF in particular, because the occurrence of future events will affect expert opinion and consequently the

UCERF model. Practically speaking, UCERF's ability to accurately forecast the next 10–20 large earthquakes in California is of little interest, because UCERF is expected to be a “living model” that emphasizes the next single or next few large earthquakes. In other words, any probability may be put under test through a RELM-like procedure, but, for a subjective probability, this test would be meaningless.

A second point to consider is the meaning of “best model.” From a scientific point of view, the RELM experiment may provide a conclusive result based on five years of observations. But if a forecast model is to be implemented for practical application, it is necessary to realize that the end users—insurance companies or a civil protection agency, for example—may have a different metric to judge what is “best.” Therefore, the UCERF attitude is not only driven by science, but also by what may be loosely called “politics.” This is not inherently bad. For example, one of the main goals of decision-makers is to minimize possible *a posteriori* critiques if a forecast model fails. This is certainly easier if a large community of scientists is involved in building the model. Note that this is not unique to UCERF, but is also characteristic of other efforts such as the Intergovernmental Panel on Climate Change (Solomon *et al.* 2007).

One of the key distinctions between RELM and UCERF is each working group's average view of the extent to which small and large earthquakes are similar. Certainly small and large earthquakes are different in terms of the appropriate fault rupture representation; a simple point source is not a complete representation of a very large earthquake, but it often suffices for microseismicity. Nevertheless, there is little evidence from systematic studies that indicates that the distributions of small and large earthquakes are fundamentally different. This is an important problem, but one that neither RELM nor UCERF directly addresses.

At first blush, and perhaps even upon further reflection, it might seem that the overall RELM and UCERF approaches are diametrically opposed: RELM is strictly hands-off and completely objective hypothesis testing, whereas UCERF would not be possible without a certain amount of data massaging and the subjectivity that inevitably accompanies expert polling. We remind the reader that, at least in this sense, RELM and UCERF are representative of systematic earthquake predictability experimentation and long-term rupture forecasting, respectively.

One reason that it is so difficult to reconcile these different approaches is that, while the technical goals of RELM and UCERF are clear—the former evaluates five-year rate forecasts and the latter yields a long-term forecast—their impact and applicability beyond their practical tasks are not clear. Moreover, the assumption that fundamentally divides them—that large earthquakes either are, or are not, “like” smaller earthquakes—is often implicit; we hope that with this article we have made the distinction more clear. Adding to the confusion, UCERF does incorporate some elements from studies of smaller earthquakes. For instance, UCERF includes spatial smoothing of moderate historical earthquakes and a recurrence model whose usage is, at

least in part, justified by studies of repeating micro-earthquakes (Ellsworth *et al.* 1999). Nevertheless, many of the subjective aspects of the UCERF approach result from the assumption that understanding small earthquakes is not a sufficient condition for forecasting large earthquakes. On the other hand, RELM participants have not been so bold as to make outright claims that the results of their experiment are widely applicable, but neither have they said the opposite. Frankly, it is unknown if RELM results are applicable in other regions of space or time or magnitude (so that we might learn something about earthquakes outside California, in another future period, or of different magnitudes, respectively).

Any forecasting endeavor is characterized by a balance between scientific and practical components. These are not necessarily opposing forces but they do influence the approach for obtaining the best model. UCERF has a strong practical component, and RELM is essentially a purely scientific exercise. This difference, when coupled with each effort's assumption regarding the relationship between small and large earthquakes, justifies the dissimilar approaches of RELM and UCERF. ❧

## ACKNOWLEDGMENTS

We thank referee Max Werner for a thoughtful and thorough commentary, which led us to clarify many points throughout the article. We also thank Ned Field for providing useful comments. We thank Lauren Cooper for helpful editorial suggestions regarding an early draft.

## REFERENCES

- Aki, K. (1989). Ideal probabilistic earthquake prediction. *Tectonophysics* **169** (1–3), 197–198; doi:10.1016/0040-1951(89)90193-5.
- American Association for the Advancement of Science (AAAS) (1989). *Science for All Americans: A Project 2061 Report on Literacy Goals in Science, Mathematics and Technology*. Washington, DC: American Association for the Advancement of Science.
- Broad, W. J. (1979). Paul Feyerabend: Science and the anarchist. *Science* **206**, 534–537; doi:10.1126/science.386510.
- De Finetti, B. (1931). Probabilism. Reprinted in English in *Erkenntnis* **31**, 169–223 (1989).
- Ellsworth, W. L., M. V. Matthews, R. M. Nadeau, S. P. Nishenko, P. A. Reasenber, and R. W. Simpson (1999). *A Physically-Based Earthquake Recurrence Model for Estimation of Long-Term Earthquake Probabilities*. USGS Open-File Report 99-522; <http://geopubs.wr.usgs.gov/open-file/of99-522/>.
- Field, E. H. (2007). Overview of the working group for the development of regional earthquake likelihood models (RELM). *Seismological Research Letters* **78** (1), 7–16.
- Field, E. H., T. E. Dawson, K. R. Felzer, A. D. Frankel, V. Gupta, T. H. Jordan, T. Parsons, M. D. Petersen, R. S. Stein, R. J. Weldon III, and C. J. Wills (2009). Uniform California Earthquake Rupture Forecast, Version 2 (UCERF 2). *Bulletin of the Seismological Society of America* **99** (4), 2,053–2,107; doi: 10.1785/0120080049.
- Gauch, H. G. (2003). *Scientific Method in Practice*. New York: Cambridge University Press.
- Gillies, D. (2000). *Philosophical Theories of Probability*. New York: Routledge.
- Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2007). High-resolution time-independent grid-based forecast for  $M > 5$  earthquakes in



- California, *Seismological Research Letters* **78**, 78–86, doi:10.1785/gssrl.82.3.78.
- Howson, C., and P. Urbach (1989). *Scientific Reasoning: The Bayesian Approach*. LaSalle, IL: Open Court.
- Jackson, D. D. (1996). Hypothesis testing and earthquake prediction. *Proceedings of the National Academy of Sciences USA* **93**, 3,772–3,775.
- Jordan, T. H. (2006). Earthquake predictability, brick by brick. *Seismological Research Letters* **77**, 3–6.
- Jordan, T. H., and L. M. Jones (2010). Operational earthquake forecasting: Some thoughts on why and how. *Seismological Research Letters* **81** (4), 571–574; doi:10.1785/gssrl.82.3.571.
- Jordan, T. H., Y.-T. Chen, P. Gasparini, R. Madariaga, I. Main, W. Marzocchi, G. Papadopoulos, G. Sobolev, K. Yamaoka, and J. Zschau (2009). *Operational Earthquake Forecasting: State of Knowledge and Guidelines for Implementation. Findings and Recommendations of the International Commission on Earthquake Forecasting for Civil Protection*, Rome, Italy: Dipartimento della Protezione Civile.
- Kantorovich, L.V., and V. I. Keilis-Borok (1991). Earthquake prediction and decision making: Social, economic, legislative and civil defense domains. In *International Conference on Earthquake Prediction: State-of-the-Art, Scientific-Technical Contributions*, 586–593. Arpajon, France: Centre Sismologique Euro-Méditerranéen / European Mediterranean Seismological Centre.
- Kuhn, T. S. (1970). *The Structure of Scientific Revolutions*, 2nd ed. Chicago: University of Chicago Press.
- Lakatos, I., (1970). *Criticism and the Growth of Knowledge*. Cambridge University Press.
- Marzocchi, W., and A. M. Lombardi (2009). Real-time forecasting following a damaging earthquake. *Geophysical Research Letters* **36**, L21302; doi:10.1029/2009GL040233.
- Marzocchi, W., L. Sandri, and E. Boschi (2003). On the validation of earthquake-forecasting models: The case of pattern recognition algorithms. *Bulletin of the Seismological Society of America* **93**, 1,994–2,004.
- Marzocchi, W., and G. Woo (2007). Probabilistic eruption forecasting and the call for evacuation. *Geophysical Research Letters* **34**; doi:10.1029/2007GL031922.
- Marzocchi, W., and G. Woo (2009). Principles of volcanic risk metrics: Theory and the case study of Mt. Vesuvius and Campi Flegrei (Italy). *Journal of Geophysical Research* **114**, B03213; doi:10.1029/2008JB005908.
- Molchan, G. M. (1997). Earthquake prediction as a decision-making problem. *Pure and Applied Geophysics* **149**, 233–247.
- Oreskes, N., K. Shrader-Frechette, and K. Belitz (1994). Verification, validation, and confirmation of numerical models in the Earth sciences. *Science* **263**, 641–646.
- Popper, K. (1968). *The Logic of Scientific Discovery*. New York: Harper & Row.
- Salt, J. D. (2008). The seven habits of highly defective simulation projects. *Journal of Simulation* **2**, 155–161.
- Schorlemmer, D., and M. C. Gerstenberger (2007). RELM Testing Center. *Seismological Research Letters* **78** (1), 30–36.
- Schorlemmer, D., J. D. Zechar, M. J. Werner, E. H. Field, D. D. Jackson, and T. H. Jordan (2010). First results of the Regional Earthquake Likelihood Models experiment. *Pure and Applied Geophysics* **167** (8/9); doi:10.1007/s00024-010-0081-5.
- Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, eds. (2007). *Climate Change 2007: The Physical Science Basis: Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge and New York: Cambridge University Press.
- Theocharis, T., and M. Psimopoulos (1987). Where Science has gone wrong. *Nature* **329**, 595–598.
- van Stiphout, T., S. Wiemer, and W. Marzocchi (2010). Are short-term evacuations warranted? Case of the 2009 L'Aquila earthquake. *Geophysical Research Letters* **37**; doi:10.1029/2009GL042352.
- Woo, G. (2008). Probabilistic criteria for volcano evacuation decision. *Natural Hazards* **45**, 87–97.
- Zechar, J. D., M. C. Gerstenberger, and D. A. Rhoades (2010). Likelihood-based tests for evaluating space-rate-magnitude earthquake forecasts. *Bulletin of the Seismological Society of America* **100** (3); doi:10.1785/0120090192.
- Zechar, J. D., and T. H. Jordan (2008). Testing alarm-based earthquake predictions. *Geophysical Journal International* **172**, 715–724; doi:10.1111/j.1365-246X.2007.03676.x.
- Zechar, J. D., D. Schorlemmer, M. Liukis, J. Yu, F. Euchner, P. J. Maechling, and T. H. Jordan (2010). The Collaboratory for the Study of Earthquake Predictability perspective on computational earthquake science. *Concurrency and Computation: Practice and Experience* **22**, 1,836–1,847; doi:10.1002/cpe.1519.

*Istituto Nazionale di Geofisica e Vulcanologia*  
 Via di Vigna Murata 605  
 000143 Rome, Italy  
 warner.marzocchi@ingv.it  
 (W. M.)