OPINION

Bayes and BOGSAT: Issues in When and How to Revise Earthquake Hazard Maps

INTRODUCTION

Recent large earthquakes that caused ground shaking larger than anticipated have generated interest in how to improve earthquake hazard mapping. Issues under discussion include how to evaluate maps' performance, how to assess their uncertainties, how to make better maps, and how to best use maps given their limitations.

An important question is what to do after an earthquake yielding shaking larger than anticipated. Hazard mappers have two choices. One is to regard the high shaking as a low-probability event allowed by the map, which used estimates of the probability of future earthquakes and the resulting shaking to predict the maximum shaking expected with a certain probability over a given time (Hanks *et al.*, 2012; Frankel, 2013). The usual choice, however, is to accept that high shaking was not simply a low-probability event consistent with the map, and revise the map to show increased hazard in the heavily shaken area (Fig. 1).

Whether and how much to revise a map is complicated, because a new map that better describes the past may or may not better predict the future. For example, increasing the predicted hazard after an earthquake on a fault will make better predictions if the average recurrence time is short compared to the map's time window but will overpredict future shaking if the average recurrence time is much longer than the map's time window.

BAYES' RULE

For insight into whether and how to remake a hazard map, imagine tossing a coin, which comes up heads four times in a row. How likely do you think it is to come up heads on the next toss? You started off assuming that the coin is fair—equally likely to land heads or tails. Should you change that assumption?

Either choice runs a risk. If the coin is severely biased, staying with the assumption that it is fair will continue to yield poor predictions. However, if the coin is fair and the four heads were just a low-probability event, changing to the assumption that the coin is biased does a better job of describing what happened in the past but will make your prediction worse.

Your choice would depend on how confident you were in your assumption, prior to the tosses, that the coin was fair. If you were confident that the coin was fair, you would not change your model and continue to assume that a head or tail is equally likely. However, if you got the coin at a magic show, your confidence that it is fair would be lower, and you would be more apt to change your model to one predicting a head more likely than a tail.

A statistical approach that combines preconceptions with observations to decide how to update forecasts as additional information becomes available uses Bayes' Rule (Rice, 2007). In this formulation

$$\begin{aligned}&\text{revised or posterior probability}\\&\quad\propto \text{likelihood of observations given the prior model}\\&\qquad\times \text{prior probability}\end{aligned}$$

omitting a normalization. This starts by assuming an initial or prior probability model based on information available prior to the additional observations, calculating how likely the observations were given that model, and using the product as the revised or posterior probability model to account for the additional observations.

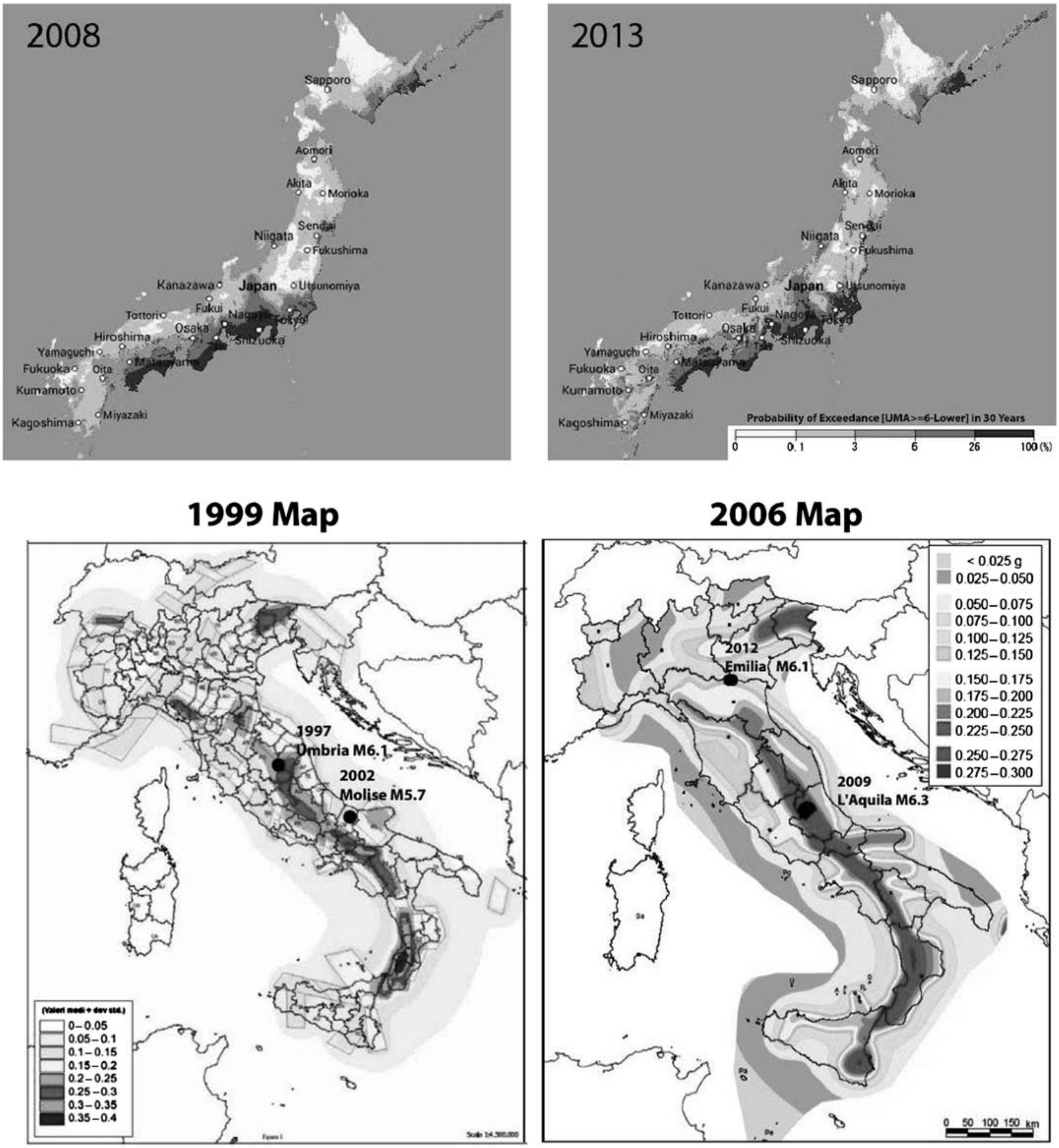
We can describe a coin's probability of landing heads by a parameter from 0 (always tails) to 1 (always heads) and represent our beliefs about the parameter by a probability distribution. If, prior to observing the four heads, we are confident the coin is fair or nearly fair, our prior probability distribution is tightly clustered around 0.5 (although to allow surprises, it assigns nonzero probability throughout the interval). If we think the coin may be biased, our prior distribution would have a much larger spread and might be skewed toward 0 or 1.

After some tosses, the revised model depends on both the observations and the prior model. If we had high confidence that the coin was fair, a few low-probability observations would not change it much. However, if we had little confidence in the prior model, these low-probability observations change it a lot.

In the Bayesian approach, probability represents our belief in how a system works based on the information we have. This probability is subjective, because given the little information we know about the coin, we have no way to know the actual probability of a head on the next toss. Once we have chosen a model, we can calculate this probability precisely. However, because this probability assumes that the model is true, it also is subjective and subject to revision after the next toss.

This view differs from the frequentist view in which an event's probability is the frequency in which it occurs in a large number of trials. After a thousand independent tosses under standard conditions, the fraction of heads would be a good estimate of the probability of a head on the next toss. However, because we only have four tosses, we factor in our preconcep-

 doi: 10.1785/0220140215



▲ **Figure 1.** (top) Japanese seismic-hazard maps before and after the 2011 Tohoku earthquake. The predicted hazard has been increased both along the east coast, where the 2011 earthquake occurred, and on the west coast. (http://www.j-shis.bosai.go.jp/map/?lang=en; last accessed December 2014.) (bottom) Comparison of successive Italian hazard maps (Stein *et al.*, 2013). The 1999 map was updated to reflect the 2002 Molise earthquake, and the 2006 map will likely be updated after the 2012 Emilia earthquake.

tions rather than automatically assume that four heads prove that the probability of one in the next toss is near 1.

Although the Bayesian approach requires assuming a prior probability distribution, this assumption's effect is reduced as more data become available, provided the prior distribution does not assign zero probability to parameter values that include the true state of nature. After enough observations, the posterior distribution does not depend on the assumed prior distribution.

EARTHQUAKE PROBABILITIES

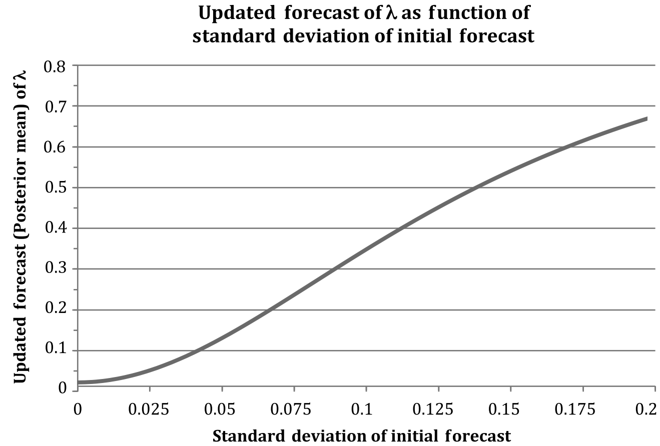
Seismologists often approach estimating earthquake hazards in the spirit of Bayes' Rule, because this involves assuming probability models based on limited data and then using new data to improve them (Marzocchi and Jordan, 2014). To see this, consider a simple example in which we assume that the probability of a large earthquake on a fault is described by a Poisson process with parameter $\lambda = 1/T$, corresponding to an average return time of T years. Following Campbell (1982), we represent our uncertainty about λ using a gamma distribution with mean μ and standard deviation σ as our prior probability distribution. If an earthquake occurs only one year after the last, the prior distribution is updated to the posterior distribution, and the prior mean μ updates to the posterior mean $\mu' = \mu(1 + \sigma^2/\mu^2)/(1 + \sigma^2/\mu)$ (Rice, 2007).

Consider μ to be specified as 0.02, that is $T = 50$ yrs. If we are highly confident about λ when the forecast is made, σ is small, so the posterior mean μ' and prior mean μ are close. We treat the new observation that did not fit the model well as a rare event that does not change our preconception much. However, if we were uncertain that λ would be near the prior mean μ, σ is large so the new observation changes our view, making the posterior mean very different (larger) than the prior mean.

Figure 2 shows how the updated forecast, described by the posterior mean, increasingly differs from the initial forecast (prior mean) when the uncertainty in the prior distribution is larger. The less confidence we have in the prior model, the more a new datum can change it.

This example is useful because inferring earthquake probabilities, which are crucial inputs for hazard mapping, is very difficult given the poorly understood faulting process and the limitations of the earthquake record (Savage, 1994; Parsons, 2008). It is unclear whether to assume earthquake recurrence is described by a Poisson process with no memory, so the probability is constant with time, or by time-dependent models based on an earthquake cycle in which the probability is small shortly after the past one, and then increases. Numerical simulation shows that these two are difficult to distinguish even in a simple case (Stein and Stein, 2013a). Moreover, using a time-dependent model requires choosing many parameters that are poorly constrained by the available earthquake history.

From a statistical view, Stark and Freedman (2003) concluded that earthquake probability estimates are "shaky." In their view, "the interpretation that probability is a property of a model and has meaning for the world only by analogy seems the most appropriate.... The problem in earthquake forecasts is that the models have not been tested against relevant data. Indeed, the models cannot be tested on a human time scale, so there is little reason to believe the probability estimate." Savage (1991) concluded that earthquake probability estimates for California are "virtually meaningless" and that it would be meaningful only to quote broad ranges, such as low (<10%), intermediate (10%–90%), or high (>90%). In other words, it seems reasonable to say that earthquakes of a given size are more likely on some faults than others, but quantifying this involves large uncertainty.



▲ **Figure 2.** Sensitivity of updated forecast of λ, initially assumed to equal 0.02, to assumed prior uncertainty. The lower our confidence in the initial forecast, the more the new datum changes it.

HAZARD MAPS

The earthquake probability example illustrates the challenge for hazard maps: choosing hundreds or thousands of parameters to predict the answers to four questions over periods of 500–2500 yr: Where will large earthquakes occur? When will they occur? How large will they be? How strong will their shaking be?

Some of the parameters required are reasonably well known, some are somewhat known, some are essentially unknown, and some may be unknowable (e.g., Stein *et al.*, 2012). As a result, mappers combine data and models with their sense of how the earth works. Stark and Freedman (2003) note that this involves "geological mapping, geodetic mapping, viscoelastic loading calculations, paleoseismic observations, extrapolating rules of thumb across geography and magnitude, simulation, and many appeals to expert opinion. Philosophical difficulties aside, the numerical probability values seem rather arbitrary."

Such models, which involve subjective assessments and choices among many poorly known or unknown parameters, are sometimes termed BOGSATs, from "Bunch Of Guys Sitting Around a Table" (Kurowicka and Cooke, 2006). Not surprisingly, sometimes the resulting maps do well at predicting what occurs in future earthquakes, and sometimes they do poorly. However, at this point, there is no way to avoid BOGSAT. Although some parameters could be better estimated, and knowledge of some will improve as new data and models become available, major uncertainties seem likely to remain (Stein and Friedrich, 2014).

Nonetheless, despite their large uncertainties, hazard maps have some useful information. From a mitigation policy standpoint, inaccurate hazard (and loss) estimates are still useful unless they involve gross misestimates (Stein and Stein, 2013b). For example, a highway department would likely use its limited funds to preferentially strengthen bridges in predicted high-hazard areas.

In our view, one should consider the BOGSAT process from a Bayesian perspective. This recognizes that the predicted hazard reflects mapmakers' view of the world based on their assessment of diverse data and models, and that when and how maps are revised once new data become available depends on the mapmakers' preconceptions. Because this is the case, how can it be done better?

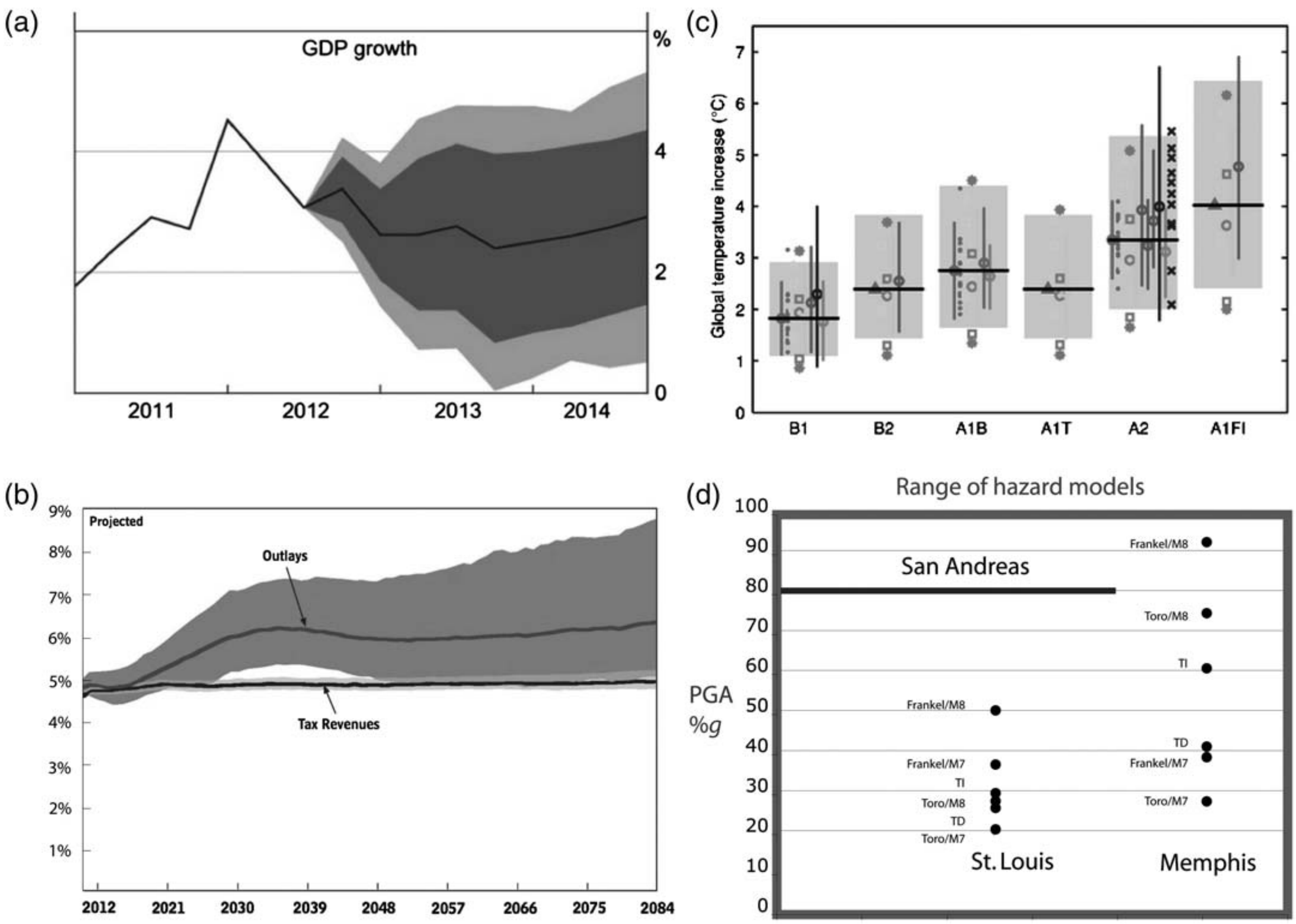
At a fundamental level, we need to learn more about when and how revising maps makes them better or worse predictors of the future. In some cases revisions should make the map work better, and in others, worse. In particular, raising the predicted hazard where a large earthquake recently occurred may improve the match of the model to past data (though this is rarely quantified using a previously defined metric) but degrade its fit to future events.

On a working level, we suggest several changes to current procedures.

First, maps should specify what they seek to predict and how their performance should be measured. Various metrics can be used, so users can know what the mappers' goals are and be able at later time to assess how well the map met them. For example, how well did the map perform compared to one that assumed a much smoother variation in the predicted hazard (Geller, 2011)?

Second, hazard map documentation should list the parameters used and estimates of their uncertainties. Often much of this information is available in the documentation (e.g., Field *et al.*, 2008). In particular, weights assigned to logic tree branches are a discretized version of the prior probability density function assumed for that parameter. It would be useful to list model assumptions in a consistent form to make changes between successive maps easier to identify and discuss.

Third, estimates of the expected uncertainty in the predicted hazard should be presented and explained. Forecasts with significant economic and policy implications typically present



▲ **Figure 3.** Presenting forecast uncertainties. (a) Forecast of Australian Gross Domestic Product (GDP) growth. Uncertainty bounds are 70% and 90% (Reserve Bank of Australia, 2013). (b) Forecast of U.S. Social Security expenditure as percentage of GDP (Congressional Budget Office, 2010) (c) Comparison of the rise in global temperature by the year 2099 predicted by various climate models. For various carbon emissions scenarios, for example, B1, the vertical band shows the predicted warming (Intergovernmental Panel on Climate Change [IPCC], 2007). (d) Comparison of earthquake hazard, described as peak ground acceleration as a percentage of the acceleration of gravity expected with 2% risk in 50 yr, predicted by various assumptions for two sites in the central United States (Stein *et al.*, 2012).

uncertainties (Fig. 3). Although forecasts sometimes miss their targets (Stein and Stein, 2014), uncertainty estimates are still useful. This would involve generating hazard curves and maps for different parameter values within their assumed uncertainties. The resulting range of estimates could be presented via uncertainty maps or tabulations at sites. These uncertainties could be factored in policy making, as is done for other forecasts.

Fourth, changes in parameter values between successive maps should be listed and explained. Some will likely reflect what happened in earthquakes after the map was made whereas others will reflect data not used in the earlier map, because they were not recognized, not appreciated, or unavailable. The criteria used to decide when parameters were changed should be defined (Ramsey, 1926).

Deciding when and how to revise hazard maps would combine Bayes and BOGSAT. Conceptually, changing parameters would reflect Bayes' Rule, because those previously thought to have greater uncertainty would be most easily changed by new data or ideas. Operationally, because most parameters are estimated via a combination of data, models, and assumptions, the actual values would come from BOGSAT rather than explicit calculation. Even so, the Bayesian approach would add value because it is systematic. If BOGSAT leads to big changes in the map, one can assess what that implies about prior confidence in the forecasts.

This approach would give users information about the uncertainties to make better decisions. Meteorologists (Hirschberg *et al.*, 2011) have adopted a goal of "routinely providing the nation with comprehensive, skillful, reliable, sharp, and useful information about the uncertainty of hydrometeorological forecasts." Although seismologists have a tougher challenge and a longer way to go, we should try to do the same.

ACKNOWLEDGMENTS

Stein thanks the U.S. Geological Survey John Wesley Powell Center for Analysis and Synthesis for hosting a working group under auspices of the Global Earthquake Model project, whose stimulating discussions inspired this work. Spencer thanks the Institute for Policy Research for supporting his research. We also thank Sandy Zabell for helpful comments.

REFERENCES

Campbell, K. W. (1982). Bayesian analysis of extreme earthquake occurrences, *Bull. Seismol. Soc. Am.* **72,** 1689–1705.

Congressional Budget Office, Washington, D.C. (2010). *Long-Term Projections for Social Security.*

Field, E., 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 II, and C. J. Wills (2008). The uniform California earthquake rupture forecast, Version 2, *U.S. Geol. Surv. Open-File Rept. 2007-1437.*

Frankel, A. (2013). Comment on "Why earthquake hazard maps often fail and what to do about it," by S. Stein, R. J. Geller, and M. Liu, *Tectonophysics* **592,** 200–206.

Geller, R. J. (2011). Shake-up time for Japanese seismology, *Nature* **472,** 407–409.

Hanks, T. C., G. C. Beroza, and S. Toda (2012). Have recent earthquakes exposed flaws in or misunderstandings of probabilistic seismic hazard analysis? *Seismol. Res. Lett.* **83,** 759–764.

Hirschberg, P., E. Abrams, A. Bleistein, W. Bua, L. D. Monache, T. W. Dulong, J. E. Gaynor, B. Glahn, T. M. Hamill, J. A. Hansen, D. C. Hilderbrand, R. N. Hoffman, B. H. Morrow, B. Philips, J. Sokich, and N. Stuart (2011). An implementation plan for generating and communicating forecast uncertainty information, *Bull. Am. Meteorol. Soc.* **92,** 1651–1666.

Intergovernmental Panel on Climate Change (IPCC) (2007). *Climate Change*, Cambridge Univ. Press, New York.

Kurowicka, D., and R. M. Cooke (2006). *Uncertainty Analysis with High Dimensional Dependence Modeling*, Wiley, Hoboken, New Jersey.

Marzocchi, W., and T. H. Jordan (2014). Testing for ontological errors in probabilistic forecasting models of natural systems, *Proc. Natl. Acad. Sci. Unit. States Am.* **111,** 11973–11978.

Parsons, T. (2008). Earthquake recurrence on the south Hayward fault is most consistent with a time dependent, renewal process, *Geophys. Res. Lett.* **35,** doi: 10.1029/2008GL035887.

Ramsey, F. P. (1926). Truth and probability. in *Foundations of Mathematics.*, R. B. Braithwaite (Editor), Littlefield, Adams & Co, Totowa, New Jersey, 156–198.

Reserve Bank of Australia, Sydney, Australia (2013). *Statement on Monetary Policy.*

Rice, J. A. (2007). *Mathematical Statistics and Data Analysis*, Duxbury, India.

Savage, J. C. (1991). Criticism of some forecasts of the national earthquake prediction council, *Bull. Seismol. Soc. Am.* **81,** 862–881.

Savage, J. C. (1994). Empirical earthquake probabilities from observed recurrence intervals, *Bull. Seismol. Soc. Am.* **84,** 219–221.

Stark, P. B., and D. A. Freedman (2003). What is the chance of an earthquake? in *Earthquake Science and Seismic Risk Reduction*, F. Mulargia and R. J. Geller (Editors), 201–213. Kluwer, Dordrecht, The Netherlands.

Stein, S., and A. Friedrich (2014). How much can we clear the crystal ball? *Astron. Geophys.* **55,** 2.11–2.17.

Stein, S., and J. L. Stein (2013a). Shallow versus deep uncertainties in natural hazard assessments, *Eos Trans AGU* **94,** no. 4, 133–140.

Stein, S., and J. L. Stein (2013b). How good do natural hazard assessments need to be? *GSA Today* **23,** nos. 4/5, 60–61.

Stein, S., and J. L. Stein (2014). *Playing Against Nature: Integrating Science and Economics to Mitigate Natural Hazards in an Uncertain World*, Wiley/AGU, Hoboken, New Jersey.

Stein, S., R. J. Geller, and M. Liu (2012). Why earthquake hazard maps often fail and what to do about it, *Tectonophysics* **562/563,** 1–25.

Stein, S., R. J. Geller, and M. Liu (2013). Reply to comment by Arthur Frankel on "Why Earthquake Hazard Maps Often Fail and What to do About It," *Tectonophysics* **592,** 207–209.

Seth Stein
Department of Earth and Planetary Sciences
and Institute for Policy Research
Northwestern University
Evanston, Illinois 60208 U.S.A.
seth@earth.northwestern.edu

Bruce D. Spencer
Department of Statistics and Institute for Policy Research
Northwestern University
Evanston, Illinois 60208 U.S.A.

Edward Brooks
Department of Earth and Planetary Sciences
Northwestern University
Evanston, Illinois 60208 U.S.A.

Published Online 31 December 2014