Probability in PSHA: Reply to "Comment on 'PSHA Validated by Quasi-Observational Means' by Z. Wang"

by R. M. W. Musson

One of the attractions of the Monte Carlo simulation approach to seismic hazard is its conceptual simplicity and directness. What a site operator wants to know is what type of earthquake shaking his plant might possibly have to endure in its lifetime. In response, the seismologist takes all the available information on regional seismogenesis, distils it into a numerical model, and then uses that model to investigate all the possible earthquake ground motions that might occur at the site. Given a sufficient number of simulations, every possible outcome will be sampled, and the most likely outcomes will be observed more often than the less likely outcomes.

Wang (2012), in commenting on the description of this method in Musson (2012a), states that randomization of an earthquake catalog is an insufficient approach to assessing seismic hazard, as it ignores the tectonic basis for seismicity, and cannot include possible future events not represented in the historical catalog, such as the 2011 Tohoku event. Such a catalog randomization approach has been suggested in the past, for instance by Ebel and Kafka (1999). But this is not what is described in Musson (2012a). Simulated catalogs used for hazard analysis are not randomizations of the historical catalog, but realizations of possible future events based on what is known about regional seismicity. Any individual simulation may have a different number of events from the historical catalog, but it will be consistent with what is known about local seismotectonics. So it may include larger earthquakes than those observed historically, up to a limit inferred from the tectonic situation.

Indeed, a catalog randomization approach would not be able to reproduce Cornell-type results as shown in Musson (2012a), since the model used as an example in that paper includes a fault which is assessed on the basis that it may be active, but which is not manifested by any earthquakes in the catalog. Randomizing the historical catalog would therefore result in completely missing potential earthquakes on this fault.

So tectonics is, in fact, central to the simulation method. Musson *et al.* (2009) suggests a framework for building the necessary model, starting from the gross kinematic situation (which plates are involved, and how they are moving), refining to seismotectonic zones where generally similar processes are operating, and then finally refining down to the level of individual seismic sources. This approach is equally applicable for any seismic source model, whether it will be processed with the Cornell method or through a stochastic approach.

In the process, it may be expedient to make some simplifications. For instance, the adoption of a memory-less Poissonian model for occurrence is a convenience. It is not necessary to adopt it. But it turns out that applying more realistic models with memory usually makes no difference to the actual results (e.g., Wong *et al.*, 2008), and there are good mathematical reasons why this is so (Khintchine, 1960).

Wang's (2012) objection to the simulation method is thus founded on a misunderstanding of what the method actually entails. His objection to the Cornell method of probabilistic seismic hazard assessment (PSHA) is the claim that Cornell's (1968) paper is based on a mathematical error, that of equating a dimensionless quantity (probability) with a dimensioned one (frequency). In fact, Cornell (1968) is simply following the mathematics of an ordinary renewal process (see for instance, Cox and Miller [1965], section 9.2). Probability in such cases represents a notional proportion of years; one can write

$$P = \mathrm{Ye}/\mathrm{Yt},$$

where Ye is the number of years with exceedances and Yt is the total number of years, in which case, clearly the inverse of the probability is a number of years.

This does bring up an interesting topic. Probability is often defined as the limit of the relative frequency of an event in a large number of trials. The concept of probability, however, actually embraces two different concepts, and this duality of probability goes back historically to writers like Condorcet in the nineteenth century, and even earlier (Hacking, 1975; Grimmett and Stirzaker, 2001).

The first aspect of probability is frequency-based. Thus, if one draws (with replacement) 20 white balls in 100 draws from an urn containing balls of different colors, one can say that the probability of drawing a white ball on the next draw is 20%. This is frequency probability. The second aspect is that one can meaningfully say that the probability of the home team

doi: 10.1785/0220120036

winning the football game next Saturday is 60%. This is belief-based or subjective probability. There is no frequency involved, as the game will only be played once. Also, the statement that the probability of a home win is 60% is not falsifiable whether the team wins or loses. Its meaning resides in the expression that 6–4 represents fair betting odds. There is a tendency in recent seismology to have a horror of anything that has the word "subjective" attached, but as de Finetti (1972, 1974) showed, belief-based probability can still be mathematically rigorous.

There exist people who are frequency dogmatists, who insist that frequency-based probability is the only correct type, and also belief dogmatists, who insist that only subjective probability is valid. The philosophical consensus, according to Howson and Urbach (2006), is that both dogmatisms are unhelpful, and one should allow that both types of probability are valid.

In the past, the debate on probability in PSHA has tended to be between frequency dogmatists and those of a more liberal outlook. An area of dispute has been the meaning of weights in a PSHA logic tree (Abrahamson and Bommer, 2005; Musson, 2005, 2012b; Scherbaum and Kuehn, 2011). If several options are given different weights, does this represent the probability that each option is the best one, or merely one's degree of certainty? This has important implications. If the weights on a logic tree are probabilities, then various fractile hazard curves are "probabilities of probabilities" which need to be collapsed (Musson, 2005). In this case, hazard maps that present the median hazard do not actually show the ground motion that has the probability claimed by the study. The median hazard for, say, a 450-yr return period is actually the expected (i.e., mean) hazard for some other return period.

Justification of hazard fractiles requires one to be a frequency dogmatist, but this runs the risk of PSHA not being probabilistic at all, in that the seismicity over the next 50 yr is not repeatable, or to put it in technical terms, does not form part of a collective, as defined by von Mises (1957). One could argue, however, that 100,000 simulations of the next 50 yr do indeed constitute a collective, and thus the validity of the Monte Carlo approach to PSHA could be taken as confirmation of the probabilistic nature of PSHA for the strict frequentist.

That it is possible to exactly reproduce Cornell-type PSHA results using a simulation approach, as demonstrated in Musson (2012a), is remarkable, given that in the past different methodological approaches to seismic hazard have tended to give differing results. As a general principle, to solve a problem by two independent methods and arrive at the same result is a strong confirmation that the results are correct. To argue that the results are still in error requires one to strike at something so fundamental that it is common to both as, in this case, is the definition of probability. Wang's (2012) objection to PSHA can thus be read as endorsing a belief dogmatist approach (i.e., denying that probability can be based on frequency), in contrast to the frequency dogmatism more often encountered. One has to reply, with Howson and Urbach (2006), that probability has always had a dual nature, and extremist positions that deny one or the other element tend to be unproductive.

In a way, it does not matter. Returning to the very opening of the present paper, from the results of the simulations in a Monte Carlo hazard study, one can present to the engineer a very complete picture of what the future could hold. One can say that given 100,000 projections of future earthquake activity, only in 10% of them did the site experience ground motion above such-and-such a value, and some higher value was only exceeded 1% of the time. One can show the statistics for the maximum ground motion experienced during the lifetime of a plant in each simulation, and present the mean, the median and the 84th percentile of these values. One can even find the absolute maximum shaking value that any of the simulations produced.

Whether one calls this probability, or frequency, or anything else is almost irrelevant—these statistics are clearly useful to anyone planning how to address earthquake ground motions at a given site. They represent the statistical range of possible seismic futures that could befall the plant, and provide an informed basis for rational decision making.

ACKNOWLEDGMENTS

This paper is published with the permission of the Executive Director of BGS (NERC).

REFERENCES

Abrahamson, N. A., and J. J. Bommer (2005). Probability and uncertainty in seismic hazard analysis, *Earthquake Spectra* **21,** no. 2, 603–608.

Cornell, C. A. (1968). Engineering seismic risk analysis, *Bull. Seismol. Soc. Am.* **58,** no. 5, 1583–1606.

Cox, D. R., and H. D. Miller (1965). *The Theory of Stochastic Processes*, Methuen, London.

de Finetti, B. (1972). *Probability, Induction and Statistics: The Art of Guessing*, Wiley, London.

de Finetti, B. (1974). *Theory of Probability: A Critical Introductory Treatment*, Wiley, London.

Ebel, J. E., and A. L. Kafka (1999). A Monte Carlo approach to seismic hazard analysis, *Bull. Seismol. Soc. Am.* **89,** 854–866.

Grimmett, G. R., and D. R. Stirzaker (2001). *Probability and Random Processes*, Oxford University Press, Oxford.

Hacking, I. (1975). *The Emergence of Probability*, Cambridge University Press, Cambridge.

Howson, C., and P. Urbach (2006). *Scientific Reasoning: The Bayesian Approach*, Open Court, Chicago.

Khintchine, A. J. (1960). *Mathematical Models in the Theory of Queuing*, Griffin, London.

Musson, R. M. W. (2005). Against fractiles, *Earthquake Spectra* **21,** no. 3, 887–891.

Musson, R. M. W. (2012a). PSHA validated by quasi-observational means, *Seismol. Res. Lett.* **23,** 130–134.

Musson, R. M. W. (2012b). On the nature of logic trees in probabilistic seismic hazard assessment, *Earthquake Spectra* (in press).

Musson, R. M. W., S. Sellami, and W. Brüstle (2009). Preparing a seismic hazard model for Switzerland: The view from PEGASOS Expert Group 3 (EG1c), *Swiss J. Geol.* **102,** 107–120.

Scherbaum, F., and N. M. Kuehn (2011). Logic tree branch weights and probabilities: Summing up to one is not enough, *Earthquake Spectra* **27,** 1237–1251.
von Mises, R. (1957). *Probability, Statistics and Truth*, Allen and Unwin, London.
Wang, Z. (2012). Comment on "PSHA validated by quasi-observational means" by R. M. W. Musson, *Seismol. Res. Lett.* **83,** no. 4, 714–716.
Wong, I., P. Thomas, and R. Youngs (2008). Time-Dependent Seismic Hazard in the San Francisco Bay Region, California, Paper 07-0110, in *Proceedings of the 14th World Conference on Earthquake Engineering*, Beijing, China.

R. M. W. Musson
British Geological Survey
West Mains Road
Edinburgh EH9 3L0041, UK
rmwm@bgs.ac.uk