A Review of the Seismic Hazard Model MPS19.S

Philip B. Stark
Department of Statistics
University of California
Berkeley, CA 94720-3860
stark@stat.berkeley.edu

12 July 2022

Executive Summary

MPS19.S relies on probabilistic seismic hazard analysis (PSHA), which has been debunked in the geophysical literature. PSHA is based on a metaphor, not on physics. It involves (i) assuming that earthquakes occur randomly, (ii) conflating past frequencies with future probabilities, and (iii) conflating ignorance with randomness (epistemic versus aleatory uncertainty). It ignores the possibility that foreshocks and aftershocks can be destructive and deadly. It is not a sound basis for setting building codes or insurance rates, protecting antiquities, prioritizing risk mitigation, or otherwise protecting the public.

Table of Contents

Introduction

I was asked to review *The seismic hazard model MPS19.S*, Edited by Carlo Meletti et al., dated April 6, 2022. I hereby request that this document be made public, with my name as author.

I do not think MPS19.S is a sound basis for policy decisions affecting public welfare, such as setting building codes and other public safety issues, protecting antiquities, setting insurance rates, or allocating funding for risk mitigation. The basis of MPS19.S is probabilistic seismic hazard analysis, PSHA. PSHA lacks any meaningful empirical, geophysical, or statistical basis, as I shall explain.

For background on the problems with PSHA, see Mulargia, Stark, and Geller (2017), which I understand is the most frequently downloaded article from *Physics of the Earth and Planetary Interiors* since it was published.

The statistical methodology and assumptions in MPS19.S are flawed. Some of the analysis assumes that (declustered) seismicity has a Poisson distribution, which it does not (Luen and Stark, 2012), and the analysis ignores putative foreshocks and aftershocks by declustering input catalogs—even though foreshocks and aftershocks can be destructive, as the report acknowledges.

Below, I focus on some of the foundational issues raised in Mulargia et al. (2017) rather than on such technical issues. As George Box (1976) wrote, "it is inappropriate to be concerned with mice when there are tigers abroad."

What is probability?

PSHA purports to estimate exceedance probabilities. What is probability? Probability has an axiomatic aspect and a philosophical aspect. Kolmogorov's axioms, the mathematical basis of modern probability, are just that: mathematics. *Theories of probability* provide the philosophical glue to connect the mathematics to the real world, allowing us to interpret probability statements.[1]

As discussed by Stark and Freedman (2010), none of the standard theories of probability makes sense in the context of assigning probabilities to future earthquakes, nor to future ground motion, including "exceedance probabilities."

In particular, the *subjective* or (*neo-*)*Bayesian theory* implicitly adopted by the MPS19.S authors is incoherent in this context. The subjective theory of probability defines probability in terms of degree of belief. According to the subjective theory, what it means to say "the probability that a coin lands heads is 50%" is that the speaker believes with equal strength that it will land heads as he or she believes that it will land tails. In this interpretation, probability measures the state of mind of the person making the probability statement. This interpretation changes the subject from geophysics to psychology. The situation is complicated further by the fact that people are not very good judges of what is going to happen, as discussed below.

LeCam (1977, pp. 134–135) offers the following observations:

> (1) The neo-Bayesian theory makes no difference between 'experiences' and 'experiments'.
>
> (2) It confuses 'theories' about nature with 'facts', and makes no provision for the construction of models.
>
> …

[1] For a more technical discussion, see Freedman (2010) and Stark and Freedman (2010); LeCam (1977). For historical discussions, see Desrosiéres (1998) and Diaconis and Skyrms (2018).

> (4) It does not provide a mathematical formalism in which one person can communicate to another the reasons for his opinions or decisions. Neither does it provide an adequate vehicle for transmission of information. …
>
> (5) The theory blends in the same barrel all forms of uncertainty and treats them alike.
>
> …
>
> In summary, the Bayesian theory is wonderfully attractive but grossly oversimplified. It should be used with the same respect and the same precautions as the kinetic theory of perfect monoatomic gases.

Another interpretation of probability implicit in PSHA is that of probability as *metaphor*, that earthquakes (and ground acceleration) occur 'as if' the Earth were a casino game. Taleb (2007, pp. 127–129) discusses *the ludic fallacy* of treating all uncertainty as if it arose from casino games:

> "The casino is the only human venture I know where the probabilities are known, Gaussian (i.e., bell-curve), and almost computable." … [W]e automatically, spontaneously associate chance with these Platonified games. … Those who spend too much time with their noses glued to maps will tend to mistake the map for the territory. … Probability is a liberal art; it is a child of skepticism, not a tool for people with calculators on their belts to satisfy their desire to produce fancy calculations and certainties. Before Western thinking drowned in its "scientific" mentality, … people prompted their brain to think—not compute.

Uncertainty and Probability

The MPS19.S authors use the word 'probability' to describe anything uncertain. A common taxonomy, which MPS19.S uses, classifies uncertainties as *aleatory* or *epistemic*. Aleatory uncertainty results from the play of chance mechanisms—the luck of the draw. Epistemic uncertainty results from ignorance. Epistemic uncertainty is 'stuff we don't know' but in principle could learn.

Canonical examples of aleatory uncertainty include coin tosses, die rolls, lotteries, radioactive decay, some kinds of measurement error, and the like. Under some circumstances, such things do behave (approximately) as if random—but generally not perfectly so. Canonical examples of epistemic uncertainty include ignorance of the physical laws that govern a system or ignorance of the values of parameters in a system.

Imagine a biased coin that has an unknown chance p of landing heads. Ignorance of the chance of heads is epistemic uncertainty. But even if we knew the chance of heads, we would not know the outcome of the next toss: it would still have aleatory uncertainty.

The standard way to combine aleatory and epistemic uncertainties involves using subjective (aka Bayesian or neo-Bayesian) prior probability to represent epistemic uncertainty. In effect, this puts individual beliefs on a par with unbiased physical measurements that have known uncertainties. We know that the chance of heads must be between 0 and 1, but we do not know more than that.

Attempting to combine aleatory and epistemic uncertainties by considering both to be 'probabilities' that satisfy Kolmogorov's axioms amounts to claiming that there are two equivalent ways to tell how much something weighs: I could weigh it on an actual physical scale or I could think hard about how much it weighs. The two are on a par: as LeCam wrote, the method used in MPS19.S does not distinguish between "experiment" and "experience." It implicitly claims that thinking hard about the question produces an unbiased measurement. Moreover, it implies that I know the accuracy of my internal 'measurement' from careful introspection. Hence, I can combine the two sources of uncertainty as if they are independent measurements of the same thing, both made by unbiased instruments.[2] Unfortunately, this does not work.

Psychology, psychophysics, and psychometrics have shown empirically that people are bad at making even rough qualitative estimates, and that quantitative estimates are usually biased. Moreover, the bias can be manipulated through processes such as *anchoring* and *priming*, as described in the seminal work of Tversky and Kahneman (1975). Anchoring, the tendency to stick close to an initial estimate, no matter how that estimate was derived, doesn't just affect individuals—it affects entire disciplines. The Millikan oil drop experiment to measure the charge of an electron (Millikan, 1913) is an example: Millikan's value was too low, supposedly because he used an incorrect value for the viscosity of air. It took about 60 years for new estimates to "drift up" towards the currently accepted value, which is about 0.8% higher (a small difference, but considerably larger than the error bars). Other examples include measurements of the speed of light and the amount of iron in spinach.[3] In these examples and others, somebody erred and it took a very long time for a discipline to correct the error because subsequent work did not stray too far from the previous estimate—perhaps because the first estimate made them doubt the accuracy of results that were far from it.

Tversky and Kahneman also showed that people are poor judges of probability and have strong biases from anchoring, *representativeness,* and *availability*, which in turn depends on the retrievability of instances—that is, on the vagaries of human memory. Their work also shows that probability judgments are insensitive to prior probabilities and to predictability, and that people ignore the regression to the mean effect, even those who have had formal training in probability. (Regression to the mean is the mathematical phenomenon that in a sequence of independent realizations of a random variable, particularly extreme values are likely to be followed by values that are closer to the mean.)

People cannot even accurately judge how much an object weighs with the object in their hands. The direct physical tactile measurement is biased by the density and shape of the object—and even its color.[4] The notion that one could just think hard about seismic risk and thereby come up with a meaningful estimate and uncertainty for that estimate is preposterous. Wrapping the estimate with computer simulations that are not grounded in physics distracts, rather than illuminates.

---

[2] When practitioners analyse complex systems such as earthquakes, the same observations they use as data are often also the basis of their beliefs as reflected in the prior. But the analysis generally treats the data and prior as if they provided "independent" measurements—another fishy aspect of this approach.

[3] It is widely believed that spinach has substantially more iron than other green vegetables. This is evidently the result of a transcription error in the 1870s that shifted the decimal, multiplying the measured value by 10 (see, e.g., http://www.dailymail.co.uk/sciencetech/article-2354580/Popeyes-legendary-love-spinach-actually-misplaced-decimal-point.html0). The fact that the original value was far too high was well known before the Popeye character became popular in the 1930s.

[4] E.g., Bicchi et al. (2008, section 4.4.3).

Humans are also bad at judging and creating randomness: we have *apophenia* and *pareidolia*, a tendency to see patterns in randomness.[5] And when we deliberately try to create randomness, what we make has fewer patterns than genuinely random processes would generate. For instance, we produce too few runs and repeats (e.g., Schulz et al., 2012; Shermer, 2008). We are over-confident about our estimates and predictions (e.g., Kahnemann, 2011; Taleb, 2007). And our confidence is unrelated to our actual accuracy (e.g., Krug, 2007; Chua et al., 2004).[6]

In discussing the "neo-Bayesian" theory, LeCam (1977, pp. 155–156) gives these examples of uncertainty:

> It is clear that we can be uncertain for many reasons. For instance, we may be uncertain because (1) we lack definite information, (2) the events involved will occur according to the results of the spin of a roulette wheel, (3) we could find out by pure logic but it is too hard. The first type of uncertainty occurs in practically every question. The second assumes a well-defined mechanism. However, the neo-Bayesian theory seems to make no real distinction between probabilities attached to the three types. It answers in the same manner the following questions.
>
> (1) What is the probability that Eudoxus had bigger feet than Euclid?
> (2) What is the probability that a toss of a 'fair' coin will result in tails?
> (3) What is the probability that the $10^{137}+1$ digit of π is a 7?
> …
> Thus, presumably, when neo-Bayesians state that a certain event *A* has probability one-half, this may mean either that he did not bother to think about it, or that he has no information on the subject, or that whether *A* occurs or not will be decided by the toss of a fair coin. The number ½ itself does not contain any information about the process by which it was obtained …

An editorial in *Nature* (1978) also pushes back on the idea that all risks can be quantified, much less quantified on the same scale:

> LORD ROTHSCHILD, speaking on British television last week, argued that we should develop a table of risks so we could compare, say, the risk of our dying in an automobile accident with the risk of Baader-Meinhoff guerillas taking over the nuclear reactor next door. Then we would know how seriously to take our risks, be they nuclear power, damage to the environment or whatever.
>
> …
>
> [] Rothschild confused two fundamental distinct kinds of risk in his table: known risks-such as car accidents-where the risk is simply calculated from past events; and unknown risks—such as the terrorists taking over a fast breeder—which are matters of estimating the future. The latter risks inevitably depend on theory. Whether the theory is a social theory of terrorism or a risk-tree analysis of fast breeder failure, it will be open to conjecture. And it ought to be remembered that the history of engineering is largely a history of unforeseen accidents. Risk estimates can be proved only by events. Thus it is easy for groups, consciously or unconsciously, to bend their

---

[5] https://en.wikipedia.org/wiki/Apophenia, https://en.wikipedia.org/wiki/Pareidolia (last accessed 16 November 2016)

[6] E.g., https://en.wikipedia.org/wiki/Overconfidence_effect (last accessed 16 November 2016)

> calculations to suit their own objectives or prejudices. With unknown risks it is as important to take these into account as to come up with a number.

In short, insisting on quantifying all kinds of uncertainty on the same scale—probability—is neither helpful nor sensible. PSHA treats destructive earthquakes as if they were auto accidents, while they are arguably more like a terrorist attacks.

Rates versus probabilities

PSHA conflates empirical rates with probabilities. Historical rates are rates, not probabilities, nor are they in general estimates of probabilities. Klemeš (1989) wrote eloquently about this false equivalence in hydrology:

> The automatic identification of past frequencies with present probabilities is the greatest plague of contemporary statistical and stochastic hydrology. It has become so deeply engrained that it prevents hydrologists from seeing the fundamental difference between the two concepts. It is often difficult to put across the fact that whereas a histogram of frequencies for given quantities [] can be constructed for any function whether it has been generated by deterministic or random mechanism, it can be interpreted as a probability distribution only in the latter case. [] Ergo, automatically to interpret past frequencies as present probabilities means *a priori* to deny the possibility of any signal in the geophysical history; this certainly is not science but sterile scholasticism.
>
> The point then arises, why are these unreasonable assumptions made if it is obvious that probabilistic statements based on them may be grossly misleading, especially when they relate to physically extreme conditions where errors can have catastrophic consequences? The answer seems to be that they provide the only conceptual framework that makes it possible to make probabilistic statements, i.e. they must be used if the objective is to make such probabilistic statements.

Any finite series of dichotomous trials has an empirical rate of success. But the outcomes of a series of trials cannot tell you whether the trials were random in the first place: the mechanism behind the trial is key. Suppose there is a series of random Bernoulli trials,[7] that each trial has the same probability p of success, and that the trials are independent—like the standard model of coin tossing, treating 'heads' as 'success.' Then the Law of Large Numbers guarantees that the rate of successes converges (in probability) to the probability of success.

If a sequence of trials *is* random and the chance of success is the same in each trial, then the empirical rate of success is an unbiased estimate of the underlying chance of success. If the trials are random *and* they have the same chance of success *and* you know the dependence structure of the trials (for example, if the trials are independent), then you can quantify the uncertainty of that estimate of the underlying chance of success. *But the mere fact that something (e.g., seismicity) has a rate does not mean that it is the result of a random process*.

For example, suppose a sequence of heads and tails results from a series of random, independent tosses of an ideal fair coin. Then the rate of heads will converge (in probability) to one half. But suppose I give you the sequence 'heads, tails, heads, tails, heads, tails, heads, tails, heads, tails, …' *ad infinitum*. The limiting rate of heads is ½. While that sequence *could*

[7] A Bernoulli trial is a random trial that can result in *failure* or *success,* generally represented as 0 or 1, respectively.

be the result of a sequence of fair random tosses, that is implausible, and the sequence certainly need not be the result of a random process. Rates are not necessarily (estimates of) probabilities.

Here are two thought experiments:

1. You are in a group of 100 people. You are told that one person in the group will die next year. What is the chance it is you?
2. You are in a group of 100 people. You are told that one of them is named Philip. What is the chance it is you?

Both scenarios involve a rate of 1% in a group. But in the first one you are invited to say, 'the chance is 1%,' while in the second you are invited to say, 'that's a silly question.' The point is that a rate is not necessarily a probability, and that probability does not capture every kind of uncertainty.

In question 1, if the mechanism for deciding who will die in the next year is to select the tallest person and shoot them, there is nothing random. There is no *probability* that you will be the person who dies—you either are or are not the tallest person, just as you either are or are not named 'Philip.' If the mechanism for deciding who will die is to draw lots and shoot whoever gets the short straw, that might be reasonably modeled as random, in which case the probability that you are the person who dies is indeed 1%. *The existence of a probability is in the method of selection, not in the existence of a rate.* Rates and probabilities are not the same, and ignorance and randomness are not the same. Not all uncertainties can be put on the same scale.

Simulation and probability

Some seismologists act as if probabilities can be estimated in a 'neutral' or 'automatic' way by doing Monte Carlo simulations: just let the computer generate the distribution. For instance, an investigator might posit a failure tree model for seismicity. The values of some parameters in the model are unknown. In one approach to uncertainty quantification, values of those parameters are drawn pseudo-randomly from an assumed joint distribution (generally treating the parameters as independent). The distribution of outputs is interpreted as the probability of various outcomes in the real world.

Setting aside other issues in numerical modeling, Monte Carlo simulation is a way to substitute computing for hand calculation. It is not a way to *discover* the probability distribution of anything; it is a way to estimate the numerical values that flow from an *assumed* distribution. It is a substitute for doing an integral, not a way to uncover laws of Nature.

Monte Carlo doesn't tell you anything that wasn't already baked into the simulation. The distribution of the output comes from assumptions in the input (modulo bugs): a probability model for the parameters that govern the simulation. It comes from what you program the computer to do. Monte Carlo reveals the consequences of your assumptions, not anything new. The randomness is an assumption.

PSHA

PSHA purports to estimate the probability of a given level of ground shaking (acceleration), for instance, a level that would damage the containment structure of a nuclear power plant. It involves modelling earthquakes as occurring at random in space, time and with random magnitude. Then it models ground motion as being random, conditional on the occurrence of an earthquake of a given magnitude in a given place. From this, PSHA claims to estimate the 'exceedance probability,' the chance that the acceleration in some particular place exceeds some threshold level within some number of years.

PSHA arose from probabilistic risk assessment, which originated in aerospace and nuclear power, primarily. A big difference between PSHA and these other applications is that a spacecraft is an engineered system. Its properties are relatively predictable even before humans had launched a manned spaceflight, as are those of the environment it is operating in. Even before a nuclear reactor was built, people knew something about nuclear physics and thermodynamics. They knew something about the physical properties of concrete and steel. They knew something about pressure vessels.

We know very little about earthquakes, other than their phenomenology. We don't really understand the physical generating processes (Geller et al., 2015). We don't know in detail how they occur. There is a big difference between an engineered system whose components can be tested and a natural system that is inaccessible to experimentation.

PSHA models earthquakes as a marked stochastic process with known parameters. The fundamental relationship in PSHA is that the probability of a given level of ground movement in a given place is the integral over space and magnitude of the conditional probability of that level of movement given that there is an event of a particular magnitude in a particular place times the probability that there is an event of a particular magnitude.

This is just the law of total probability and the multiplication rule for conditional probabilities—but where does the probability come from in the first place? What justifies treating destructive earthquakes as random? That earthquakes occur at random is an *assumption*, not a matter of physics. Seismicity is complicated and unpredictable: *haphazard*, but not necessarily *random*. The standard argument to calibrate the PSHA fundamental relationship requires conflating rates with probabilities, the fallacy discussed at length above. For instance, suppose a magnitude eight event has been observed to occur about once a century in a given region. PSHA would assume that, therefore, the chance of a magnitude 8 event is 1% per year.

That is wrong, for a variety of reasons. First, there is an epistemic leap from a rate to the existence of an underlying, stationary random process that generated the rate, as discussed above (see the quotation from Klemeš in particular). Second, it involves the assumption that seismicity is uniform, which contradicts the observed clustering of seismicity in space and time. Third, it ignores the fact that even if seismicity were stationary and random, the historical rate is at best an *estimate* of a probability, not the exact value of the probability.

Among other infelicities, PSHA conflates frequencies with probabilities in treating relationships such as the Gutenberg-Richter (G-R) law, the historical spatial distribution of seismicity, and ranges of ground acceleration given the distance and magnitude of an earthquake as probability distributions. For instance, the G-R law says that the log of the number of earthquakes of a given magnitude in a given region is approximately proportional to the magnitude. Magnitude is a logarithmic scale, so the G-R law says that the relationship between "size" and frequency is approximately linear on a log-log plot (at least over some range of magnitude). While the G-R law is empirically approximately true, PSHA involves the additional assumptions that the magnitudes of future earthquakes are drawn *randomly* and independently from the G-R law. There is no basis for that assumption.

PSHA relies on the metaphor that earthquakes occur as if in a casino game. According to the metaphor, there is a special deck of earthquake cards. The game involves dealing one card per time period. If the card is blank, there is no earthquake. If the card is an eight, there is a magnitude 8 earthquake. If the card is a six, there is a magnitude 6 earthquake, and so forth.

There are tens of thousands of journal pages that amount to arguing about how many cards of each value are in the deck, how well the deck is shuffled, whether after each draw the card is replaced and the deck is shuffled again before dealing the next card, whether to add high-numbered cards to the deck if no high card has been drawn in a while, and so on. The literature, and the amount of money spent on this kind of work, are enormous—especially given that it has been unsuccessful scientifically. Three recent destructive earthquakes were in regions that seismic hazard maps said were relatively safe (Stein et al., 2012; see also Panza et al., 2014; Kossobokov et al., 2015). This should not be surprising, because PSHA is based on a metaphor, not on physics.

Here is a different metaphor: earthquakes occur like terrorist bombings. We don't know when or where they're going to happen. We know that they could be be large enough to hurt people when they do happen, but not how large. We know that some places are easier targets than others (e.g., places near active faults), and that some are more vulnerable than others (e.g., places subject to soil liquefaction and structures made of unreinforced masonry). But there is no probability *per se*. We might choose to *invent* a probability model to try to improve law enforcement or prevention, but that is different from a generative model according to which terrorists decide when and where to strike by rolling dice. In principle, the predictions of such a model could be tested—but fortunately, such events are sufficiently rare that no meaningful test is possible.

What would justify using the casino metaphor for earthquakes? It might be apt if the physics of earthquakes were stochastic (i.e., random)—but it isn't. It might be apt if stochastic models provided a compact, accurate representation of earthquake phenomenology, but they don't: the data show that the models are no good (see, e.g., Luen and Stark, 2012; Luen, 2010). The metaphor might be apt if the models led to useful predictions of future seismicity, but they don't (Luen, 2010).

PSHA suffers from two of the issues above, *viz.*, forcing all uncertainties to be on the same scale and conflating rates with probabilities. Cornell (1968), the foundational PSHA paper, writes:

> In this paper a method is developed to produce [various characteristics of ground motion] and their average return period for his site. The minimum data needed are only the seismologist's best estimates of the average activity levels of the various potential sources of earthquakes … The technique to be developed provides the method for integrating the individual influences of potential earthquake sources, near and far, more active or less, into the probability distribution of maximum annual intensity (or peak-ground acceleration, etc.). The average return period follows directly.
>
> …
>
> In general the size and location of a future earthquake are uncertain. They shall be treated therefore as random variables.

Whatever intuitive appeal and formulaic simplicity PSHA might have, there is no justification for treating everything that is uncertain as if it were random, with distributions that are known but for the values of a few parameters. Moreover, the method does not work in practice (Mulargia et al., 2017).

I believe we (seismologists and engineers) are better at ranking risks than quantifying them, and better at engineering calculations for human-built structures than we are at predicting ground motion from future earthquakes. Thus, society might be better off if we address seismic risk by starting with a financial budget rather than a risk budget. Instead of asking, "how can we make this structure have a 90 percent chance of lasting 100 years?," we might be better off asking, "if we were willing to spend $10 million to harden this structure, how should we spend it?"

In summary, PSHA is based on metaphors and fallacies, not on physics. The numbers it provides are not estimates of a quantity that even has a coherent definition, except by analogy to idealized casino games that have little in common with geophysical systems. The numbers are not meaningful, nor are they a rational basis for public policy.

References

Bicchi, A., M. Buss, M.O. Ernst, and A. Peer (2008). *The Sense of Touch and Its Rendering: Progress in Haptics Research,* Springer-Verlag, Berlin Heidelberg

Box, G.E.P., (1976). Science and Statistics, *Journal of the American Statistical Association*, *71,* 791–799.

Chua, E.F., E. Rand-Giovannetti, D.L. Schacter, M.S. Albert, and R.A. Sperling (2004). Dissociating Confidence and Accuracy: Functional Magnetic Resonance Imaging Shows Origins of the Subjective Memory Experience, *Journal of Cognitive Neuroscience 16*, 1131–1142.

Cornell, C.A. (1968). Engineering seismic risk analysis, *Bulletin of the Seismological Society of America*, *58* (5): 1583–1606. doi: 10.1785/BSSA0580051583

Desrosières, A. (1998). *The Politics of Large Numbers : A History of Statistical Reasoning*, Harvard University Press, Cambridge.

Diaconis, P. and B. Skyrms (2018). *Ten Great Ideas About Chance*, Princeton University Press, Princeton and Oxford.

Freedman, D.A. (2010). Issues in the Foundations of Statistics: Probability and Statistical Models, in *Statistical Models and Causal Inference: A Dialogue with the Social Sciences*, D. Collier, J. Sekhon, and P.B. Stark, eds., Cambridge University Press, NY.

Geller, R.J., F. Mulargia, and P.B. Stark (2015). Why we need a new paradigm of earthquake occurrence, in *Subduction Dynamics: From Mantle Flow to Mega Disasters*, Morra, G., D.A. Yuen, S.D. King, S.-M. Lee, and S. Stein, eds. American Geophysical Union, doi: 10.1002/9781118888865.ch10

Kahneman, D. (2011). *Thinking, Fast and Slow*, Farrar, Strauss, and Giroux, NY.

Klemeš, V. (1989). The Improbable Probabilities of Extreme Floods and Droughts, in *Hydrology of Disasters: Proceedings of the World Meteorological Organization*, Starosolszky, O., and O.M. Melder, eds., Routledge.

Kossobokov, V.G., A. Peresan and G. F. Panza (2015). Reality Check: Seismic Hazard Models You Can Trust, *EOS Trans. AGU.*, 96, 9–11. doi: 10.1029/2015EO031919.

Krug, K. (2007). The Relationship Between Confidence And Accuracy: Current Thoughts of the Literature and a New Area of Research, *Applied Psychology in Criminal Justice, 2007, 3,* 7–41.

Le Cam, L. (1977). A Note on Metastatistics, or 'An Essay toward Stating a Problem in the Doctrine of Chances,' *Synthese*, *36*, 133–160, at 155–156.

Luen, B. (2010). *Earthquake prediction: Simple methods for complex phenomena*, Ph.D. Dissertation, Department of Statistics, University of California, Berkeley, UMI Number 3449030.

Luen, B. and P.B. Stark (2008). Testing Earthquake Predictions. IMS Lecture Notes—Monograph Series. *Probability and Statistics: Essays in Honor of David A. Freedman*, 302–315. Institute for Mathematical Statistics Press, Beachwood, OH.

Luen, B., and Stark, P.B. (2012). Poisson tests of declustered catalogs, *Geophysical Journal International, 189,* 691–700

Mayo, D. (2018). *Statistical Inference as Severe Testing: How to Get Beyond the Statistics Wars*, Cambridge University Press.

Millikan, R.A. (1913). On the Elementary Electrical Charge and the Avogadro Constant, *Physical Review, 2,* 109–143. http://history.aip.org/history/exhibits/gap/PDF/millikan.pdf (last accessed 16 November 2016)

Mulargia, F., P.B. Stark, and R.J. Geller (2017). Why is probabilistic hazard analysis still used? *Physics of the Earth and Planetary Interiors, 264*, 63–75, doi: 10.1016/j.pepi.2016.12.002

Nature (1978). Rothschild's numerate arrogance, *Nature*, *276*, 429. doi: 10.1038/276429a0

Panza, G., V. G. Kossobokov, A. Peresan and A. Nekrasova (2014). Why are the Standard Probabilistic Methods of Estimating Seismic Hazard and Risks Too Often Wrong? Chapter 12, in M. Wyss and J.F. Shroder (eds.), *Earthquake Hazard, Risk, and Disasters*, Elsevier Inc., pp 309-357. doi: 10.1016/B978-0-12-394848-9.00012-2

Schulz, M.-A., B. Schmalbach, P. Brugger, and K. Witt (2012). Analyzing humanly generated random number sequences: a pattern-based approach. *PLoS ONE 7*, e41531, doi:10.1371

Shermer, M. (2008). Patternicity: Finding Meaningful Patterns in Meaningless Noise, *Scientific American,* December.

Stark, P.B., and D.A. Freedman (2010). What is the Chance of an Earthquake, in *Statistical Models and Causal Inference: A Dialogue with the Social Sciences*, D. Collier, J. Sekhon, and P.B. Stark, eds., Cambridge University Press, NY.

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

Stein, S., and Stein, J. (2013). Shallow versus deep uncertainties in natural hazard assessments, *EOS, 94,* 133–140.

Stigler, S.M. (1986). *The History of Statistics: The Measurement of Uncertainty Before 1900*, Harvard University Press, Cambridge, MA

Taleb, N.N. (2007). *The Black Swan: The Impact of the Highly Improbable*, Random House, NY.

Tversky, A., and D. Kahneman (1975). Judgment under uncertainty: heuristics and biases. *Science*, *185*, 1124–1131.