

## Review Article

## Why earthquake hazard maps often fail and what to do about it

Seth Stein <sup>a,\*</sup>, Robert J. Geller <sup>b</sup>, Mian Liu <sup>c</sup><sup>a</sup> Department of Earth and Planetary Sciences, Northwestern University, Evanston IL 60208, USA<sup>b</sup> Dept. of Earth and Planetary Science, Graduate School of Science, University of Tokyo, Tokyo 113-0033, Japan<sup>c</sup> Department of Geological Sciences, University of Missouri, Columbia, MO 65211, USA

## ARTICLE INFO

## Article history:

Received 16 December 2011

Received in revised form 21 June 2012

Accepted 24 June 2012

Available online 4 July 2012

## Keywords:

Earthquake hazards

Seismology

Faulting

Hazard mitigation

## ABSTRACT

The 2011 Tohoku earthquake is another striking example – after the 2008 Wenchuan and 2010 Haiti earthquakes – of highly destructive earthquakes that occurred in areas predicted by earthquake hazard maps to be relatively safe. Here, we examine what went wrong for Tohoku, and how this failure illustrates limitations of earthquake hazard mapping. We use examples from several seismic regions to show that earthquake occurrence is typically more complicated than the models on which hazard maps are based, and that the available history of seismicity is almost always too short to reliably establish the spatiotemporal pattern of large earthquake occurrence. As a result, key aspects of hazard maps often depend on poorly constrained parameters, whose values are chosen based on the mapmakers' preconceptions. When these are incorrect, maps do poorly. This situation will improve at best slowly, owing to our limited understanding of earthquake processes. However, because hazard mapping has become widely accepted and used to make major decisions, we suggest two changes to improve current practices. First, the uncertainties in hazard map predictions should be assessed and clearly communicated to potential users. Recognizing the uncertainties would enable users to decide how much credence to place in the maps and make them more useful in formulating cost-effective hazard mitigation policies. Second, hazard maps should undergo rigorous and objective testing to compare their predictions to those of null hypotheses, including ones based on uniform regional seismicity or hazard. Such testing, which is common and useful in similar fields, will show how well maps actually work and hopefully help produce measurable improvements. There are likely, however, limits on how well hazard maps can ever be made because of the intrinsic variability of earthquake processes.

© 2012 Elsevier B.V. All rights reserved.

## Contents

1.	Introduction	2
2.	What went wrong at Tohoku	2
3.	Why hazard maps matter	5
4.	Lessons from the Tohoku failure for hazard maps	6
4.1.	What can go wrong	6
4.1.1.	Bad physics	6
4.1.2.	Bad assumptions	6
4.1.3.	Bad data	6
4.1.4.	Bad luck	6
4.2.	Too many black swans	7
5.	Hazard map challenges	8
5.1.	Forecasting and prediction	8
5.2.	Defining the hazard	9
5.2.1.	How?	10
5.2.2.	Where?	11
5.2.3.	When?	12
5.2.4.	How big?	15
5.2.5.	How much shaking?	17

\* Corresponding author. Tel.: +1 847 491 5265; fax: +1 847 491 8060.

E-mail addresses: [seth@earth.northwestern.edu](mailto:seth@earth.northwestern.edu) (S. Stein), [bob@eps.s.u-tokyo.ac.jp](mailto:bob@eps.s.u-tokyo.ac.jp) (R.J. Geller), [lium@missouri.edu](mailto:lium@missouri.edu) (M. Liu).

6.	What to do . . . . .	18
6.1.	Assess and present uncertainties . . . . .	20
6.2.	Test hazard maps . . . . .	20
7.	Mission impossible? . . . . .	22
	Acknowledgments . . . . .	23
	References . . . . .	23

*"It's a problem that physicists have learned to deal with: They've learned to realize that whether they like a theory or they don't like a theory is not the essential question. Rather, it is whether or not the theory gives predictions that agree with the experiment. It is not a question of whether a theory is philosophically delightful, or easy to understand, or perfectly reasonable from the point of view of common sense." (Feynman, 1986)*

*"My colleagues had the responsibility of preparing long-range weather forecasts, i.e., for the following month. The statisticians among us subjected these forecasts to verification and found they differed in no way from chance. The forecasters themselves were convinced and requested that the forecasts be discontinued. The reply read approximately like this: The commanding general is well aware that the forecasts are no good. However, he needs them for planning purposes." Nobel Prize winner Kenneth Arrow describing his experience as a military weather forecaster in World War II (Gardner, 2010)*

## 1. Introduction

Until March 11, 2011, residents of Japan's Tohoku coast were proud of their tsunami defenses (Onishi, 2011a,b,c). The 10-meter high sea walls that extended along a third of the nation's coastline – longer than the Great Wall of China – cost billions of dollars and cut off ocean views. However, these costs were considered a small price to pay for eliminating the threat that had cost many lives over the past centuries. In the town of Taro, people rode bicycles, walked, and jogged on top of the impressive wall. A school principal explained, "For us, the sea wall was an asset, something we believed in. We felt protected."

The defenses represented what an affluent technological society could do. Over a period of years, most recently in 2010, an agency of the Japanese government, advised by some of Japan's leading seismologists, had calculated precisely what kinds of earthquakes could be expected in different parts of the country. The largest hazard was assumed to be from thrust fault earthquakes to the east, where the Pacific plate subducts at the Japan Trench and the Philippine Sea plate subducts at the Nankai Trough. For the area of the Japan Trench off Miyagi prefecture on the Tohoku coast, the hazard mappers stated that there was a 99% probability that a magnitude 7.5 earthquake would occur in the next 30 years (Earthquake Research Committee, 2009, 2010). This forecast, as well as similar detailed seismicity forecasts for all other regions, was used to produce the national seismic hazard map that predicted the probability that the maximum ground acceleration (shaking) in any area would exceed a particular value during the next 30 years. Larger expected shaking corresponds to higher predicted seismic hazard. A similar approach was used to forecast the largest expected tsunami. Engineers, in turn, used the results to design tsunami defenses and build structures to survive earthquake shaking.

All this planning proved inadequate on March 11, when a magnitude 9 earthquake offshore generated a huge tsunami that overtopped the sea walls, causing over 19,000 deaths (including missing; official

police data as of December 2011) and at least \$200 billion damage (Normile, 2012), including crippling nuclear power plants. This earthquake released about 150 times the energy of the magnitude 7.5 quake that was expected for the Miyagi-oki region by the hazard mappers. Somehow, the mapping process significantly underpredicted the earthquake hazard. The complex decision making process involved for the Fukushima nuclear power plant is reviewed by Nöggerath et al. (2011).

The hazard map, whose 2010 version is shown in Fig. 1, predicted less than a 0.1% probability of shaking with intensity "6-lower" (on the Japan Meteorological Agency intensity scale) in the next 30 years. In other words, such shaking was expected on average only once in the next 30/0.001 or 30,000 years. However, within two years, such shaking occurred.

How this discrepancy arose has become a subject of extensive discussion among seismologists (Kerr, 2011). We raised three issues in a recent short opinion article (Stein et al., 2011): 1) What went wrong for Tohoku? 2) Was this failure an exceptional case, or does it indicate systemic difficulties in earthquake hazard mapping? 3) How to improve this situation? Here we discuss these issues in more detail.

## 2. What went wrong at Tohoku

Analysis of the Japanese national seismic hazard map (Fig. 1) after the earthquake (Geller, 2011) pointed out that the Tohoku area was shown as having significantly lower hazard than other parts of Japan, notably the Tokai, Tonankai, and Nankai districts to the south. This assessment arose for several interrelated reasons. We use this example to illustrate how, owing to limited knowledge, hazard maps often depend crucially on mapmakers' preconceptions, which can lead to significant overprediction or underprediction of hazards.

The map reflects the widespread view among Japanese seismologists that M 9 earthquakes would not occur on the Japan Trench off Tohoku (Chang, 2011; Sagiya, 2011; Yomogida et al., 2011). The largest future earthquakes along different segments of the trench there were expected to have magnitude between 7 and 8 (Fig. 2) (Earthquake Research Committee, 2009, 2010). The model assumed that different segments of the trench would not break simultaneously.

However, the March 2011 earthquake broke five segments, yielding a magnitude 9 earthquake. As illustrated in Fig. 3a, an M 9 earthquake involves a larger average slip over a larger fault area, resulting in a larger tsunami because the maximum tsunami run-up height is typically about twice the fault slip (Okal and Synolakis, 2004). Thus the March earthquake generated a huge tsunami that overtopped 10-meter high sea walls.

Such a giant earthquake was not anticipated off Tohoku due to several incorrect assumptions that reinforced one another. The available earthquake history that appeared to show no record of such giant earthquakes seemed consistent with an incorrect hypothesis that the subduction dynamics precluded M 9 earthquakes in the Japan Trench.

Specifically, the fact that the available history had no record of giant earthquakes seemed plausible, given an analysis (Ruff and Kanamori, 1980) of the largest known earthquakes at various subduction zones. These data (Fig. 3b) appeared to show a striking pattern –

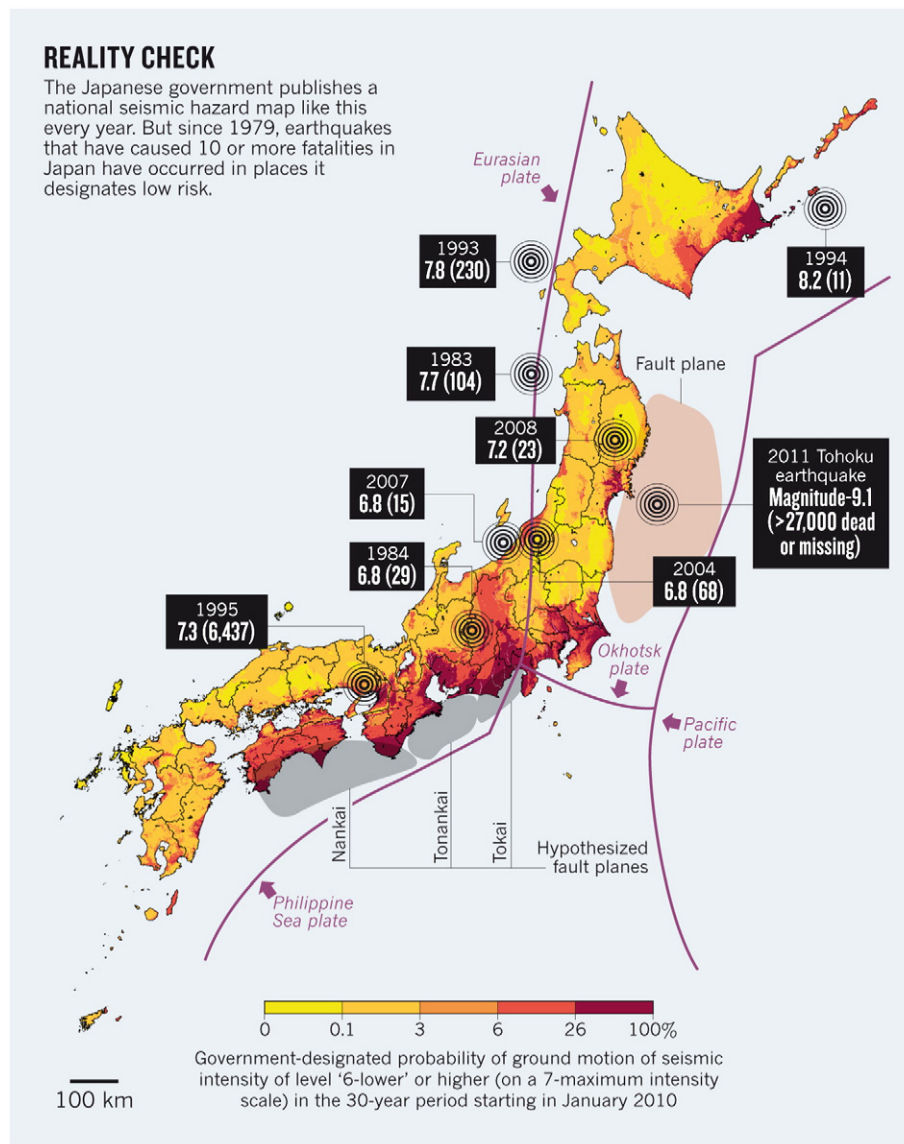


Fig. 1. Comparison of Japanese government hazard map to the locations of earthquakes since 1979 that caused 10 or more fatalities (Geller, 2011).

magnitude 9 earthquakes only occurred where lithosphere younger than 80 million years old was subducting rapidly, faster than 50 mm/yr. This result made intuitive sense, because both young age and speed could favor strong mechanical coupling at the interface between the two plates (Fig. 3c). Because oceanic lithosphere cools as it moves away from a ridge and ages, young lithosphere is less dense and thus more buoyant. Similarly, faster subducting lithosphere should increase friction at the plate interface. The stronger coupling was, in turn, assumed to produce larger earthquakes when the interface eventually slips. Hence the age of the subducting plate and the rate of plate convergence were used to predict the maximum expected earthquake size in subduction zones.

This model was widely accepted (Satake and Atwater, 2007) until the 2011 Tohoku earthquake (Chang, 2011). However, the giant December 26, 2004 magnitude 9.3 Sumatra earthquake that generated the devastating Indian Ocean tsunami, should, according to the model, have generated at most an earthquake with a magnitude of about 8. Hence studies after the 2004 Sumatra earthquake (McCaffrey, 2007, 2008; Stein and Okal, 2007) showed that a fundamental rethinking was needed for several reasons. Better rates of plate motion were available from new Global Positioning System data. Additional information

on maximum earthquake sizes came from new observations, including paleoseismic estimates of the size of older earthquakes such as the 1700 event at the Cascadia subduction zone (Satake and Atwater, 2007). Moreover, it was recognized that although the largest trench earthquakes are typically thrust fault events, this is not always the case. With the newer data the proposed correlation between earthquake size and the rate and age of subducting slabs vanished, as the 2011 Tohoku earthquake subsequently confirmed (Fig. 3d).

Thus instead of only some subduction zones being able to generate magnitude 9 earthquakes, it now looks like many or all can (McCaffrey, 2007, 2008). The apparent pattern had resulted from the fact that magnitude 9 earthquakes are rare – on average there is less than one per decade (Stein and Wyssession, 2003). They are about ten times rarer than magnitude 8. Thus the short seismological record (the seismometer was invented in the 1880s and seismograms that allow magnitude 9 events to be reliably quantified have only been available since about 1950) misled seismologists into assuming that the largest earthquakes known for each subduction zone were the largest that could occur there. Moreover, until recently seismologists tended to downplay geological evidence for past extremely large earthquakes, such as that of Minoura et al. (2001) for Tohoku.

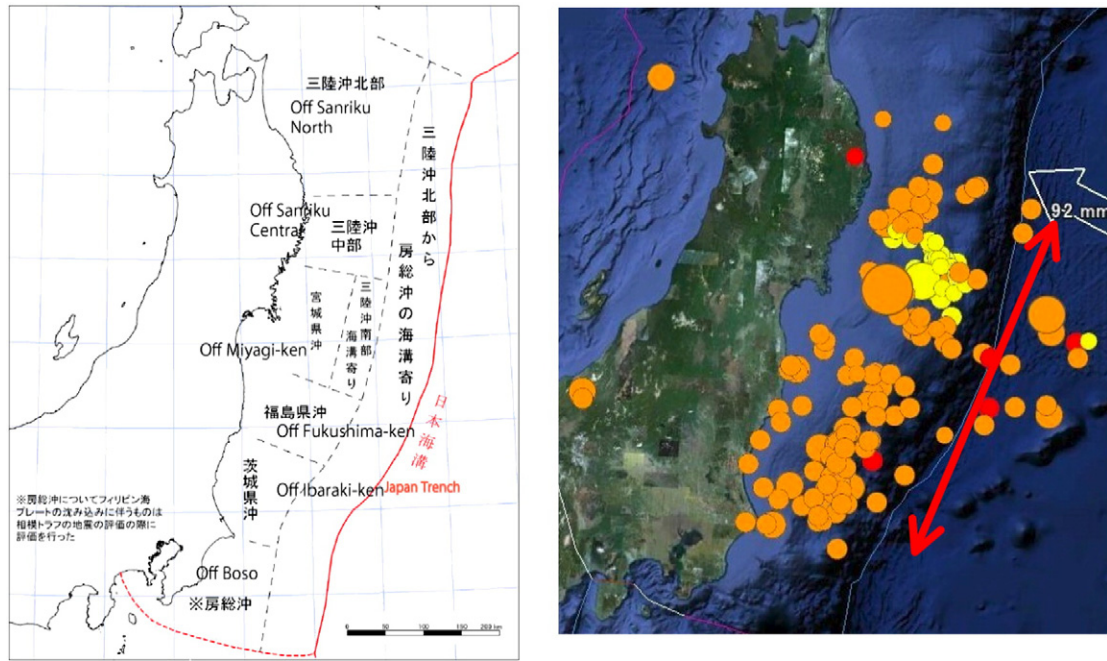


Fig. 2. Comparison of the trench segments assumed in the Japanese hazard map to the aftershock zone of the March 11, 2011 earthquake, which broke five segments (left: Earthquake Research Committee, right: U.S. Geological Survey).

These results also illustrated the weakness of the concept of strong interplate coupling leading to larger earthquakes. Although the frictional properties of the plate interface are hypothesized to control earthquake rupture geometry (Scholz, 2002), efforts to directly relate maximum earthquake size to the physical processes at the plate interface have been unsuccessful (Stein and Wyssession, 2003). The magnitude of earthquakes depends on their seismic moment, which is the product of the shear modulus, average slip, and area of fault rupture (Kanamori, 1977a). The ruptured area has the largest effect, because subduction zone earthquakes break various segments of a trench, as shown in Fig. 3e for the Nankai Trough. Sometimes one segment ruptures, and other times more than one does. The more segments rupture, the bigger the earthquake. Whether these segments are fixed and represent long-term properties of the interface, or are simply transient effects due to the history of slip in earlier earthquakes, remains unknown. It is similarly unclear what controls the variation in the amount of slip along the rupture, as strikingly illustrated by the surprisingly high values observed on part of the Tohoku rupture (Simons et al., 2011).

Before December 2004, seismologists only knew of earthquakes along the Sumatra trench with magnitude less than 8 due to short ruptures (Bilham et al., 2005), so the much larger multi-segment rupture came as a surprise. Plate motion calculations show that earthquakes like 2004's should happen about 500 years apart (Stein and Okal, 2005), so the short history available did not include any. Paleoseismic studies have since found deposits from a huge tsunami about 600 years ago in northern Sumatra (Monecke et al., 2008).

Similar variability is found at other trenches (Satake and Atwater, 2007). For example, the magnitude 9.5 1960 Chilean earthquake, the largest ever seismologically recorded, was a multisegment rupture much bigger than typical on that trench (Stein et al., 1986). Similarly, it appears that the very large Cascadia subduction zone earthquake in 1700 was a multi-segment rupture, and smaller events occur in the intervals between the larger ones (Kelsey et al., 2005).

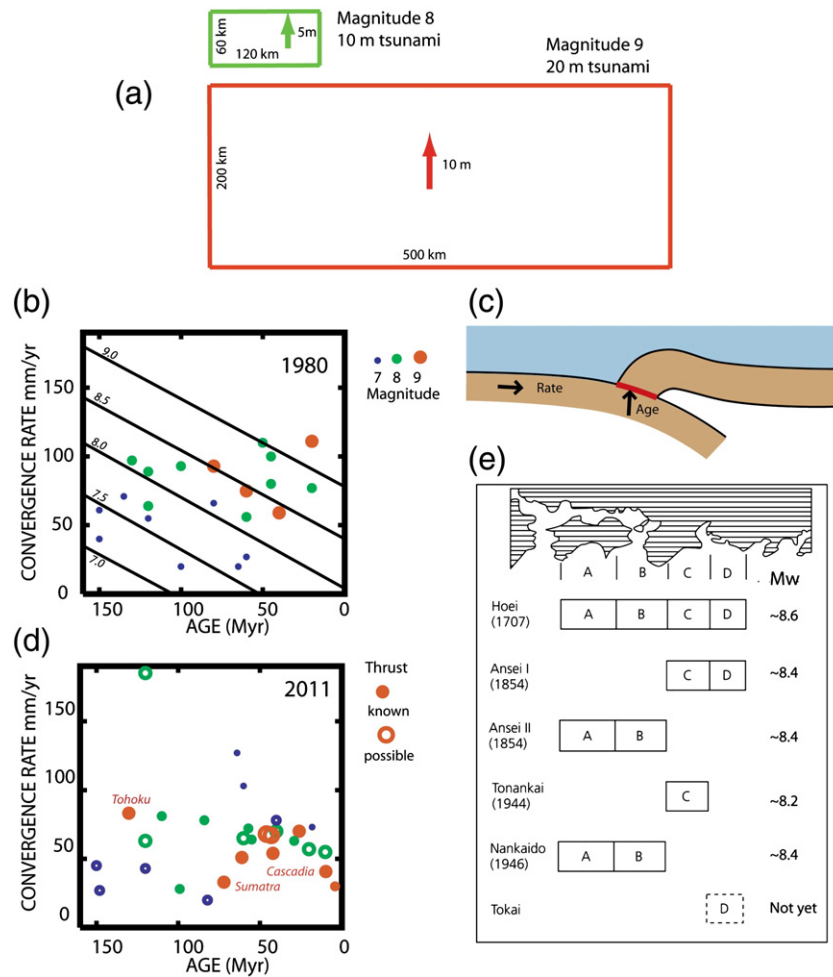
The presumed absence of giant earthquakes on the Japan Trench was implicitly interpreted as indicating that much of the subduction occurred aseismically. The Kurile trench, just to the north, seemed to show this discrepancy. The largest seismologically recorded earthquakes there are magnitude 8, which only account for about one third of the plate motion. Hence it had been assumed that most of the

subduction occurred aseismically (Kanamori, 1977b). However, more recently discovered deposits from ancient tsunamis show that much larger earthquakes had happened in the past (Nanayama et al., 2003), accounting for much of the subduction that had been thought to occur aseismically. In hindsight, the same applied off Tohoku.

In the decade prior to the March 2011 earthquake, increasing attention was also being paid to data showing that large tsunamis had struck the area in 869 (Minoura et al., 2001), 1896, and 1933. Some villages had stone tablets marking the heights reached by previous tsunamis and warning “Do not build your homes below this point” (Fackler, 2011). GPS data also were recognized as showing a much higher rate of strain accumulation on the plate interface than would be expected if a large fraction of the subduction occurred aseismically (Loveless and Meade, 2010). Including these data would have strengthened the case for considering the possibility of large earthquakes.

However, the revised ideas about maximum earthquake and tsunami size were not yet fully appreciated and incorporated into the Japanese hazard map. Thus, as summarized by Sagiya (2011) “If historical records had been more complete, and if discrepancies between data had been picked up, we might have been alert to the danger of a magnitude-9 earthquake hitting Tohoku, even though such an event was not foreseen by the Japanese government.” Instead, the hazard map focused on the Nankai Trough area (Geller, 2011). Based on the seismic gap model and the time sequence of earthquakes there (Fig. 3e), large earthquakes are expected on the Nankai (segments A–B), Tonankai (segment C), and especially Tokai (segment D) portions of the trench. Thus the 2010 hazard map (Fig. 1) and its predecessors show this area as the most dangerous in Japan. As noted by Geller (2011):

*“The regions assessed as most dangerous are the zones of three hypothetical ‘scenario earthquakes’ (Tokai, Tonankai and Nankai; see map). However, since 1979, earthquakes that caused 10 or more fatalities in Japan actually occurred in places assigned a relatively low probability. This discrepancy – the latest in a string of negative results for the characteristic earthquake model and its cousin, the seismic-gap model – strongly suggests that the hazard map and the methods used to produce it are flawed and should be discarded.”*



**Fig. 3.** What went wrong at Tohoku. (a) Illustration of the relative fault dimensions, average fault slip, and average tsunami run-up for magnitude 8 and 9 earthquakes. (b) Data available in 1980, showing the largest earthquake known at various subduction zones. Magnitude 9 earthquakes occurred only where young lithosphere subducts rapidly. Diagonal lines show predicted maximum earthquake magnitude. (Ruff and Kanamori, 1980). (c) Physical interpretation of this result in terms of strong mechanical coupling and thus large earthquakes at the trench interface. (d) Data available today, updated from Stein and Okal (2007) by including 2011 Tohoku earthquake. (e) Earthquake history for the Nankai trough area (Ando, 1975) illustrating how different segments rupturing cause earthquakes of different magnitudes (Stein and Okal, 2011).

### 3. Why hazard maps matter

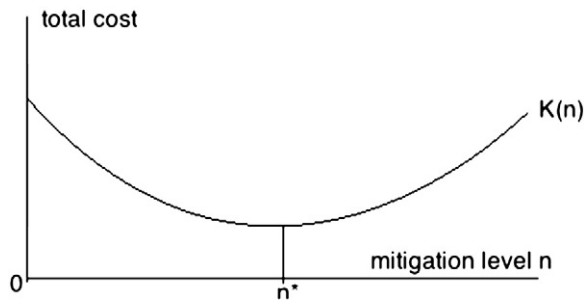
The Tohoku example illustrates how earthquake hazard maps are crucial in developing hazard mitigation strategies. Society faces the challenge of deciding how much of its resources to spend on natural hazard mitigation. More mitigation can reduce losses in possible future disasters, at increased cost. Less mitigation reduces costs, but can increase potential losses.

The discussion in Japan about reconstruction of the Tohoku coast, which suffered enormous damage from the tsunami generated by the earthquake of March 11, 2011, illustrates this tradeoff. Because the tsunami overtopped 5–10 m high sea walls, destroying most of the seawalls in the impacted area (Normile, 2012), the extent to which the seawalls and other defenses should be rebuilt is a difficult and debated question. The issue is illustrated by the city of Kamaishi (Onishi, 2011c). The city, although already declining after its steel industry closed, was chosen for protection by a \$1.6 billion breakwater. A song produced by the government “Protecting Us for a Hundred Years” praised the structure “It protects the steel town of Kamaishi, it protects our livelihoods, it protects the people’s future.” However, the breakwater collapsed when struck by the tsunami. 935 people in the city died, many of whom could have evacuated once warnings were given but did not, believing they were safe. Although the breakwater is being rebuilt, critics argue that it would be more efficient to relocate such communities inland, because their

populations are small and decreasing. Otherwise “in 30 years there might be nothing here but fancy breakwaters and empty houses.”

Because building coastal defenses adequate to withstand tsunamis as large as March 2011’s is too expensive, those planned are about 12 m high, only a few meters higher than the older ones (Cyranski, 2012; Normile, 2012). These are planned to provide protection for the largest tsunamis expected every 200–300 years, augmented with land-use planning to provide some protection against much larger tsunamis. The defenses should reduce economic losses, while improved warning and evacuations should reduce loss of lives.

Although such policy issues are complicated and must be decided politically, their economic aspects can be conceptualized by considering a simple model for deciding how high a seawall to construct (Stein and Stein, 2012) based on economic modeling approaches (Stein, 2012). As shown in Fig. 4, the optimal level of mitigation – in this case the height of a seawall – minimizes the total cost to society, which is the sum of the cost of constructing tsunami defenses and the expected property and indirect losses from tsunamis. The expected loss is the sum of the losses expected for tsunamis of different heights times the probability of a tsunami of that height. Because both terms depend on the mitigation level, their sum has a minimum at the optimal level of mitigation. Less mitigation decreases construction costs but increases the expected loss and thus total cost, whereas more mitigation decreases the expected loss but increases the total cost.



**Fig. 4.** Variation in total cost, the sum of expected loss and mitigation cost, as a function of mitigation level. The optimal level of mitigation,  $n^*$ , minimizes the total cost. The expected loss depends on the hazard model, so the better the hazard model, the better the mitigation policy (Stein and Stein, 2012).

A hazard model is crucial, because it is used to predict the probabilities of tsunamis of different heights and hence the expected loss. A similar analysis can be done for earthquake ground shaking, in which the expected loss is predicted from a hazard model's predicted shaking and data about building stock and vulnerability (Leonard et al., 2007).

Society's goal is to choose a level of safety that makes economic sense, because such mitigation diverts resources from other uses. Ideally mitigation should not be too weak, permitting undue risks, or too strong, imposing unneeded costs. These questions are complex, and involve assessing both the economic costs and benefits and comparing the relative cost per life saved by earthquake hazard mitigation to that required to save a life through other improved health or safety measures (Stein, 2010).

As a result, the best decisions come when a hazard map neither overpredicts nor underpredicts the hazard. Naturally, a hazard map has uncertainty, but this can be included in the analysis. It is thus crucial to understand how well hazard maps are performing, how to assess their uncertainties, and how to improve them to the extent possible.

#### 4. Lessons from the Tohoku failure for hazard maps

In some cases, earthquake hazard maps have done well at predicting the shaking from a major earthquake; in other cases they have done poorly (Kossobokov and Nekrasova, 2012). Is the Tohoku failure a rare exception? Or does it illustrate major problems with current hazard mapping?

##### 4.1. What can go wrong

To explore this issue, we use the Tohoku earthquake to identify four partially overlapping factors that can cause a hazard map to fail. Although all are much easier to identify after a map failure than before one, they are useful to bear in mind when trying to assess either how a map failed or how useful a map will be in the future.

###### 4.1.1. Bad physics

Some map failures result from incorrect physical models of the faulting processes. The basic models underlying the hazard maps are those of "characteristic earthquakes", which assumes that parts of a fault or fault segment will rupture in a predictable fashion, producing characteristic earthquakes with quasi-regular recurrence intervals, and of "seismic gaps", where the fault has not ruptured recently relative to other parts of the fault and are thus most likely to rupture in the future. The Japanese hazard map shows high hazard along the Nankai Trough, especially the Tokai segment ("D" in Fig. 3e), because the mappers assume that it is a seismic gap where earthquakes are "due." Although this concept underlies a great deal of earthquake seismology and has had some success, it remains controversial (Kagan, 1996). The physics that causes some ruptures to

stop at short distances while others to rip through many segments is not well understood. Kagan and Jackson (1991, 1995) found that the seismic-gap model does not identify future earthquake locations significantly better than assuming earthquakes occur randomly on these faults. Although these negative results have never been refuted, many hazard mapping studies implicitly or explicitly continue to use the seismic-gap model.

In addition to these general presumptions, specific ones are often made for individual areas. As discussed, the Tohoku area was mapped to have low hazard based on the presumption that M 9 earthquakes could only occur where young lithosphere was subducting rapidly, which we know now is not the case.

###### 4.1.2. Bad assumptions

Many map failures result from assumptions about which faults are present, which are active, how fast they are accumulating strain, and how it will be released. Although these somewhat overlap with the earlier category, they can be viewed as model parameter choices that prove inaccurate, rather than as assumptions about the underlying processes. The Japanese hazard map's presumption that giant earthquakes did not occur in the Tohoku area (bad physics) led to the view that the different segments would not break together. Interactions between faults (Luo and Liu, 2010; McCloskey et al., 2005; Stein, 1999) and unsteady loading rates are among many other factors that can complicate assumptions for other areas. Hazard maps also require assumptions about when and/or how often large earthquakes will recur. These estimates have large uncertainties and often prove incorrect, because (as discussed later), reliably estimating earthquake probabilities is very difficult (Freedman and Stark, 2003; Savage, 1991).

###### 4.1.3. Bad data

Often crucial data are "wrong" in the sense that they are lacking, incomplete, or misinterpreted. Earthquake recurrence in space and time is highly variable and still not well understood. As a result, earthquake hazard mapping is challenging even in areas where there are good earthquake records from instrumental seismology (which began in about 1900), historical accounts and paleoseismic (including in some cases paleotsunami) data going back further in time, and GPS data showing the presence or absence of strain accumulation. For Tohoku, the data showing very large past earthquakes and ongoing strain accumulation were available but not yet incorporated in the map.

The challenge is even greater in most other areas, where far less data are available, typically because the earthquake history available from instrumental and paleoseismic records is too short compared to the long and variable recurrence time of large earthquakes. In such cases, the locations, magnitudes of the largest future earthquakes and the resulting shaking expected are poorly known. In particular, some earthquakes occur on faults that had not been previously identified, in many cases because they are "blind" and have no clear surface expression (Stein and Yates, 1989; Valensise and Pantosti, 2001). Insufficient information about the shaking in earthquakes that occurred prior to the advent of modern seismometers also makes it difficult to accurately predict the shaking in future earthquakes.

###### 4.1.4. Bad luck

Because most hazard maps predict the maximum shaking expected with some probability in some time interval, earthquakes can produce higher shaking without invalidating a map. A much larger earthquake and resulting shaking than predicted can be regarded as an operational failure in terms of the map's providing information for hazard mitigation. However, formally it need not indicate a map failure. Instead, it can be considered a rare event – a "black swan" (Taleb, 2007) – that should not be used to judge the map as unsuccessful. Although the Japanese map predicted a less than a 0.1% probability of shaking with intensity 6 in the next 30 years, the fact that it occurred within two years could simply reflect a very

low probability event. Moreover, when large subduction thrust earthquakes eventually occur in the Tokai, Tonankai and Nankai portions of the trench, they could be used to declare the map successful.

#### 4.2. Too many black swans

The recent steady stream of large earthquakes around the world that produce higher shaking than predicted (Kerr, 2011) increasingly raise doubts about whether they can usefully be regarded as isolated “black swans”. If such operational failures are more common than predicted by hazard maps, they would represent a systemic failure of the maps.

This may be the case, because analyses of large earthquakes worldwide subsequent to the 1999 publication of the Global Seismic Hazard Map (Kossobokov and Nekrasova, 2012) find that the shaking in these earthquakes is often significantly higher than predicted. This effect is especially large for the largest ( $M > 7.5$ ) earthquakes, which thus cause many more fatalities than expected (Wyss et al., 2012). Naturally, this analysis is biased against the maps because it considers large earthquakes that did occur rather than sample an area uniformly, which would also consider areas where little or no shaking occurred. However, it suggests the advantage of analyzing these cases, rather than dismissing them as “black swans”, to explore what caused the discrepancies and how to improve the maps. To do this, we first consider a few examples that illustrate the four factors just discussed.

The 2008 Wenchuan earthquake ( $M$  7.9) in Sichuan Province, China, caused more than 80,000 deaths. It occurred on the Longmenshan fault (Fig. 5), which was assessed, based on the lack of recent seismicity (Fig. 5b), to have low hazard. The green and yellow colors in the hazard map show that the maximum ground acceleration predicted to have a 10% chance of being exceeded once in 50 years, or on average once about every  $50/0.1 = 500$  years, was less than  $0.8\text{--}1.6\text{ m/s}^2$ . This value is less than 16% of the acceleration of gravity ( $9.8\text{ m/s}^2$ ), so little building damage would be expected (Fig. 6).

However, a different view would have come from considering the geology (Witze, 2009), namely that this is where the Tibetan plateau is thrusting over the Sichuan basin, as illustrated by the dramatic relief. GPS data also show  $1\text{--}3\text{ mm/yr}$  of motion across the Longmenshan Fault (Meng et al., 2008; Zhang et al., 2004). Although this seems slow, over  $500\text{--}1000$  years such motion would accumulate enough for a magnitude 7 earthquake, and longer intervals would permit even larger earthquakes. Given the lack of evidence for large earthquakes on the Longmanshan fault in the past thousand years or so before the Wenchuan earthquake, the accumulated moment on the Longmanshan was enough to produce an  $M$  8 event (Wang et al., 2011). This situation is comparable to Utah’s Wasatch fault, where GPS shows similar motion and the paleoseismic record shows large earthquakes, although there is little present seismicity (Chang and Smith, 2002).

Another example is the convergent boundary between Africa and Eurasia in North Africa. The 1999 Global Seismic Hazard Map, which shows peak ground acceleration expected at 10% probability in 50 years, features a prominent hazard “bull’s-eye” at the site of the 1980  $M$  7.3 El Asnam earthquake. The largest subsequent earthquakes to date, the 2003  $M$  6.8 Algeria and 2004  $M$  6.4 Morocco events, did not occur in the bull’s-eye or regions designated as having high hazard levels (Fig. 7). Instead, they occurred in areas shown as having low hazard.

The 2010  $M$  7.1 Haiti earthquake similarly occurred on a fault mapped in 2001 as having low hazard, producing ground motion far greater than the map predicted (Fig. 8). This situation arose because the map was based on recent seismicity. A much higher hazard would be predicted by considering the long term earthquake history of faults in the area and GPS data showing strain accumulating across them (Manaker et al., 2008).

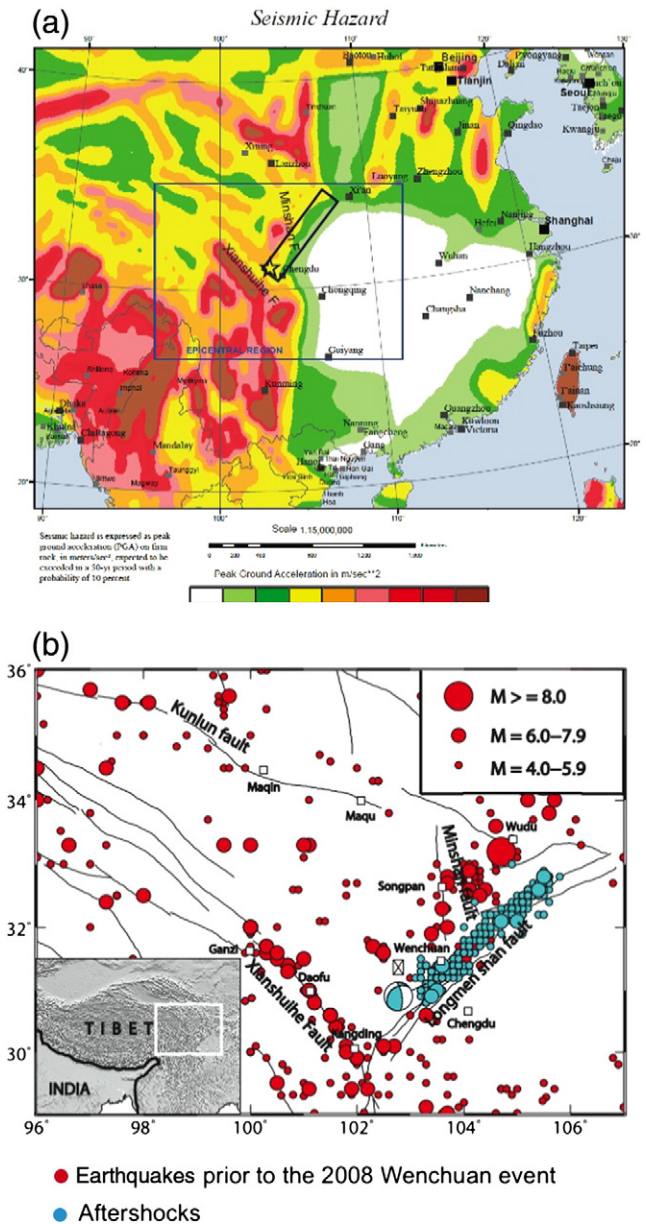
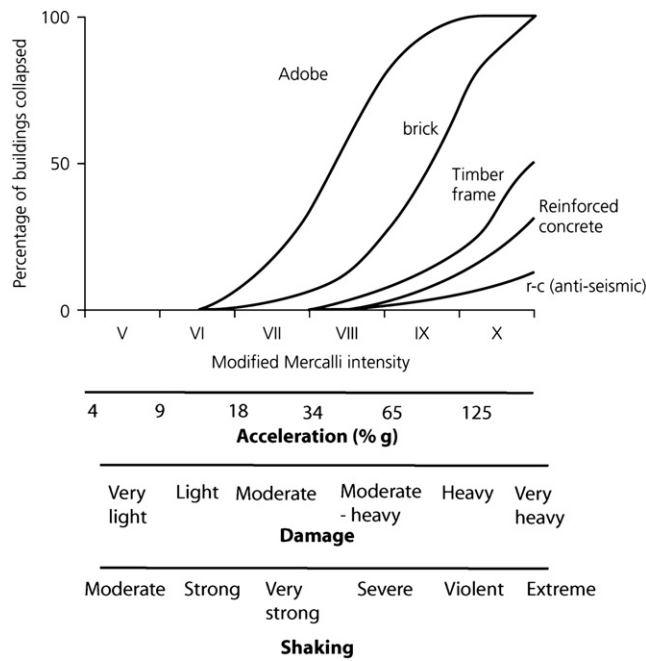


Fig. 5. (a) USGS seismic hazard map for China produced prior to the 2008 Wenchuan earthquake, which occurred on the Longmenshan Fault (black rectangle). (b) Seismicity in the region. Note that the hazard map showed low hazard on the Longmenshan fault, on which little instrumentally recorded seismicity had occurred before the Wenchuan earthquake, and higher hazard on faults nearby that showed more seismicity.

A different, but also illuminating, problem arose for the February 22, 2011  $M$  6.3 earthquake that did considerable damage in Christchurch, New Zealand. This earthquake, an aftershock of a  $M$  7.1 earthquake on a previously unrecognized fault, produced much stronger ground motion than the hazard map predicted would occur in the next 10,000 years (Reyners, 2011).

These examples illustrate that although hazard maps are used in many countries to provide a foundation for earthquake preparation and mitigation policies, they often underpredict or overpredict what actually happens. These difficulties are endemic to the hazard mapping process, rather than any specific group or nation’s implementation of it. As these examples show, the maps often depend dramatically on unknown and difficult-to-assess parameters, and hence on the mapmakers’ preconceptions. As a result, hazard maps often have large uncertainties that are often unrecognized and rarely communicated to public and other users.



**Fig. 6.** Approximate percentage of buildings that collapse as a function of the intensity of earthquake-related shaking. The survival of buildings differs greatly for constructions of weak masonry, fired brick, timber, and reinforced concrete (with and without anti-seismic design) (Stein and Wyssession, 2003).

In hindsight, the problem is that present hazard mapping practices were accepted and adopted without rigorous testing to show how well they actually worked. This acceptance was analogous to what took place since the 1970s in finance, where sophisticated mathematical models were used to develop arcane new financial instruments (Overbye, 2009). Few within the industry beyond their practitioners, termed “quants,” understood how the models worked. Nonetheless, as described by economist Fischer Black (Derman, 2004), their theoretical bases were “accepted not because it is confirmed by conventional empirical tests, but because researchers persuade one another that the theory is correct and relevant.” This wide acceptance was illustrated by the award in 1997 of the Nobel Prize in economics to Myron Scholes and Robert Merton for work based upon Black’s, who died a few years earlier. Only a year later, Long Term Capital Management, a hedge fund whose directors included Scholes and Merton, collapsed and required a \$3.6 billion bailout (Lowenstein, 2000). Unfortunately, this collapse did not lead to reassessment of the financial models, whose continued use in developing mortgage backed securities contributed significantly to the 2008 financial crisis (Stein, 2012). The Tohoku earthquake is a

similar indicator of difficulties with earthquake hazard maps. Hence our goal here is to review some of the factors causing these difficulties and propose approaches to address them.

## 5. Hazard map challenges

Making earthquake hazard maps is an ambitious enterprise. Given the complexities of the earthquake process and our limited knowledge of it, many subjective choices are needed to make a map. As a result, maps depend heavily on their makers’ preconceptions about how the earth works. When these preconceptions prove correct, a map fares well. When they prove incorrect or inadequate, a map does poorly. Predicting earthquake hazard has been described as playing “a game of chance of which we still don’t know all the rules” (Lomnitz, 1989). Not surprising, nature often wins. As in any game of chance, the best we can do is to maximize our expectation value by understanding the game as well as possible. The better we understand the limitations of the hazard maps, the better we can use them and hopefully improve them.

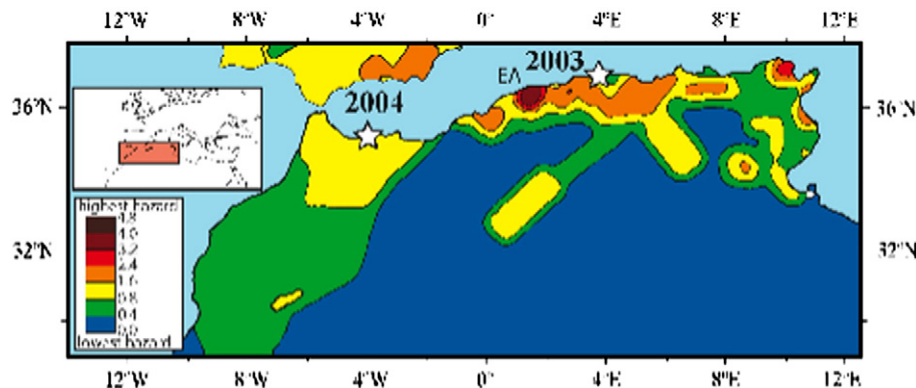
### 5.1. Forecasting and prediction

To understand the challenge of earthquake hazard mapping, it is useful to view it as an extension of unsuccessful attempts to predict earthquakes (Geller, 1997, 2011; Geschwind, 2001; Hough, 2009). In this paper, following Geller (1997) among others, we use “prediction” to refer to attempts to issue warnings of imminent earthquakes on a time scale of a few days or at most weeks, and “forecasting” for the much longer time scales involved in hazard maps.

In the 1960’s and 1970’s, prediction programs began in the U.S., China, Japan, and the USSR. These programs relied on two basic approaches. One was based on laboratory experiments showing changes in the physical properties of rocks prior to fracture, implying that earthquake precursors could be identified. A second was the idea of the seismic cycle, in which strain accumulates over time following a large earthquake. Hence areas on major faults that had not had recent earthquakes could be considered seismic gaps likely to have large earthquakes.

Optimism for prediction was high. For example, Louis Pakiser of the U.S. Geological Survey announced that if funding were granted, scientists would “be able to predict earthquakes in five years.” California senator Alan Cranston, prediction’s leading political supporter, told reporters that “we have the technology to develop a reliable prediction system already at hand.” Such enthusiasm led to well-funded major programs like the U.S. National Earthquake Hazards Reduction Program and Japan’s Large-Scale Earthquake Countermeasures Act.

In the U.S., prediction efforts culminated in 1985 when the USGS launched an official earthquake prediction experiment. Part of the



**Fig. 7.** Portion of Global Seismic Hazard Map (1999) for North Africa, showing peak ground acceleration in  $m/s^2$  expected at 10% probability in 50 years. Note prominent “bull-eye” at site of the 1980  $M_s$  7.3 El Asnam earthquake. The largest subsequent earthquakes to date, the May 2003  $M_s$  6.8 Algeria and February 2004  $M_s$  6.4 Morocco events (stars) did not occur in the predicted high hazard regions (Swafford and Stein, 2007).



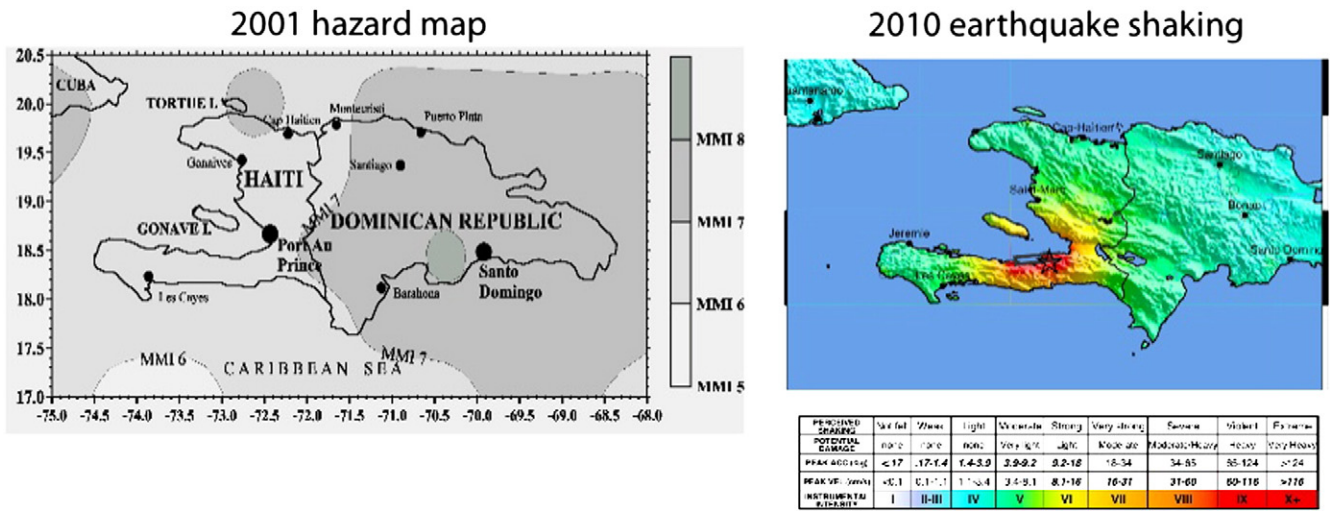


Fig. 8. Left: Seismic hazard map for Haiti produced prior to the 2010 earthquake ([http://www.oas.org/cdmp/document/seismap/haiti\\_dr.htm](http://www.oas.org/cdmp/document/seismap/haiti_dr.htm)) showing maximum shaking (Modified Mercalli Intensity) expected to have a 10% chance of being exceeded once in 50 years, or on average once about every 500 years. Right: USGS map of shaking in the 2010 earthquake.

San Andreas fault near Parkfield, California, had had magnitude 6 earthquakes about every 22 years, with the last in 1966. Thus the USGS predicted at the 95% confidence level that the next such earthquake would occur within five years of 1988, or before 1993. The USGS's National Earthquake Prediction Evaluation Council endorsed the prediction. Equipment was set up to monitor what would happen before and during the earthquake. The *Economist* magazine commented, "Parkfield is geophysics' Waterloo. If the earthquake comes without warnings of any kind, earthquakes are unpredictable and science is defeated. There will be no excuses left, for never has an ambush been more carefully laid."

Exactly that happened. The earthquake did not occur by 1993, leading *Science* magazine (Kerr, 1993) to conclude, "Seismologists' first official earthquake forecast has failed, ushering in an era of heightened uncertainty and more modest ambitions." Although a USGS review committee criticized "the misconception that the experiment has now somehow failed" (Hager et al., 1994), the handwriting was on the wall. An earthquake occurred near Parkfield in 2004, eleven years after the end of the prediction window, with no detectable precursors (Bakun et al., 2005). Thus searches for precursors did not result in a short-term prediction of the 2003 event. It is unclear whether the 2004 event should be regarded as "the predicted Parkfield earthquake" (although much too late) or merely a random earthquake (Jackson and Kagan, 2006; Savage, 1993).

Attempts elsewhere have had the same problem. Although various possible precursors have been suggested, no reliable and reproducible precursors have been found (Geller, 1997). For example, despite China's major national prediction program, no anomalous behavior was identified before the 2008 Wenchuan earthquake (Chen and Wang, 2010). Similarly, the gap hypothesis has not yet proven successful in identifying future earthquake locations significantly better than random guessing (Kagan and Jackson, 1991, 1995), and the "characteristic earthquake" model also fails to explain observed seismicity beyond random chance (Kagan, 1996).

As a result, seismologists have largely abandoned efforts to predict earthquakes on time scales less than a few years, although the ongoing government program in Japan to try to issue a prediction within three days of an anticipated magnitude 8 earthquake in the Tokai region is a notable exception (Geller, 2011). Instead, effort has turned to trying to make longer-term forecasts (e.g., Jordan et al., 2011; Lee et al., 2011). This approach melded with efforts initiated by earthquake engineers to develop earthquake hazard maps. The hazard maps are used in the formulation of seismic design maps, which are used in turn to develop building codes. Given that typical buildings have useful lives of 50–100 years, the long-term approach is natural.

### 5.2. Defining the hazard

Hazard maps predict the effects of future earthquakes of different magnitudes by assuming how likely areas are to have earthquakes and using ground motion attenuation relations to specify how shaking decreases with distance from the epicenter. The hazard in a given location is described by the maximum shaking due to earthquakes that are expected to happen in a given period of time.

Such maps can be made in various ways, whose relative advantages are debated (e.g., Bommer, 2009; Castanos and Lomnitz, 2002; Mucciarelli et al., 2008; Panza et al., 2010; Wang, 2011; Wang and Cobb, in press). One is to specify the largest earthquake of concern for each area. That means assuming where it will be, its magnitude, and how much shaking it will cause. This is called deterministic seismic hazard assessment or DSHA.

The most common approach is to consider all the possible earthquakes that could cause significant shaking at a place. This method, called probabilistic seismic hazard assessment or PSHA, involves estimating the probable shaking from the different earthquakes and producing an estimate of the combined hazard. PSHA uses the probabilities and uncertainties of factors like the location and times of earthquakes and how much shaking will result from an earthquake of a given magnitude. It was developed by Cornell (1968) and is widely applied in engineering design (e.g., McGuire, 1995). An overview of PSHA is given by Hanks and Cornell (1994), who note that "to the benefit of just about no one, its simplicity is deeply veiled by user-hostile notation, antonymous jargon, and proprietary software."

However the hazard is calculated, it is important to remember that earthquake hazard is not a physical quantity that can be measured. Instead, it is something that map makers define and then calculate using a set of computer algorithms. Thus, as the Japan example shows, a hazard map depends on its makers' assumptions. The predictions of maps made under various assumptions can be combined using a "logic tree," but they must still be subjectively weighted. Although the specifics vary for probabilistic and deterministic maps, many of the same uncertainties arise in either approach because they reflect limitations in knowledge of future earthquakes. For example, the Tohoku map underpredicted the shaking and tsunami, due to the assumption that M9 earthquakes would not happen. Each of the fault segments that broke was assumed to have maximum magnitude less than 8, and it was assumed that they would not fail together. Any map made with these assumptions would also have underpredicted what happened. Hence our goal here is not to argue for or against PSHA or DSHA, but to show how these limitations –

and thus in the parameters used in map making – produce uncertainties in hazard maps.

As a result, evaluating a hazard map and assessing how much confidence to place in it involves considering the assumptions made, how well constrained they are by data, and their effects on the map. A good way to do this is to look at the map's robustness – how do maps change due to map makers's assumptions. There are five main assumptions in making hazard maps: how, where, when, how big, and how strong?

### 5.2.1. How?

The most crucial issue in defining the hazard is the probability or time window, sometimes called return period, used. The simplest approach assumes that the probability  $p$  that earthquake shaking at a site will exceed some value in the next  $t$  years, assuming this occurs on average every  $T$  years, is

$$p = 1 - \exp(-t/T)$$

which is approximately  $t/T$  for  $t \ll T$  (Stein and Wyssession, 2003). Lower probabilities correspond to longer time windows. Thus shaking that there is a 10% chance of exceeding at least once in 50 years will occur on average once about every  $50/0.1 = 500$  years (actually 475 using the more accurate exponential). However shaking with a 2% chance of being exceeded in 50 years will occur on average only every  $50/0.02 = 2500$  (actually 2475) years.

The effect of the return period is illustrated in Fig. 9, contrasting hazard maps for the U.S. made in 1982 (Algermissen et al., 1982) and 1996 (Frankel et al., 1996). The maps show the hazard in terms of peak ground acceleration (PGA). The San Andreas fault system appears similarly in the two maps. Given that this area has many large earthquakes and is well studied, the high hazard mapped there seems sensible. In contrast, the seismic hazard in the New Madrid (central U.S.) seismic zone was increased from approximately 1/3 that of the San Andreas in the 1982 map to greater than the San Andreas' in the 1996 map. This change resulted largely from a change in the return period used to define the hazard for building codes. The older map shows the maximum shaking predicted to have a 10% chance of being exceeded at least once in 50 years, or once in 500 years, which is the return period used in most nations' maps. However, the new map shows the maximum shaking predicted to have a 2% chance of occurring at least once in 50 years, or on average at least once about every 2,500.

To see why using a longer return period increases the hazard, consider Fig. 10, which approximates the New Madrid seismic zone in the central U.S. Earthquakes can be thought of as darts thrown at the map, with the same chance of hitting anywhere in the area shown. About every 150 years a magnitude 6 earthquake hits somewhere, causing moderate shaking in an area that is assumed to be a circle with a radius of 50 km. Over time, more earthquakes hit and a larger portion of the area gets shaken at least once. Some places get shaken a few times. Thus the longer the time period the map covers, the higher

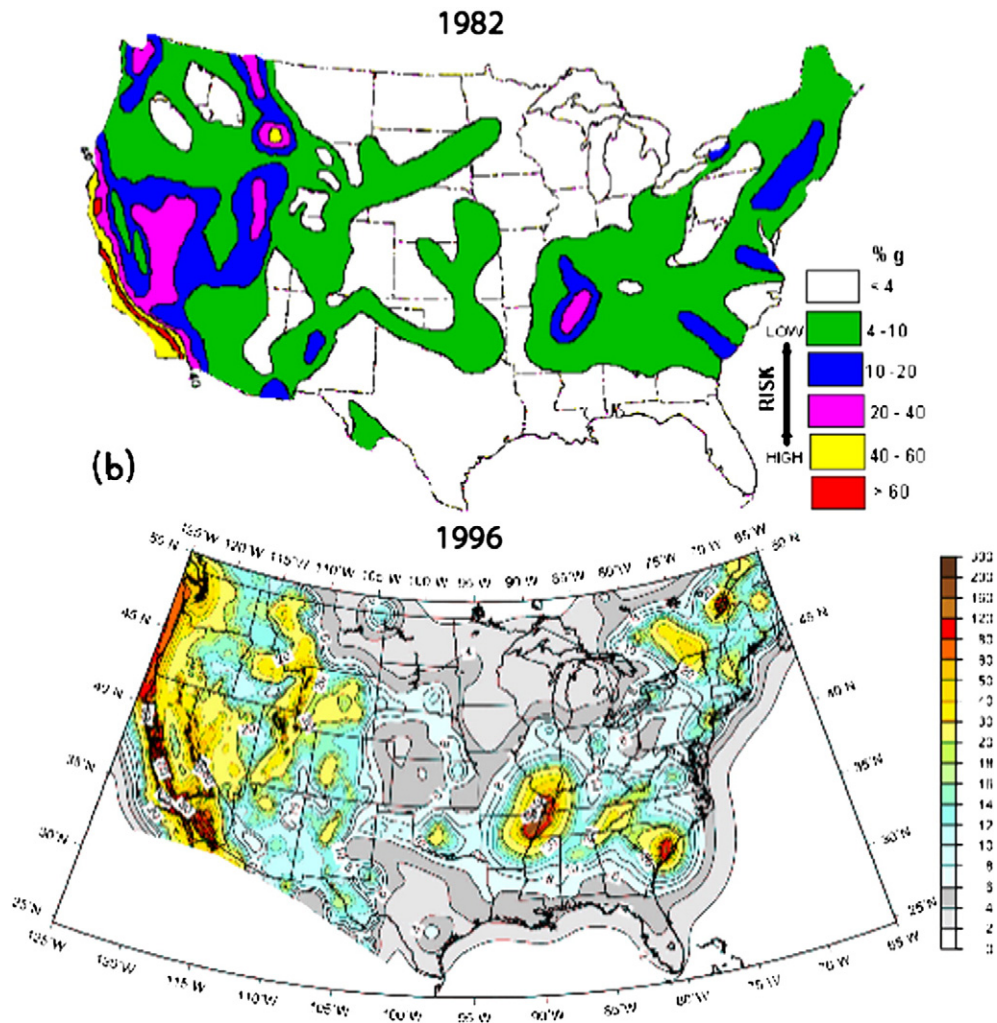


Fig. 9. Comparison of the 1982 and 1996 USGS earthquake hazard maps for the US. The predicted hazard is shown as a percentage of the acceleration of gravity. Redefining the hazard raised the predicted hazard in the Midwest from much less than in California to even greater than California's.

the predicted hazard. Whether constructing typical buildings (as opposed to highly critical structures like nuclear power plants) to withstand such rare shaking makes economic sense is debatable (Searer et al., 2007), given that typical buildings have a useful life of 50–100 years. The new definition in fact increased California's hazard to a level that would have made earthquake resistant construction too expensive, so the hazard there was "capped."

This example brings out the crucial point that the definition of hazard involves not only scientific, but also political and economic, issues. Redefining the hazard using a 2500-year return period rather than 500 years is a decision to require a higher level of seismic safety, via stronger building codes, at higher cost. This involves choosing how to allocate resources between earthquake safety and other societal needs (Stein, 2010). There is no right or unique answer. Making building codes weaker lowers costs, but allows construction that is less safe if a future earthquake occurs. More stringent codes impose higher costs on the community and use resources that might be better used otherwise. For example, communities have to balance putting steel in schools with hiring teachers. An important aspect of defining the hazard is that different levels may be appropriate for different structures. As shown by the Fukushima nuclear accident, extremely stringent hazard requirements appear appropriate for nuclear power plants (Nöggerath et al., 2011).

How successful the U.S. map will be over time – hundreds of years – is an interesting question. Given that California deforms at a rate 50 times faster than New Madrid, and is 30–100 more seismically active, treating the New Madrid hazard as comparable to California, as in the hazard map based on 2500 years return period (Fig. 9b), is likely to have vastly overestimated the hazard (Stein, 2010; Stein et al., 2003).

5.2.2. Where?

As the Tohoku, Wenchuan, and Haiti examples showed, the choice of where to assume large earthquakes will happen is crucial in hazard mapping. Fig. 11 illustrates this issue by comparing hazard maps for Canada made in 1985 and 2005. The older map shows concentrated high hazard bull's-eyes along the east coast at the sites of the 1929 M 7.3 Grand Banks and 1933 M 7.4 Baffin Bay earthquakes, assuming there is something especially hazardous about these locations. The alternative is to assume that similar earthquakes can occur anywhere along the margin, presumably on the faults remaining from the rifting (Stein et al., 1979, 1989). The 2005 map makes this assumption, and thus shows a "ribbon" of high hazard along the coast, while still retaining the bull's-eyes. Thus quite different maps result from using only the instrumentally recorded earthquakes, or including geological assumptions.

Only time will tell which map fared better. To see this, consider the simple simulation shown in Fig. 12. The simulation assumes that M 7 earthquakes occur randomly along the margin at the rate they have in the past 100 years. As shown, the simulations yield apparent concentrations of large earthquakes and seismic gaps for earthquake records up to thousands of years long. Approximately 8000–11,000 years of record is

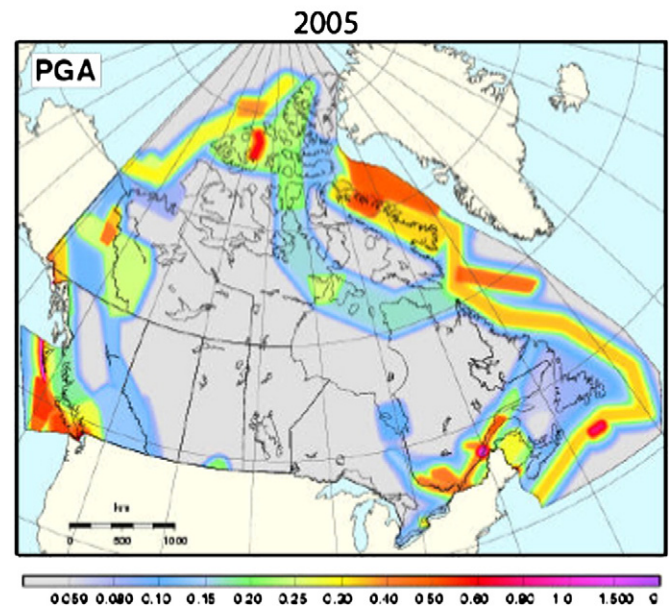
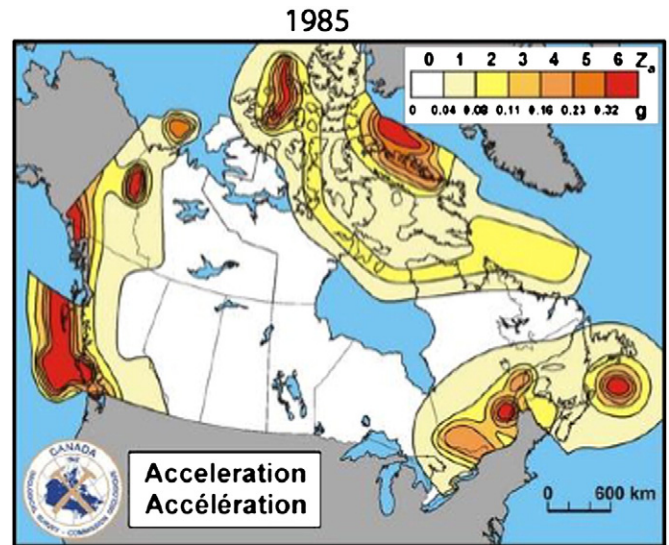


Fig. 11. Comparison of the 1985 and 2005 Geological Survey of Canada earthquake hazard maps. The older map shows concentrated high hazard bull's-eyes along the east coast at the sites of the 1929 Grand Banks and 1933 Baffin Bay earthquakes, whereas the new map assumes that similar earthquakes can occur anywhere along the margin.

needed to show that the seismicity is uniform. Any shorter sample – like that available today – would give a biased view. Hence if the seismicity and thus hazard are uniform, a hazard map produced using the seismic record alone will overestimate the hazard where previous large

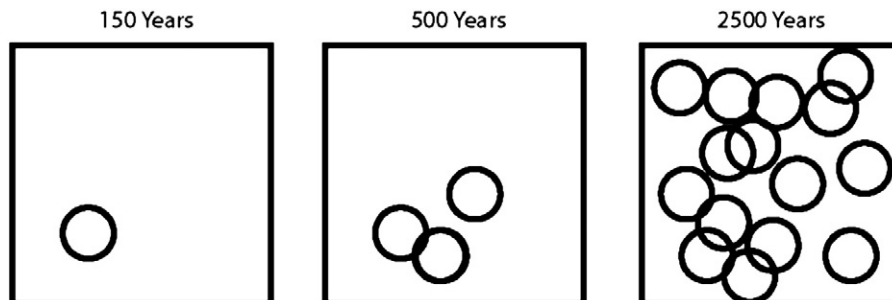


Fig. 10. Schematic illustration showing how the predicted earthquake hazard increases for longer time window. The circles show areas within which shaking above a certain level will occur (Stein, 2010).

earthquakes occurred and underestimate it elsewhere, as appears to be happening in North Africa (Fig. 7).

This continental margin example illustrates the challenge for what is essentially a one-dimensional seismic zone. The situation is even more complicated for plate boundary zones and mid-continent where the spatiotemporal patterns of seismicity are even more irregular. For example, although GPS data indicate ~3 mm/yr of extension across Utah's Wasatch front region, the strain pattern is quite diffuse. This led Friedrich et al. (2003) to suggest that strain might be distributed across a number of faults in the area, each one of which would have a small rate, and hence a smaller seismic hazard. On the other hand, Malservisi et al. (2003) suggest that the diffuse pattern of GPS strain can be fit by having all strain accommodated on the main Wasatch fault, but with the fault late in its earthquake cycle. This difference illustrates the importance of including detailed knowledge about the fault in question and the regional tectonic setting in hazard assessment.

In plate interiors, the zones over which seismicity is distributed can be even larger. A 2000-year record from North China shows migration of large earthquakes between fault systems spread over a broad region, such that no large earthquake ruptured the same fault segment twice in this interval (Fig. 13). Hence a map made using any short subset of the record would be biased. For example, a map using the 2000 years earthquake record prior to 1950 would miss the recent activity in the North China plain, including the 1976 Tangshan earthquake (Mw 7.8), which occurred on a previously unknown fault and killed nearly 240,000 people.

The China example illustrates the difficulty in hazard assessment for continental interiors, where such variable fault behavior is being widely recognized (Camelbeeck et al., 2007; Clark et al., 2011; Crone et al., 2003; Newman et al., 1999; Stein et al., 2009). In many places large earthquakes cluster on specific faults for some time and then migrate to others. Some faults that appear inactive today, such as the Meers Fault in Oklahoma, have clearly been active within the past few thousand years. Thus mid-continent faults “turn on” and

“turn off” on timescales of hundreds or thousands of years, causing episodic, clustered, and migrating large earthquakes, and making it even harder to assess the extent to which the seismological record reflects the location of future earthquakes. Moreover, while the locations of past small earthquakes are good predictors of the locations of future ones (Kafka, 2007), they can be poor indicators for where future large earthquakes will occur because intraplate earthquakes can have aftershock sequences that last hundreds of years or even longer. As a result, many small earthquakes may be aftershocks of previous large quakes (Stein and Liu, 2009). In such cases, treating small earthquakes as indicating the location of future large ones can overestimate the hazard in presently active areas and underestimate it elsewhere.

Consequently, some countries are changing from hazard maps based only on historical and recorded seismicity to ones including geological data, which predict lower and more diffuse hazard (Fig. 14). An interesting example contrasting U.S. and Canadian approaches by comparing map predictions near the border is given by Halchuk and Adams (1999). As in all such things, the mappers' choices reflect their view of how seismicity in the areas in question works. As discussed in the section about map testing, a long time will typically be needed to tell how good each view was.

### 5.2.3. When?

After assuming where earthquakes will happen, hazard mappers then have to assume when they will happen. As discussed earlier, the Japanese government map (Fig. 1) reflected the assumption that the Nankai Trough area was the area most likely to have a major earthquake soon. How to describe this possibility is difficult. As Fig. 3e shows, the recurrence history in the area is a complicated function of space and time, whose underlying dynamics is not understood.

Fig. 15, showing a longer earthquake history for the Nankai A-B segment reported by Ishibashi (1981), illustrates that even on this pair of segments the recurrence time is quite variable, with a mean

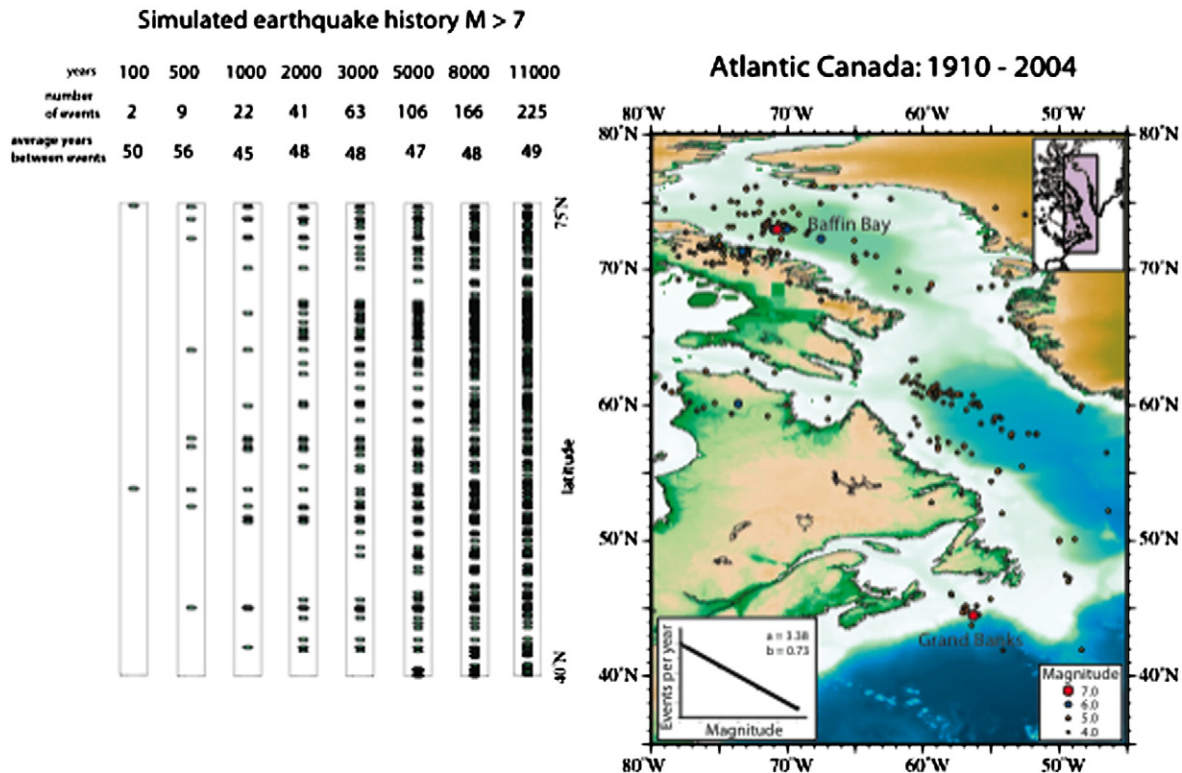
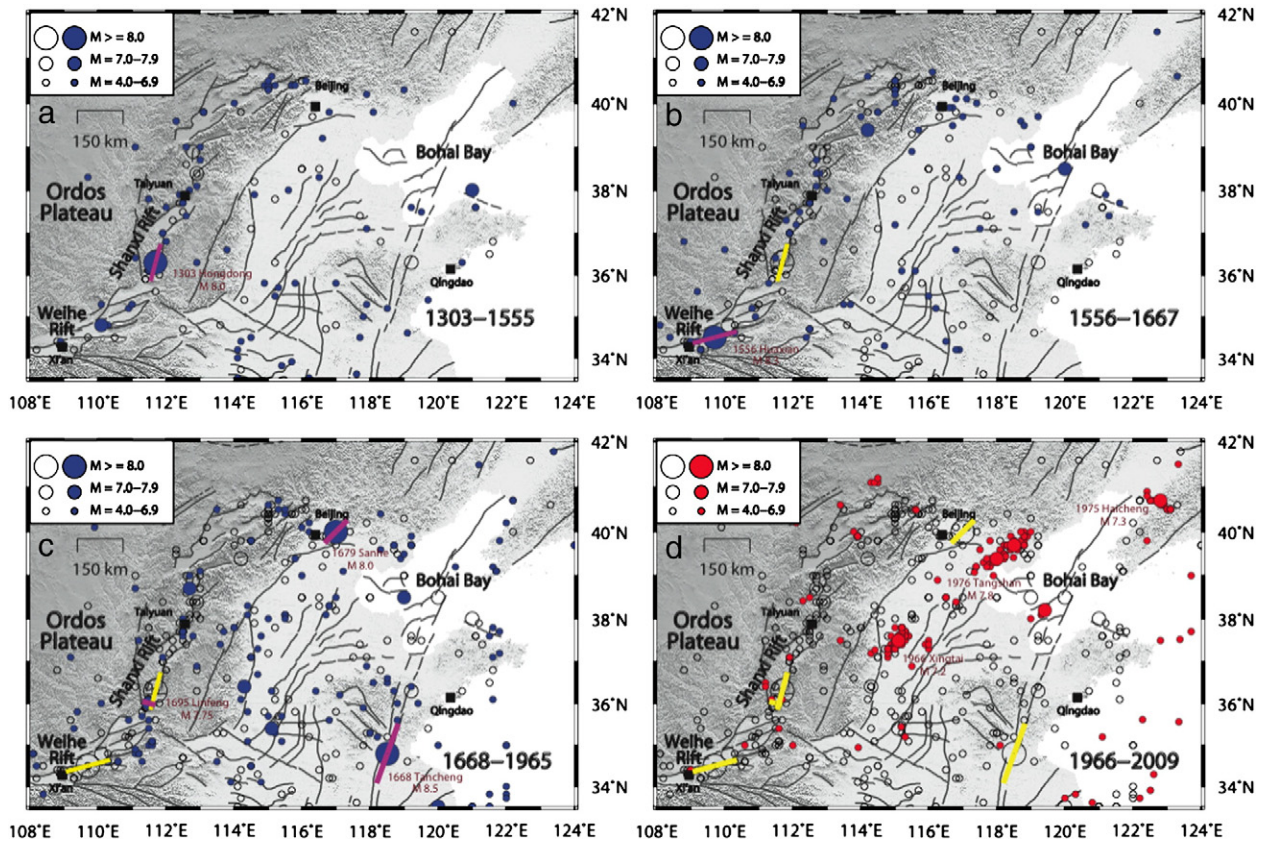


Fig. 12. Intraplate seismicity along the eastern coast of Canada. Simulations using a frequency-magnitude relation derived from these data predict that if seismicity is uniform in the zone, about 11,000 years of record is needed to avoid apparent concentrations and gaps. (Swafford and Stein, 2007).



**Fig. 13.** Earthquake history of North China, showing that seismicity has migrated such that no fault segment has ruptured twice in 2000 years. Solid circles are locations of events during the period shown in each panel; open circles are circles are the location of events from 780 BCE to the end of the previous period (1303 CE for panel A). Bars show the rupture lengths for selected large events (Liu et al., 2011).

of 180 years and a standard deviation of 72 years. Moreover, which segments break at the same time is also quite variable (Ishibashi and Satake, 1998).

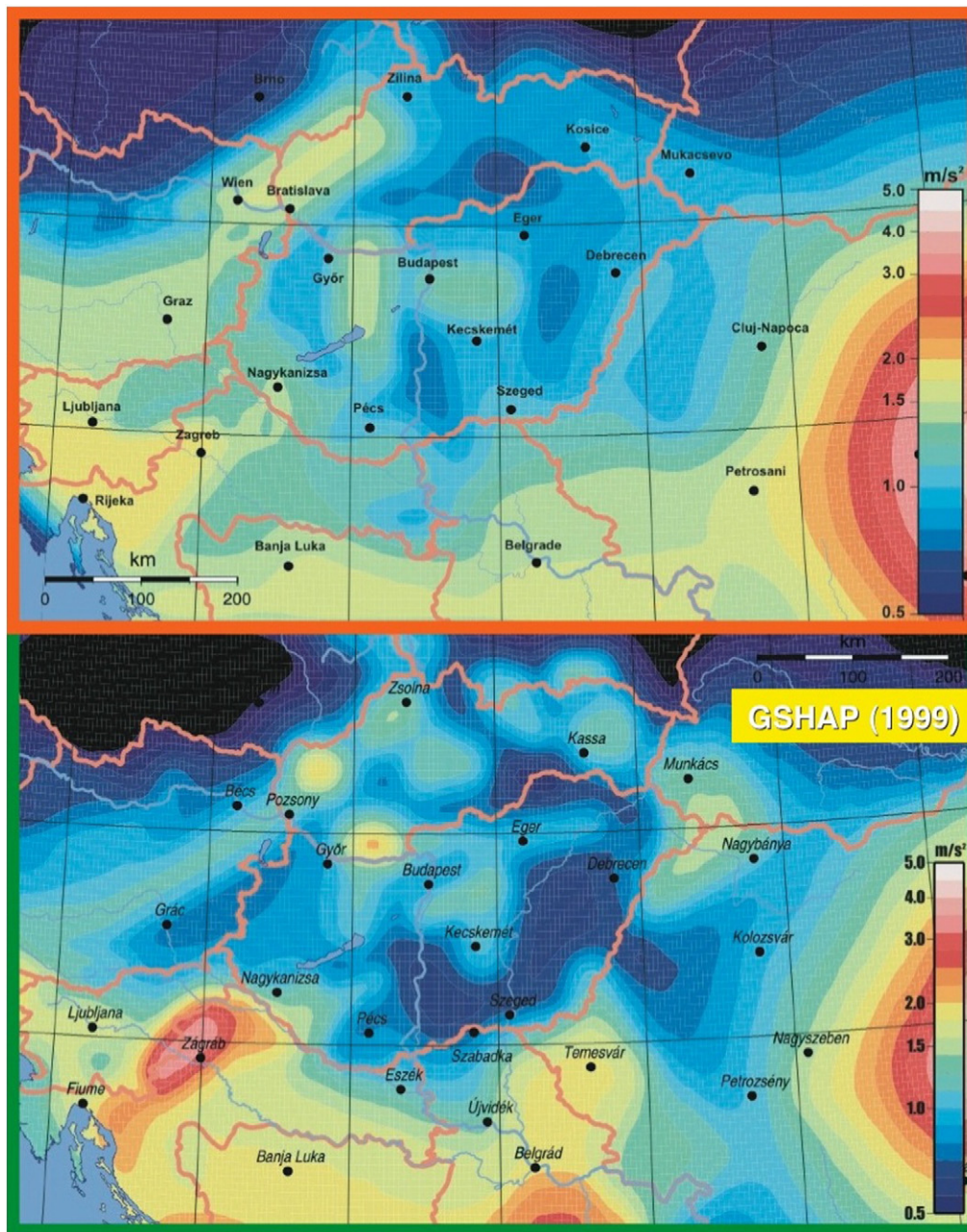
Such variability is the norm, as illustrated in Fig. 15 by the paleo-earthquake record at Pallett Creek, on the segment of the San Andreas that broke in the 1857 Fort Tejon earthquake (Sieh et al., 1989). The nine recurrence intervals have a mean of 132 years and a standard deviation of 105 years. This large variability results from the presence of several clusters of large earthquakes, which together with the observational uncertainties, make it difficult to characterize the sequence and estimate earthquake probabilities. Hence Sieh et al. (1989)'s estimates of the probability of a similar earthquake before 2019 ranged from 7% to 51%. Moreover, using different subsets of the series will yield different results (Stein and Newman, 2004). As a result, the actual variability is greater than inferred from studies that use short earthquake sequences, typically 2–4 recurrences (Nishenko and Buland, 1987).

Hazard mapping requires assuming some probability density function that describes the distribution of future earthquake recurrence intervals. However, as these examples illustrate, even long paleoseismic and historic earthquake records often cannot resolve the probability density function very well (Parsons, 2008a; Savage, 1991, 1992, 1994).

A crucial choice among the possible probability density functions is between ones representing two models of earthquake recurrence (Stein and Wysession, 2003). In one, the recurrence of large earthquakes is described by a time-independent Poisson process that has no “memory.” Thus a future earthquake is equally likely immediately after the past one and much later, so earthquakes often cluster in time. Under this assumption, the probability that an earthquake will occur in the next  $t$  years is approximately  $t/T$ , where  $T$  is the assumed mean recurrence time, and an earthquake cannot be “overdue.”

The alternative is to use some time-dependent recurrence model in which a probability distribution describes the time between earthquakes. In this model earthquakes are quasi-periodic, with the standard deviation of recurrence times small compared to their mean. In such models the conditional probability of the next large earthquake, given that it has not yet happened, varies with time. The probability is small shortly after the past one, and then increases with time. For times since the previous earthquake less than about 2/3 of the assumed mean recurrence interval, time-dependent models predict lower probabilities. Eventually, if a large earthquake has not occurred by this time, the earthquake is “overdue” in the sense that time-dependent models predict higher probabilities. Fig. 16 (top) illustrates this effect for faults on which the time since the past large earthquake is a different fraction of the assumed mean recurrence interval.

The difference is shown in Fig. 16 (bottom) for the New Madrid seismic zone, where M 7 earthquakes occurred in 1811–1812. Assuming that such earthquakes occur on average every 500 years, the probability of having one in the next 50 years is 50/500 or 10% in a time-independent model. Alternative models assume that recurrence times have Gaussian distributions with a mean of 500 years and a standard deviation of 100 or 200 years. The “crossover” time is the year 2144, 333 years after 1811. The time-dependent models predict that the probability of a major earthquake in the next hundred years is much smaller than does the time-independent model. Similar results arise for a time-dependent model with recurrence times described by a lognormal probability distribution. An important aspect of using a time-dependent model is that it requires choosing a larger number of parameters, in that both the model used for recurrence times and its parameters are needed, but are usually poorly constrained by the available earthquake history (Parsons, 2008a; Savage, 1991, 1992, 1994).



**Fig. 14.** Alternative hazard maps for Hungary and the surrounding area, showing peak ground acceleration in  $\text{m/s}^2$  expected at 10% probability in 50 years (Swafford and Stein, 2007). The GSHAP model, based only on historic and recorded seismicity (bottom), predicts more concentrated hazard near sites of earlier earthquakes, compared to a model (top) including geological data that predicts more diffuse hazard (Toth et al., 2004).

Despite the crucial difference between the time-independent and time-dependent models, it remains unresolved which one better describes earthquake recurrence. Seismological instincts favor earthquake cycle models, in which strain builds up slowly from one major earthquake and so gives quasi-periodic events, as observed on some faults (Parsons, 2008b). However, other studies find clustered events (Kagan and Jackson, 1991). As a result, both types of models are used in hazards mapping, often inconsistently. For example, the U.S. Geological Survey uses time-independent models for New Madrid, whereas their maps for California increasingly use time-dependent models (Field et al., 2008; Petersen et al., 2007).

The effect of the model choice on a hazard map is illustrated in Fig. 17 by alternative maps for the New Madrid zone. The biggest effect is close to the three faults used to model the jagged geometry of the earthquakes of 1811–1812, where the largest hazard is predicted.

These are assumed to have a moment magnitude ( $M_w$ ) of 7.3 (Hough et al., 2000).

Compared to the hazard predicted by the time-independent model, the time-dependent model predicts noticeably lower hazard for the 50-year periods 2000–2050 and 2100–2150. For example in Memphis, the time-dependent model predicts hazards for 2000–2050 and 2100–2150 that are 64% and 84% of those predicted by the time-independent model. However if a large earthquake has not occurred by 2200, the hazard predicted in the next 50 years would be higher than predicted by the time-independent model.

Given the limitations of knowledge, choosing how to model the recurrence on faults in a hazard map largely reflects the mappers' preconceptions. Thus the Japan map (Fig. 1) reflected the mappers' view that a large earthquake would happen much sooner on the Nankai Trough than off Tohoku.

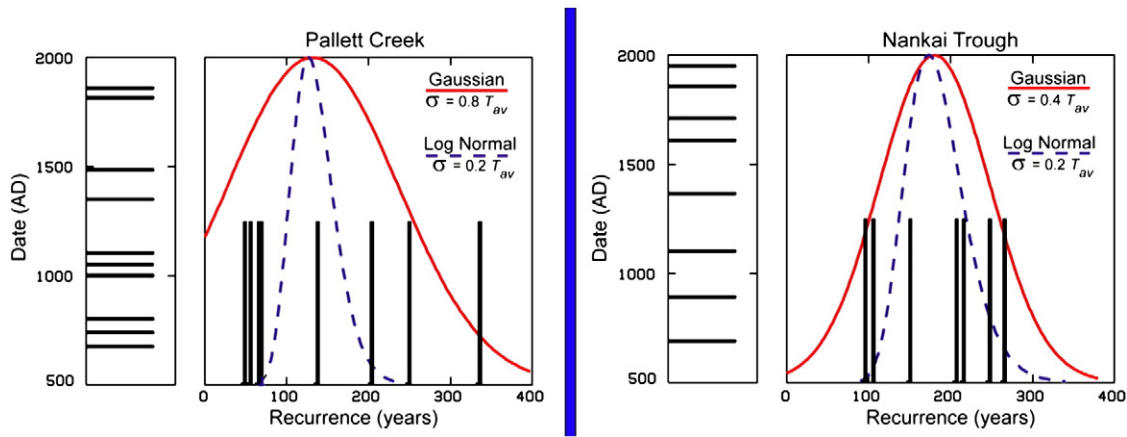


Fig. 15. Observed variability in recurrence intervals for two long sequences of large earthquakes (Stein and Newman, 2004).

5.2.4. How big?

Hazard maps depend also dramatically on the assumed magnitude of the largest earthquakes expected in each area. Because this is unknown and not at present effectively predictable on physical grounds, mappers predict the size and rate of future large earthquakes from an earthquake record containing seismological, geological, and historical earthquake records. The Tohoku, Wenchuan, and Haiti examples indicate that this process is far from straightforward and prone to a variety of biases.

Some biases arise from the assumption used in these analyses that earthquake recurrence approximately follows a log-linear, or Gutenberg–Richter, relation,

$$\log N = a - bM,$$

with  $b \approx 1$ , such that the logarithm of the annual number ( $N$ ) of earthquakes above a given magnitude ( $M$ ) decreases linearly with magnitude (Gutenberg and Richter, 1944). However, studies of specific areas, which commonly address the short history of seismological observations by combining seismological data for smaller earthquakes with paleoseismic data or geologic inferences for larger earthquakes, sometimes infer that large earthquakes occur more or less commonly than expected from the log-linear frequency-magnitude relation observed for smaller earthquakes (Fig. 18).

These deviations are important in seismic hazard assessment. In particular, higher seismic hazard is assumed when larger earthquakes are presumed to be more common than expected from the small earthquakes. Such large earthquakes are termed “characteristic earthquakes” (Schwartz and Coppersmith, 1984). This usage is somewhat confusing, in that other applications (mentioned earlier) use this term for the “typical” largest earthquake on a fault (Jackson and Kagan, 2006). By analogy, large earthquakes less common than expected from the small earthquakes could be termed “uncharacteristic.”

This problem is common, because earthquake records in many regions are shorter than would be adequate to show the true size and rate of the largest earthquakes. This is especially so when the length of the earthquake record under consideration is comparable to the mean recurrence time of large earthquakes.

This possibility is illustrated by simple numerical simulations, which involved generating synthetic earthquake histories of various lengths assuming that the seismicity followed a log-linear frequency-magnitude relation. The recurrence times of earthquakes with  $M \geq 5, 6,$  and  $7$  were assumed to be samples of a Gaussian (normal) parent distribution with a standard deviation of  $0.4$  times the mean recurrence for each of the three magnitudes.

Fig. 19 (top) shows the results for 10,000 synthetic earthquake sequences whose length is half of  $T_{av}$ , the mean recurrence time of earthquakes with  $M \geq 7$ . The left panel shows the actual log-linear frequency-magnitude relation and dots marking the “observed” mean recurrence rates in the simulation for  $M \geq 5, 6,$  and  $7$  events in each sequence. The center panel shows the parent distribution of recurrence times for  $M \geq 7$  that was sampled, and a histogram of the “observed” mean recurrence times for the sequences. Apparent characteristic (more frequent than expected) earthquakes, for which the observed recurrence time is less than  $T_{av}$ , plot above the log-linear

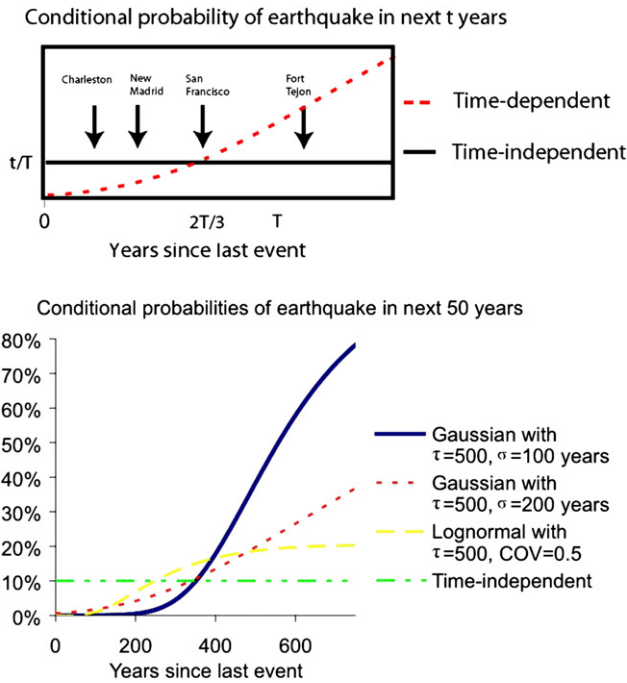


Fig. 16. Top: Schematic comparison of time-independent and time-dependent models for different seismic zones. Charleston and New Madrid are “early” in their cycles, so time-dependent models predict lower hazards. The two model types predict essentially the same hazard for a recurrence of the 1906 San Francisco earthquake, and time-dependent models predict higher hazard for the nominally “overdue” recurrence of the 1857 Fort Tejon earthquake. The time-dependent curve is schematic because its shape depends on the probability distribution and its parameters. Bottom: Comparison of the conditional probability of a large earthquake in the New Madrid zone in the next 50 years, assuming that the mean recurrence time is 500 years. In the time-independent model the probability is 10%. Because the time since 1811 is less than  $2/3$  of the assumed mean recurrence interval, time-dependent models predict lower probabilities of a large earthquake for the next hundred years (Hebden and Stein, 2009).

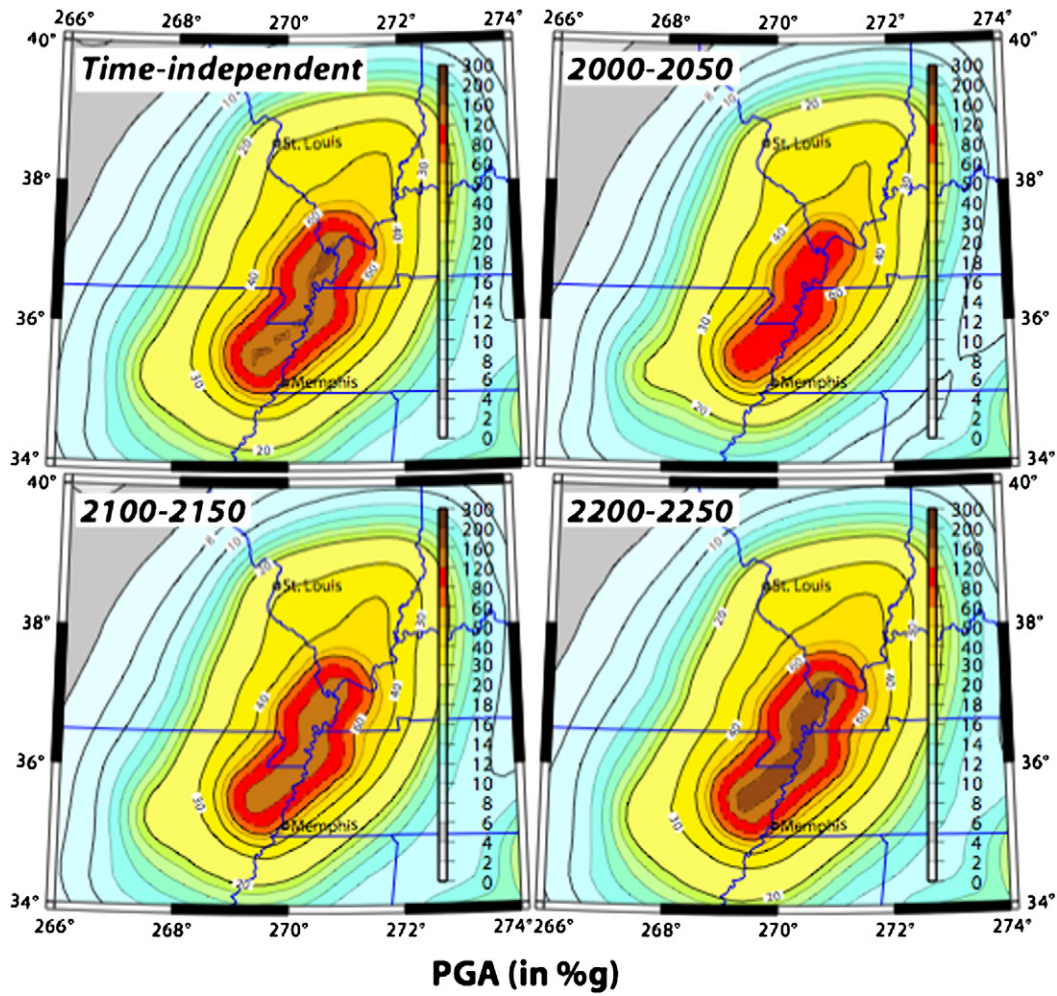


Fig. 17. Comparison of hazard maps for the New Madrid zone. Colors show peak ground acceleration as percentages of 1 g. Compared to the hazard predicted by the time-independent model, the time-dependent model predicts noticeably lower hazard for the 50-year periods 2000–2050 and 2100–2150, but higher hazard if a large earthquake has not occurred by 2200 (Hebden and Stein, 2009).

frequency-magnitude relation in the left panels, and to the left of 1 in the center panels. Due to their short length, 46% of the sequences contain no earthquakes with  $M \geq 7$ , 52% have only one, all but one of the

remaining 2% have two earthquakes, and one has three. The mean inferred recurrence times for sequences with one, two, or three earthquakes are  $T_{av}/2$ ,  $T_{av}/4$ , and  $T_{av}/6$ . Earthquakes with mean recurrence

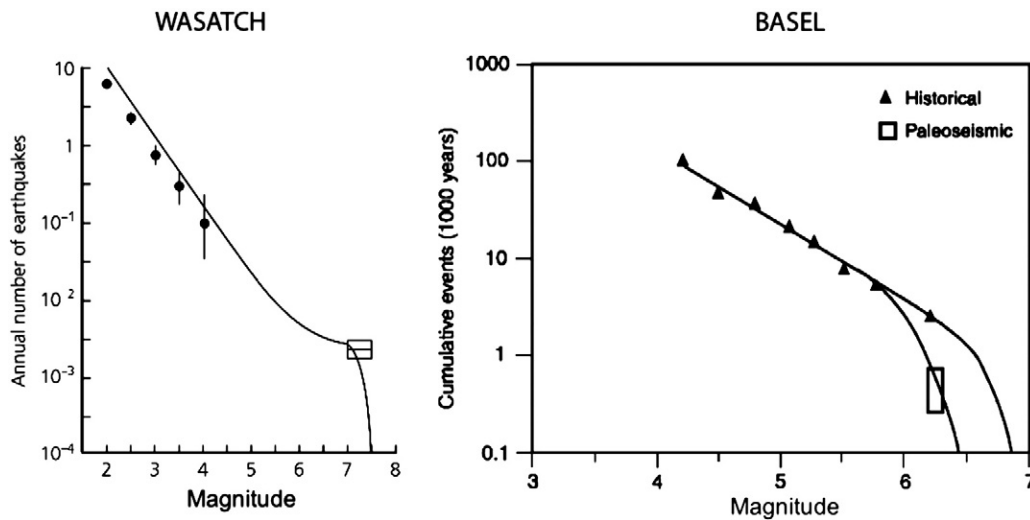
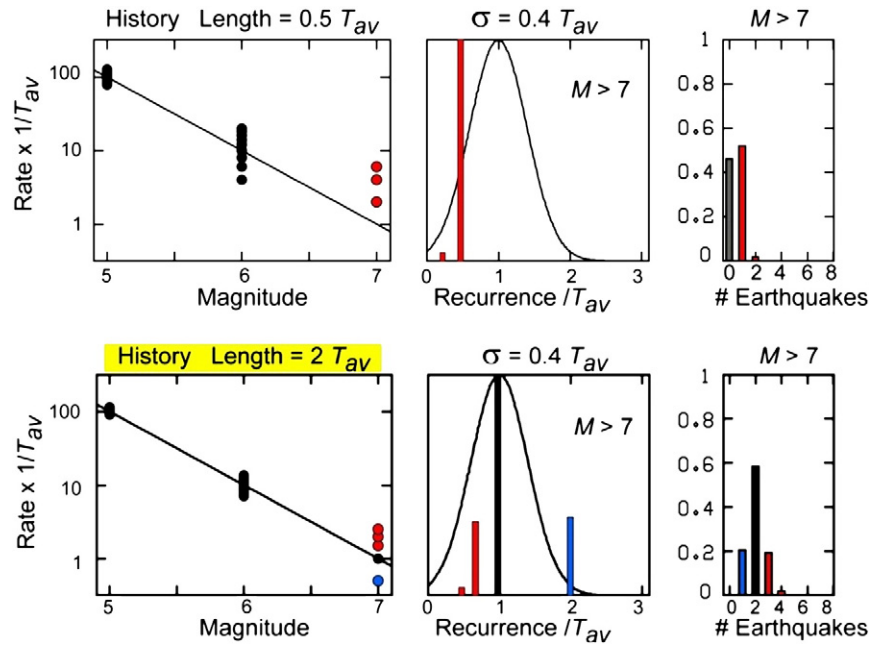


Fig. 18. Frequency-magnitude plots for various sets of earthquake data. Left: Seismological (dots) and paleoseismic (box) data for the Wasatch fault (Youngs and Coppersmith, 1985), showing large earthquakes more common than expected from the small ones. Right: Historical and paleoseismic data for the greater Basel (Switzerland) area (Meghraoui et al., 2001). The paleoseismic data imply that large earthquakes occur at a lower rate than predicted from smaller ones (Stein and Newman, 2004).





**Fig. 19.** Results of numerical simulations of earthquake sequences. Rows show results for sequences of different lengths. Left panels show the log-linear frequency–magnitude relation sampled, with dots showing the resulting mean recurrence times and the observed mean recurrence times (bars). Right panels show the fraction of sequences in which a given number of  $M \geq 7$  earthquakes occurred. In each panel, red and blue circles and bars represent characteristic and uncharacteristic earthquakes, respectively. The grey bar in the upper right panel shows where no  $M \geq 7$  events occurred, thus not appearing in the left and center panels (Stein and Newman, 2004).

interval greater than or equal to  $T_{av}$  are not observed in a sequence half that length.

Hence, due to the short sampling interval, in about half the cases we “observe” characteristic earthquakes, whereas in the other half no large earthquakes are observed. This is because we cannot observe half an earthquake. Hence we either overestimate the rate of the largest earthquakes and thus the seismic hazard, or underestimate the size of the largest earthquakes that can occur and thus underestimate the hazard.

Similar biases persist for longer sequences (Fig. 19, bottom). For example, if the sequence length is twice  $T_{av}$ , we underestimate the rate of the largest earthquakes 20% of the time and overestimate it another 20% of the time.

Another bias can arise from spatial sampling. Fig. 20 shows frequency–magnitude data for Utah’s Wasatch Front area. Here, the rate of small instrumentally recorded earthquakes is consistent with the paleoseismically inferred rate of large earthquakes. However, data from only the part of the area around the Wasatch Fault show a different pattern, in that the rate of small earthquakes underpredicts that of large paleo-earthquakes. This difference arises because the larger earthquakes occur on the fault, whereas the smaller ones occur all over the front. Hence whether one infers the presence of characteristic earthquakes depends on the sampling region.

Other biases can arise from the challenge of estimating the magnitude and rate of historic earthquakes and paleoearthquakes. Apparent characteristic earthquakes could occur if the magnitudes of historic earthquakes or paleoearthquakes were overestimated. Conversely, the mean recurrence time  $T_{av}$  would be overestimated if some paleoearthquakes in a series were not identified. These issues arise in many cases, including inferring the size of paleoearthquakes from tsunami deposits.

### 5.2.5. How much shaking?

Hazard mapping also has to assume how much shaking future earthquakes will produce. This involves adopting a ground motion attenuation relation, which predicts the ground motion expected at a given distance from earthquakes of a given size. These relations reflect the combined effects of the earthquake source spectrum and

propagation effects including geometric spreading, crustal structure, and anelastic attenuation.

Fig. 21 illustrates various ground motion models for the central U.S. In general, the predicted shaking decreases rapidly with distance for a given magnitude. This effect occurs in most places, as illustrated by the fact that much less damage resulted from shaking in the giant Tohoku and Sumatra earthquakes than from the tsunamis, because the earthquakes occurred offshore. The models also show how a smaller earthquake nearby can cause more damage than a larger one farther away.

In some areas, seismological data can be used to develop models describing average ground motion as a function of distance, about which actual motions scatter owing to variations in crustal structure (e.g., Burger et al., 1987) and source properties. However, in many areas, including the central U.S., there are no seismological records of shaking from large earthquakes. In such cases, mappers choose between various relations derived using data from smaller earthquakes and earthquake source models, which predict quite different ground motion and thus hazard.

For example, the relation of Frankel et al. (1996) predicts significantly higher ground motion than those of Atkinson and Boore (1995) and Toro et al. (1997). In fact, the ground motion for an M 7 earthquake predicted by the Frankel et al. (1996) relation at distances greater than 100 km is comparable to that predicted for an M 8 earthquake by the other relation. This difference occurs both for Peak Ground Acceleration (PGA) and 1 Hz motion, a lower-frequency parameter more useful than PGA in describing the hazard to major structures. The differences are greater for the lower frequency because the predicted shaking is more sensitive to the differences in the assumed source spectra.

The effect of these choices is shown in Fig. 22 by four possible maps. The two in each row are for the same ground motion model, but different values of the maximum magnitude – the magnitude of the largest earthquake on the main faults. Raising this magnitude from 7 to 8 increases the predicted hazard at St. Louis by about 35%. For Memphis, which is closer to the main faults, the increase is even greater. This is because the assumed maximum magnitude of the largest earthquake on the main faults affects the predicted hazard especially near those faults.

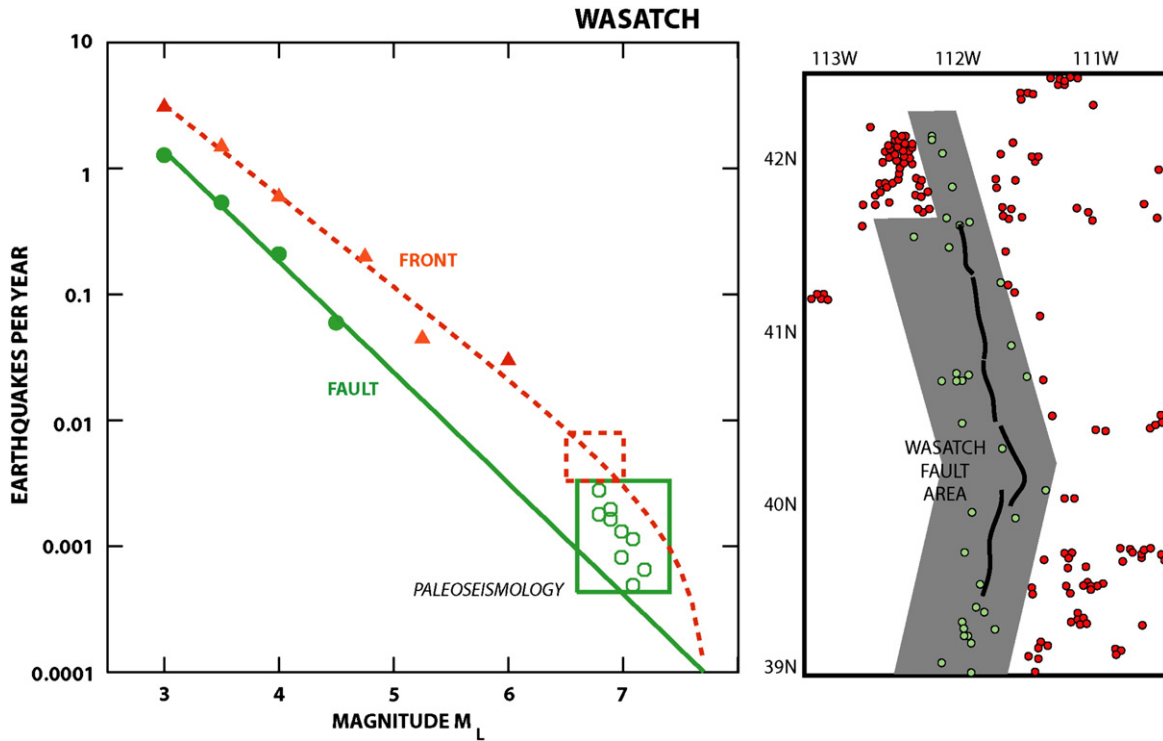


Fig. 20. Illustration of the effect of spatial sampling on frequency–magnitude curves (Stein et al., 2005). Left: Comparison of instrumental and paleoseismic earthquake recurrence rate estimates for the Wasatch seismic zone. For the Wasatch front triangles are binned earthquake rates, dashed line is fit to them, and dashed box is estimated range from paleoearthquakes (Pechmann and Arabasz, 1995). For the Wasatch Fault closed circles are binned earthquake rates (Chang and Smith, 2002 and pers. comm.), solid line is fit to them, closed circles are from paleoearthquakes, and solid box gives their range. Instrumental data for the front are consistent with the paleoseismic results and so do not imply the presence of characteristic earthquakes, whereas those for the fault underpredict the paleoseismic rate and so imply the presence of characteristic earthquakes. Right: Comparison of seismicity and paleoseismicity sampling areas for the Wasatch front (entire map area) and Wasatch fault (grey area). Solid line denotes Wasatch fault (After Chang and Smith, 2002).

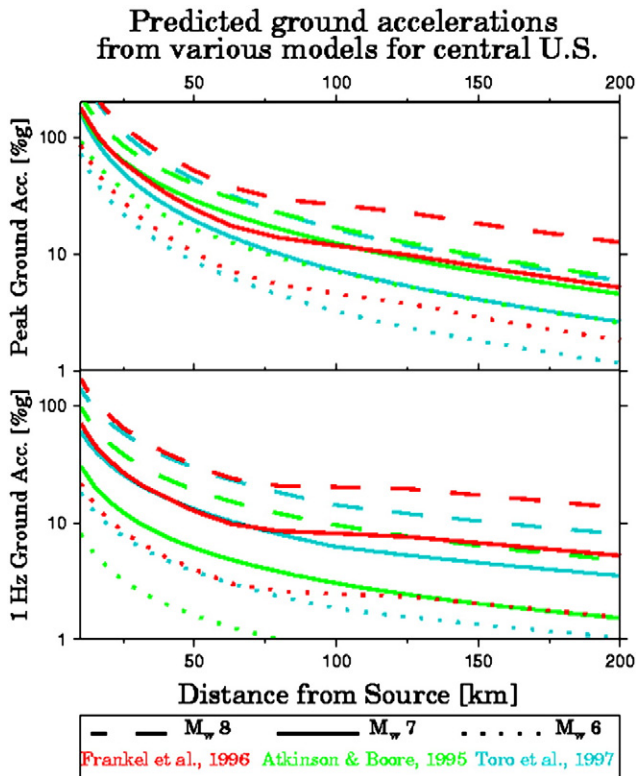


Fig. 21. Comparison of ground motion (peak ground acceleration and 1 Hz) as a function of distance for different earthquake magnitudes predicted by three attenuation relations for the central U.S. For  $M_w$  8, the Frankel et al. (1996) relation predicts significantly higher values than the others (Newman et al., 2001).

The two maps in each column have the same maximum magnitude but different ground motion models. The Frankel et al. (1996) predicts a hazard in St. Louis about 80% higher than that predicted by the Toro et al. (1997) model. For Memphis, this increase is about 30%. The ground motion model affects the predicted hazard all over the area, because shaking results both from the largest earthquakes and from smaller earthquakes off the main faults.

These models assume time-independent recurrence for the largest earthquakes, following USGS practice. As shown in Fig. 17, assuming time-independent recurrence would lower the predicted hazard.

The different maps give quite different views of the hazard, which is the subject of ongoing debate. A key parameter in the debate has been what to assume for the magnitude of a future large earthquake, which is assumed to be similar to the large events of 1811–1812. Analyses of historical accounts give values ranging from  $M$  7 (Hough et al., 2000; Nuttli, 1973) to  $M$  8 (Johnston, 1996), with recent studies (Hough and Page, 2011) favoring the lower values.

The maps in Fig. 22 also indicate that the choice of the ground motion model is as significant as that of the maximum magnitude. Thus without good constraining data, the predicted hazard depends significantly on the mappers' assumptions. Frankel et al. (1996) state that “workshop participants were not comfortable” with the source model used in the Atkinson and Boore (1995) ground motion relation, and so “decided to construct” a new relation, which predicted higher ground motions. They similarly considered but did not use an alternative recurrence distribution because it “produced substantially lower probabilistic ground motions.”

## 6. What to do

The Tohoku, Wenchuan, and Haiti earthquakes show that hazard mapping has many limitations and a long way to go. In any given area, additional research will improve hazard mapping, as more

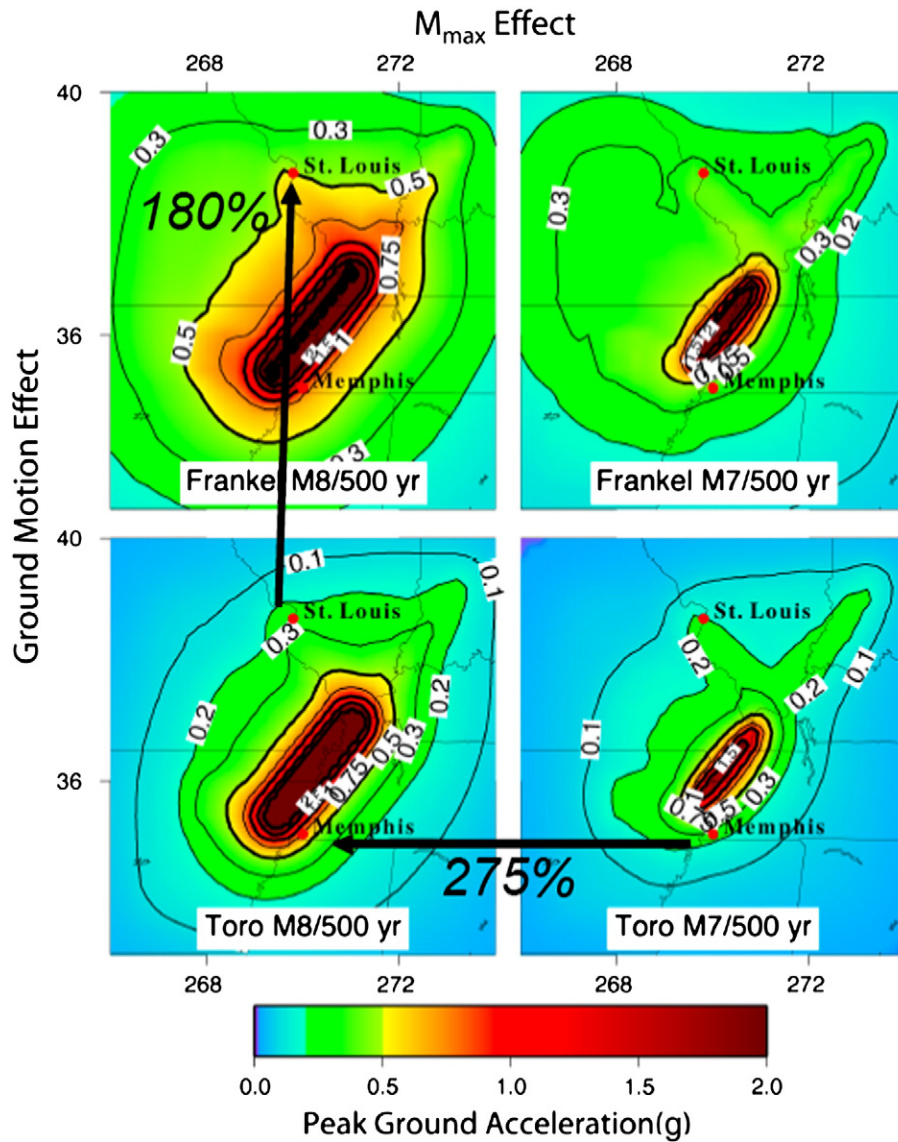


Fig. 22. Comparison of the predicted hazard (2% probability in 50 years) showing the effect of different ground motion relations and maximum magnitudes of the New Madrid fault source (Newman et al., 2001).

data are acquired from paleoearthquake records, geodesy, and other approaches. Modeling of fault processes will also help. However, we still do not know how to effectively use these data for anything beyond relatively general forecasts. For example, even had the GPS data showing strain accumulation off Tohoku been appreciated, there was no good way to forecast how large an earthquake might occur or how soon. This limitation is illustrated by the fact that communities inland from the Nankai Trough are now being warned of much larger tsunamis than previously anticipated, assuming that a future earthquake could be as large as March's Tohoku earthquake (Cyranoski, 2012). These communities face the challenge of deciding what to do for a possible 20-meter tsunami whose probability cannot be usefully estimated beyond saying it would be rare, perhaps once in a millennium.

Despite the limitations in our ability to assess future hazards, earthquake hazard mapping has grown into a large and widely-used enterprise worldwide. Although much of the approach in making hazard maps sounds sensible, fundamental advances in knowledge – which may or may not be achievable – will be needed for it to become fully effective and reliable. At this point, hazard maps in many areas seem inadequate for making multi-billion dollar decisions. We need to know whether earthquake recurrence is time dependent or

independent, or whether there is a reasonably regular recurrence interval for some faults. We need to know what controls fault segmentation, and why segments rupture sometimes singly and other times multiply. We need to know whether seismic activity on faults within continents remains steady for long periods, or switches on and off. We need reliable ways to predict strong ground shaking in areas where there are no such seismological records. Absent these advances, we should not be surprised that hazard maps often do poorly.

However, hazard mapping was developed by engineers out of necessity. Buildings must be built, so some seismic safety requirements need to be specified. In many active areas, such as California, the hazard maps seem sensible and in reasonable accord with experience. However, a sense of humility and caution is also required, as some seismologists in Japan might have made similar statements before the Tohoku earthquake. Kanamori's (2011) advice that we should "be prepared for the unexpected" is wise counsel.

In assessing the state of hazard mapping, an engineer might consider the glass half full, whereas it seems half empty to a seismologist. Engineers operate in a world of legal requirements. To avoid liability, they need a recognized standard, whatever its flaws. Depending on circumstances, this can create explicit or implicit pressure on hazard

mappers to arrive at either lower or higher values. The existence of the map allows engineers to get on with their work, and failures only become undeniable when “unexpected” earthquakes occur (e.g., March 11, 2011 Tohoku earthquake). The non-occurrence of “expected” earthquakes (e.g., Tokai in Japan) has not become an issue, because they can always be regarded as “impending” no matter how long we wait. In contrast, to a scientist the map failures are interesting because they point out what we do not understand and would like to.

Because many of hazard mapping's limitations reflect the present limited knowledge about earthquakes and tectonics, we anticipate advances both from ongoing studies and new methods (e.g. Newman, 2011). Still, we suspect that many of the challenges are unlikely to be resolved soon, and some may be inherently unresolvable. Hence in addition to research on the seismological issues, we suggest two changes to current hazard mapping practices.

### 6.1. Assess and present uncertainties

Hazard maps clearly have large uncertainties. When a map fails, it is often clear in hindsight that key parameters were poorly estimated. The sensitivity analyses shown in this paper illustrate the same point – the maps are uncertain in the sense that their predictions vary significantly depending on the choice of many poorly known parameters.

Assessing and communicating these uncertainties would make the maps more useful. At present, most users have no way to tell which predictions of these maps are likely to be reasonably well constrained, and which are not. Having this information would help users make better decisions. For example, estimates of the uncertainty can be used to develop better mitigation strategies (Stein and Stein, 2012).

The case for such an approach has been eloquently articulated in other applications. Perhaps the most famous is in Richard Feynman's (1988) report after the loss of the space shuttle Challenger: “NASA owes it to the citizens from whom it asks support to be frank, honest, and informative, so these citizens can make the wisest decisions for the use of their limited resources.”

Relative to hazard predictions, Sarewitz et al. (2000) argue similarly that

*“Above all, users of predictions, along with other stakeholders in the prediction process, must question predictions. For this questioning to be effective, predictions must be as transparent as possible to the user. In particular, assumptions, model limitations, and weaknesses in input data should be forthrightly discussed. Institutional motives must be questioned and revealed... The prediction process must be open to external scrutiny” Openness is important for many reasons but perhaps the most interesting and least obvious is that the technical products of predictions are likely to be “better” – both more robust scientifically and more effectively integrated into the democratic process – when predictive research is subjected to the tough love of democratic discourse... Uncertainties must be clearly understood and articulated by scientists, so users understand their implications. If scientists do not understand the uncertainties – which is often the case – they must say so. Failure to understand and articulate uncertainties contributes to poor decisions that undermine relations among scientists and policy makers.”*

Useful lessons for seismology can come from the atmospheric sciences. Meteorologists are much more candid in predicting hazards due to weather, which in the U.S. causes about 500 deaths per year compared to about 20 per year due to earthquakes. One key is comparing predictions by different groups using different assumptions. For example, on February 2, 2000 the Chicago Tribune weather page stated:

*“Weather offices from downstate Illinois to Ohio advised residents of the potential for accumulating snow beginning next Friday. But forecasters were careful to communicate a degree of uncertainty on the*

*storm's precise track, which is crucial in determining how much and where the heaviest snow will fall. Variations in predicted storm tracks occur in part because different computer models can infer upper winds and temperatures over the relatively data-sparse open Pacific differently. Studies suggest that examining a group of projected paths and storm intensities – rather than just one – helps reduce forecast errors.”*

The newspaper's graphics compared four models' predicted storm tracks across the Midwest and seven precipitation estimates for Chicago. This nicely explained the models' uncertainties, their limitations due to sparse data, and the varying predictions.

Hurricane researchers also often clearly explain what they can and cannot do. As Hurricane Irene threatened the U.S. East Coast, Emanuel (2011) explained to the public that “We do not know for sure whether Irene will make landfall in the Carolinas, on Long Island, or in New England, or stay far enough offshore to deliver little more than a windy, rainy day to East Coast residents. Nor do we have better than a passing ability to forecast how strong Irene will get. In spite of decades of research and greatly improved observations and computer models, our skill in forecasting hurricane strength is little better than it was decades ago.” The article described the causes of this uncertainty and approaches being taken to address it.

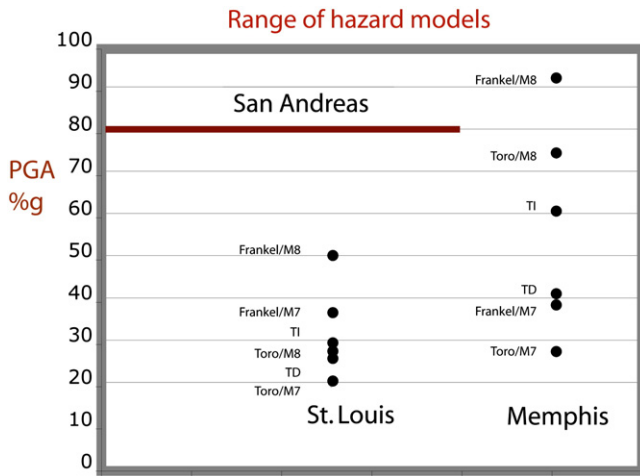
Such candor is also common in climate modeling, in which the results of a suite of different models developed by different groups using different methods and assumptions are typically presented and discussed. For example, the Intergovernmental Panel on Climate Change (2007) report compares the predictions of all 18 available models for the expected raise in global temperature, showing a factor of six variation. It further notes that the models “cannot sample the full range of possible warming, in particular because they do not include uncertainties in the carbon cycle. In addition to the range derived directly from the multi-model ensemble, Figure 10.29 depicts additional uncertainty estimates obtained from published probabilistic methods using different types of models and observational constraints.”

Similar presentations should be done for seismic hazard maps. One approach is to contrast crucial predictions. For example, Fig. 23 compares the predictions of the models in Figs. 17 and 22 for the hazard at Saint Louis and Memphis. The predictions vary by a factor of more than three. This representation shows the effects of the three factors. At Memphis, close to the main faults, the primary effect is that of magnitude, with the two M 8 models predicting the highest hazard. At Saint Louis, the ground motion model has the largest effect, so the Frankel models predict the highest hazard. Most models show hazard well below that predicted for California. The predictions for a maximum magnitude of 7 are similar to ones in which the large earthquake sequence has ended and the hazard reflects continuing aftershocks (Stein, 2010).

One of the most crucial reasons for discussing and presenting uncertainties is that experience in many applications shows that the real uncertainty is usually much greater than assumed. The general tendency toward overconfidence is shown by the fact that in many applications 20–45% of actual results are surprises, falling outside the previously assumed 98% confidence limits (Hammit and Shlyakhter, 1999). A famous example is the history of measurements of the speed of light, in which new and more precise measurements are outside the estimated error bars of the older ones much more frequently than expected (Henrion and Fischhoff, 1986). This effect has been observed in predicting river floods (Merz, 2012) and, as we have seen, is likely occurring for earthquake ground motion.

### 6.2. Test hazard maps

Clearly, we need a way to judge hazard maps' performance. Currently, there are no generally agreed criteria. A basic principle of science is that methods should only be accepted after they are shown



**Fig. 23.** Comparison of the hazard at St. Louis and Memphis predicted by hazard maps of the New Madrid zone shown in Figs. 17 and 22. For example, Frankel/M8 indicates the Frankel et al. (1996) ground motion model with a maximum magnitude of 8 in Fig. 22, and TI indicates the time-independent model in Fig. 17.

to be significantly more successful than ones based on null hypotheses, which are usually based on random chance. Otherwise, they should be rejected, regardless of how appealing their premises might seem.

Results from other fields, such as evidence-based medicine, which objectively evaluates widely used treatments, are instructive. For example, Moseley et al. (2002) found that although more than 650,000 arthroscopic knee surgeries at a cost of roughly \$5,000 each were being performed each year, a controlled experiment showed that “the outcomes were no better than a placebo procedure.”

Weather forecasts, which are conceptually similar to earthquake hazard mapping, are routinely evaluated to assess how well their predictions matched what actually occurred (Stephenson, 2000). A key part of this assessment is adopting agreed criteria for “good” and “bad” forecasts. Murphy (1993) notes “it is difficult to establish well-defined goals for any project designed to enhance forecasting performance without an unambiguous definition of what constitutes a good forecast.” Forecasts are also tested against various null hypotheses, including seeing if they do better than using the average of that date in previous years, or assuming that today’s weather will be the same as yesterday’s. Over the years, this process has produced measurable improvements in forecasting methods and results, and yielded much better assessment of uncertainties. Although the process of testing is easier than for earthquakes, because forecasts can be tested on shorter timescales, it illustrates how such testing can be done.

The recent examples of large earthquakes producing shaking much higher than predicted by the hazard maps indicate the need for an analogous process. This would involve developing objective criteria for testing such maps by comparison to the shaking that actually occurred after they were published. Such testing would show how well the maps worked, give a much better assessment of their true uncertainties, and indicate whether or not changes in the methodology over time resulted in improved performance.

Various metrics that reflect both overpredictions and underpredictions – neither of which are desirable – could be used. A natural one is to compare the maximum acceleration observed over the years in regions within the hazard map to that predicted by the map and by some null hypotheses. This could be done via the skill score used to test weather forecasts (Murphy, 1988). It would consider a region divided into  $N$  subregions. In each subregion  $i$ , over some time interval, we would compare the maximum observed shaking  $x_i$  to the map’s

predicted maximum shaking  $p_i$ . We then compute the Hazard Map Error

$$HME(p, x) = \sum_i (x_i - p_i)^2 / N$$

and assess the map’s skill by comparing it to the misfit of a reference map produced using a null hypothesis

$$HME(r, x) = \sum_i (x_i - r_i)^2 / N$$

using the skill score

$$SS(p, r, x) = 1 - HME(p, x) / HME(r, x).$$

The skill score would be positive if the map’s predictions did better than those of the map made with the null hypothesis, and negative if they did worse. We could then assess how well maps have done after a certain time, and whether successive generations of maps do better.

One simple null hypothesis is that of a regionally uniformly distributed seismicity or hazard. Fig. 1 suggests that the Japanese hazard map is performing worse than such a null hypothesis. Another null hypothesis is to start with the assumption that all oceanic trenches have similar  $b$ -value curves (Kagan and Jackson, in press) and can be modeled as the same, including the possibility of an M9 earthquake (there is about one every 20 years somewhere on a trench).

The idea that a map including the full detail of what is known about an area’s geology and earthquake history may not perform as well as assuming seismicity or hazard are uniform at first seems unlikely. However, it is not inconceivable. An analogy could be describing a function of time composed of a linear term plus a random component. A detailed polynomial fit to the past data describes them better than a simple linear fit, but can be a worse predictor of the future than the linear trend (Fig. 24). This effect is known as overparameterization or overfitting. A test for this possibility would be to smooth hazard maps over progressively larger footprints. There may be an optimal level of smoothing that produces better performing maps.

It is important to test maps using as long a record as possible. As a result, a major challenge for such testing is the availability of only a relatively short earthquake shaking record. A number of approaches could be used to address this issue. One would be to jointly test maps from different areas, which might give more statistically significant results than testing maps from individual areas.

It is also crucial to have agreed upon map testing protocols. When maps are published, their authors should specify the protocol they will use to test the map’s performance. In addition, testing should be also conducted by other groups using other methods. For this purpose, datasets of maximum shaking in areas over time should be compiled and made publicly available. Similarly, it is important to avoid biases due to new maps made after a large earthquake that earlier maps missed (Fig. 25). Statisticians refer to such a *posteriori* changes to a model as “Texas sharpshooting,” in which one first shoots at the barn and then draws circles around the bullet holes. In some cases assessing whether and how much better a new map predicts future events than an older one may take a while – sometimes hundreds of years – to assess.

Until recently, such testing was rare, although the value of such studies was illustrated by Kagan and Jackson (1991). Recently, however, interest has been growing. In addition to initial tests of hazard model performance (Albarelo and D’Amico, 2008; Kossobokov and Nekrasova, 2012; Miyazawa and Mori, 2009; Stirling and Gerstenberger, 2010; Wyss et al., 2012) there are ongoing projects to test earthquake predictability (CSEP, 2011; Kagan and Jackson, 2012; Marzocchi and Zechar, 2011; Schorlemmer et al., 2010). Because this process is just beginning, the available hazard map tests have limitations. Some use

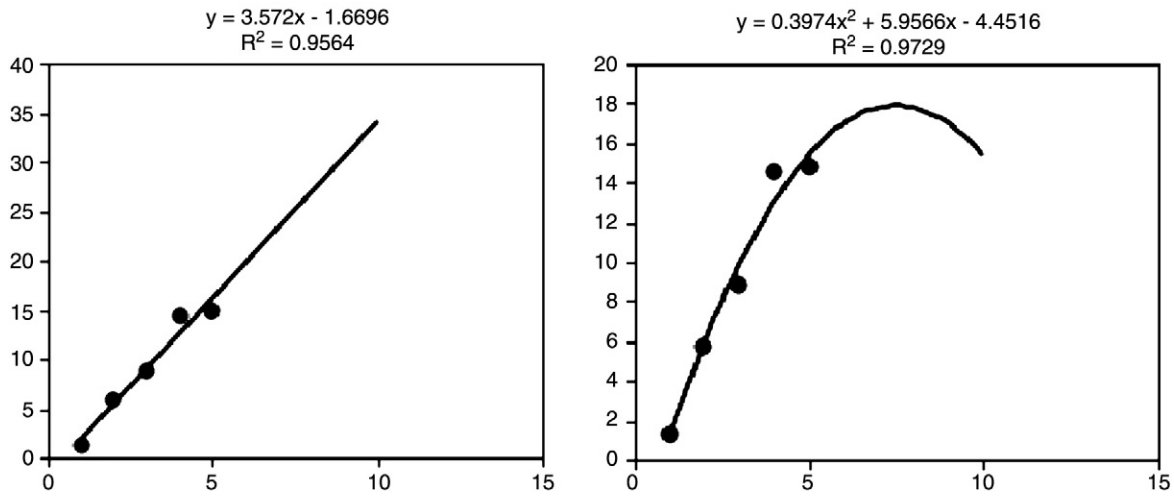


Fig. 24. Illustration of overfitting by comparison of linear and quadratic fits to a set of data. The quadratic gives a better fit to the points but a poorer representation of the trend.

shaking data from earthquakes before (rather than after) the map was made. This process favors a map in that even if these data were not explicitly used in the map making, the mappers were aware of these earthquakes' dates, locations, magnitude, and shaking. Similarly, because the mapping process involves a large number of subjective assumptions, it is difficult to use part of a data set to develop a map and test it with the other part. Conversely, tests using only data from large earthquakes after a map was made are biased against the map, because they do not include areas where little or no shaking occurred. In years to come, more comprehensive tests should be forthcoming. Until then, we have no way of knowing how well or poorly hazard maps are performing.

Hypothesis testing is the heart of the scientific method. Notwithstanding the difficulties, it is essential that a continuing process of serious and objective testing be conducted for the methods used to produce seismic hazard maps.

## 7. Mission impossible?

The present state of hazard mapping reflects the general paradox that humans desire to predict the future so strongly that we are reluctant to ask how well or poorly predictions do. This tendency, including the example in the epigram, is explored by Dan Gardner's (2010) book "Future Babble: Why expert predictions fail and why we believe them anyway." Pilkey and Pilkey-Jarvis (2006) show many examples

of models in the earth sciences that remain in common use despite repeated failures.

An interesting assessment is given by historian of science Naomi Oreskes' (2000) analysis:

*"As individuals, most of us intuitively understand uncertainty in minor matters. We don't expect weather forecasts to be perfect, and we know that friends are often late. But, ironically, we may fail to extend our intuitive skepticism to truly important matters. As a society, we seem to have an increasing expectation of accurate predictions about major social and environmental issues, like global warming or the time and place of the next major hurricane. But the bigger the prediction, the more ambitious it is in time, space, or the complexity of the system involved, the more opportunities there are for it to be wrong. If there is a general claim to be made here, it may be this: the more important the prediction, the more likely it is to be wrong."*

With these caveats in mind, it is worth noting that truly successful hazard maps would have to be very good. Ideally, they would neither underpredict the hazard, leading to inadequate preparation, nor overpredict it, unnecessarily diverting resources. Advances in our knowledge about earthquakes and objective testing of successive generations of hazard maps should improve their performance.

However, there are almost certainly limits on how well hazard maps can ever be made. Some are imposed by lack of knowledge

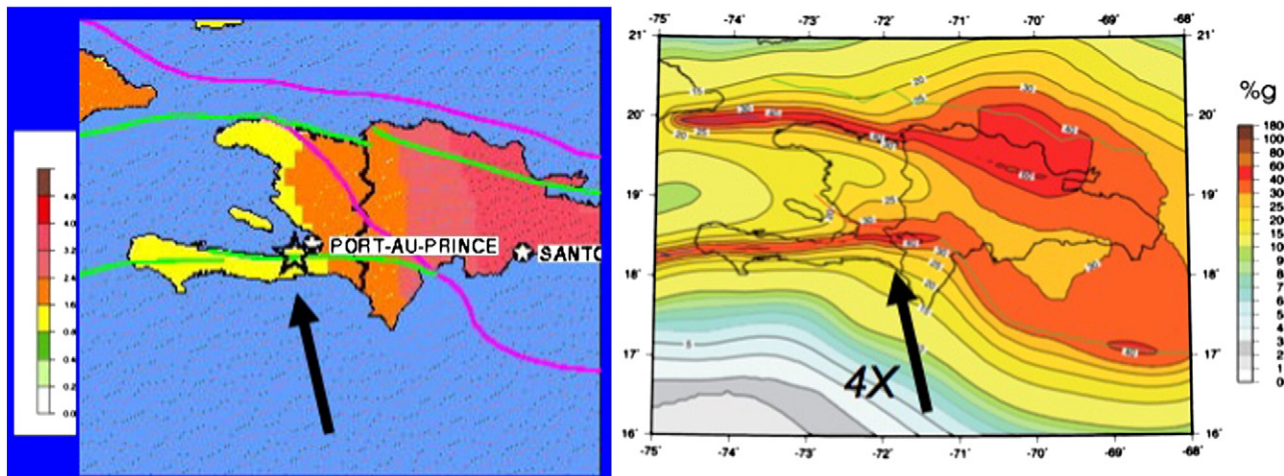


Fig. 25. Comparison of seismic hazard maps for Haiti made before (GSHAP, 1999) and after (Frankel et al., 2010) the 2010 M 7.1 earthquake. The newer map shows a factor of four higher hazard on the fault that had recently broken in the earthquake.

and the intrinsic variability of earthquake processes. As Kanamori (2011) notes in discussing why “the 2011 Tohoku earthquake caught most seismologists by surprise,” “even if we understand how such a big earthquake can happen, because of the nature of the process involved we cannot make definitive statements about when it will happen, or how large it could be.”

A similar caveat comes from the fact that inferring earthquake probabilities, which is crucial for such mapping, is very difficult (Parsons, 2008a,b; Savage, 1992, 1994). Freedman and Stark (2003) explore the question in depth, and conclude that estimates of earthquake probabilities and their uncertainties 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, unlike the models for coin-tossing, 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 estimates.” Savage (1991) similarly concluded that earthquake probability estimates for California are “virtually meaningless.” Although this situation presumably will improve somewhat as longer paleoseismic histories and other data become available, it will remain a major challenge.

Other limitations may reflect the fact that maps are produced on the basis of postulates, such as the seismic cycle model. If these models fundamentally diverge from the actual non-linear physics of earthquake occurrence – as may well be the case – then no amount of tweaking and tuning of models can produce hazard maps that come close to the ideal of predicting the shaking that will actually occur. Such a development might seem discouraging, but may prove to be the case. Objective analysis and testing of hazard maps is the only way to find out.

## Acknowledgments

We thank Bruce Spenser and Jerome Stein for valuable discussions, and Antonella Peresan and an anonymous reviewer for helpful comments.

## References

- Albarelo, D., D'Amico, V., 2008. Testing probabilistic seismic hazard estimates by comparison with observations: an example in Italy. *Geophysical Journal International* 175, 1088–1094.
- Algermissen, S.T., Perkins, D.M., Thenhaus, P.C., Hanson, S.L., Bender, B.L., 1982. Probabilistic estimates of maximum acceleration and velocity in rock in the contiguous United States. U.S. Geol. Surv. Open-File Report, pp. 82–1033.
- Ando, M., 1975. Source mechanisms and tectonic significance of historical earthquakes along the Nankai Trough, Japan. *Tectonophysics* 27, 119–140.
- Atkinson, G.M., Boore, D.M., 1995. Ground-motion relations for Eastern North America. *Bulletin of the Seismological Society of America* 85, 17–30.
- Bakun, W.H., Aagaard, B., Dost, B., Ellsworth, W.L., Hardebeck, J.L., Harris, R.A., Ji, C., Johnston, M.J.S., Langbein, J., Lienkaemper, J.J., Michael, A.J., Murray, J.R., Nadeau, R.M., Reasenber, P.A., Reichle, M.S., Roeloffs, E.A., Shakal, A., Simpson, R.W., Waldhauser, F., 2005. Implications for prediction and hazard assessment from the 2004 Parkfield earthquake. *Nature* 437, 969–974.
- Bilham, R., Engdahl, R., Feldl, N., Satyabala, S.P., 2005. Partial and complete rupture of the Indo-Andaman plate boundary 1847–2004. *Seismological Research Letters* 76, 299–311.
- Bommer, J.J., 2009. Deterministic vs. probabilistic seismic hazard assessment: an exaggerated and obstructive dichotomy. *Journal of Earthquake Engineering* 6, 43–73.
- Burger, R.W., Somerville, P.G., Barker, J.S., Herrmann, R.B., Helmberger, D.V., 1987. The effect of crustal structure on strong ground motion attenuation relations in eastern North America. *Bulletin of the Seismological Society of America* 77, 420–439.
- Camelbeek, T., Vanneste, K., Alexandre, P., Verbeeck, K., Petermans, T., Rosset, P., Everaerts, M., Wannant, R., Van Camp, M., 2007. Relevance of active faulting and seismicity studies to assess long term earthquake activity in Northwest Europe. In: Stein, S., Mazzotti, S. (Eds.), *Continental Intraplate Earthquakes: Science, Hazard, and Policy Issues*, Special Paper 425. GSA, Boulder, CO, pp. 193–224.
- Castanos, H., Lomnitz, C., 2002. PSHA: is it science? *Engineering Geology* 66 (315–317) 2002.
- Chang, K., Blindsided by ferocity unleashed by a fault, *New York Times*, March 21, 2011.
- Chang, W.-L., Smith, R.B., 2002. Integrated seismic-hazard analysis of the Wasatch front, Utah. *Bulletin of the Seismological Society of America* 92, 1904–1922.
- Chen, Q.-F., Wang, K., 2010. The 2008 Wenchuan earthquake and earthquake prediction in China. *Bulletin of the Seismological Society of America* 100, 2840–2857.
- Clark, D., McPherson, A., Collins, C., 2011. Australia's seismogenic neotectonic record: a case for heterogeneous intraplate deformation. *Geoscience Australia Record* 72–88 (2011/11).
- Cornell, C.A., 1968. Engineering seismic risk analysis. *Bulletin of the Seismological Society of America* 58, 1583–1606.
- Crone, A.J., De Martini, P.M., Machette, M.N., Okumura, K., Prescott, J.R., 2003. Paleoseismicity of two historically quiescent faults in Australia: implications for fault behavior in stable continental regions. *Bulletin of the Seismological Society of America* 93, 1913–1934.
- CSEP (Collaboratory for the Study of Earthquake Predictability), 2011. <http://www.cseptesting.org/home>.
- Cyranoski, D., 2012. Tsunami simulations scare Japan. *Nature* 484, 296–297.
- Derman, E., 2004. *My Life As A Quant: Reflections on Physics and Finance*. Wiley, New York.
- Earthquake Research Committee, 2009. Long-term forecast of earthquakes from Sanriku-oki to Boso-oki.
- Earthquake Research Committee, 2010. National seismic hazard maps for Japan.
- Emanuel, K., 2011. Why are hurricane forecasts still so rough? [http://articles.cnn.com/2011-08-25/opinion/emanuel.weather.predict\\_1\\_forecast-hurricane-irene-national-hurricane-center?\\_s=PM:OPINION](http://articles.cnn.com/2011-08-25/opinion/emanuel.weather.predict_1_forecast-hurricane-irene-national-hurricane-center?_s=PM:OPINION).
- Fackler, M., Tsunami warnings written in stone, *New York Times*, April 20, 2011.
- Feynman, R.P., 1986. *QED: The Strange Theory of Light and Matter*. Princeton University Press, Princeton, NJ.
- Feynman, R.P., 1988. *What Do You Care What other People Think*. W. W. Norton, New York, NY.
- Field, E., Dawson, T., Felzer, K., Frankel, A., Gupta, V., Jordan, T., Parsons, T., Petersen, M., Stein, R., Weldon, R., Wills, C., 2008. The Uniform California Earthquake Rupture Forecast, Version 2. U.S. Geol. Surv. Open File Report 2007-1437.
- Frankel, A., Mueller, C., Barnhard, T., Perkins, D., Leyendecker, E., Dickman, N., Hanson, S., Hopper, M., 1996. National Seismic Hazard Maps Documentation. U.S. Geol. Surv. Open-File Report 96-532. U.S. Government Printing Office, Washington, D.C.
- Frankel, A., Harmsen, S., Mueller, C., Calais, E., Haase, J., 2010. Documentation for initial seismic hazard maps for Haiti. Open-File Report 2010-1067. U.S. Government Printing Office, Washington, D.C.
- Freedman, D., Stark, P., 2003. What is the chance of an earthquake? Mulargia, F., Geller, R.J. (Eds.), *Earthquake Science and Seismic Risk Reduction*, NATO Science Series IV: Earth and Environmental Sciences, 32. Kluwer, Dordrecht, The Netherlands, pp. 201–213.
- Friedrich, A.M., Wernicke, B.P., Niemi, N.A., Bennett, R.A., Davis, J.L., 2003. Comparison of geodetic and geologic data from the Wasatch region, Utah, and implications for the spectral character of Earth deformation at periods of 10 to 10 million years. *Journal of Geophysical Research* 108, 2199. <http://dx.doi.org/10.1029/2001JB000682>.
- Gardner, D., 2010. Future Babble: Why Expert Predictions Fail – and Why We Believe Them Anyway. McClelland & Stewart, Toronto.
- Geller, R.J., 1997. Earthquake prediction: a critical review. *Geophysical Journal International* 131, 425–450.
- Geller, R.J., 2011. Shake-up time for Japanese seismology. *Nature* 472, 407–409.
- Geschwind, C.-H., 2001. *California Earthquakes: Science, Risk, and the Politics of Hazard Mitigation*. Johns Hopkins University Press, Baltimore.
- GSHAP (Global Seismic Hazard Assessment Program), 1999. <http://www.seismo.ethz.ch/static/GSHAP>.
- Gutenberg, B., Richter, C.F., 1944. Frequency of earthquakes in California. *Bulletin of the Seismological Society of America* 34, 185–188.
- Hager, B., Cornell, C.A., Medigovich, W.M., Mogi, K., Smith, R.M., Tobin, L.T., Stock, J., Weldon, R., 1994. Earthquake Research at Parkfield, California, for 1993 and Beyond. Report of the NEPEC [National Earthquake Prediction Evaluation Council] Working Group. U.S. Geological Survey Circular, 1116.
- Halchuk, S., Adams, J., 1999. Crossing the border: assessing the difference between new Canadian and American seismic hazard maps. *Proc. 8th Canad. Conf. Earthquake Engineering*, pp. 77–82.
- Hammit, J.K., Shlyakhter, A.I., 1999. The expected value of information and the probability of surprise. *Risk Analysis* 19, 135–152.
- Hanks, T.C., Cornell, C.A., 1994. Probabilistic seismic hazard analysis: a beginner's guide. Fifth Symposium on Current Issues Related to Nuclear Power Plant Structures, Equipment, and Piping.
- Hebden, J., Stein, S., 2009. Time-dependent seismic hazard maps for the New Madrid seismic zone and Charleston, South Carolina areas. *Seismological Research Letters* 80, 10–20.
- Henriion, M., Fischhoff, B., 1986. Assessing uncertainty in physical constants. *American Journal of Physics* 54, 791–798.
- Hough, S., 2009. *Predicting the Unpredictable*. Princeton University Press, Princeton, NJ.
- Hough, S., Page, M., 2011. Toward a consistent model for strain accrual and release for the New Madrid Seismic Zone, central United States. *Journal of Geophysical Research* 116. <http://dx.doi.org/10.1029/2010JB007783>.
- Hough, S., Armbruster, J.G., Seeber, L., Hough, J.F., 2000. On the Modified Mercalli Intensities and magnitudes of the 1811/1812 New Madrid, central United States, earthquakes. *Journal of Geophysical Research* 105, 23,839–23,864.
- Intergovernmental Panel on Climate Change, 2007. *Climate Change 2007: Working Group I: The Physical Science Basis*.
- Ishibashi, K., 1981. Specification of a soon-to-occur seismic faulting in the Tokai district, central Japan, based upon seismotectonics. In: Simpson, D., Richards, P. (Eds.), *Earthquake Prediction*, Maurice Ewing Ser. 4. AGU, Washington, D.C., pp. 297–332.
- Ishibashi, K., Satake, K., 1998. Problems on forecasting great earthquakes in the subduction zones around Japan by means of paleoseismology (in Japanese). *Zisin (J. Seismol. Soc. Japan)* 50 (special issue), 1–21.
- Jackson, D.D., Kagan, Y.Y., 2006. The 2004 Parkfield earthquake, the 1985 prediction, and characteristic earthquakes: lessons for the future. *Bulletin of the Seismological Society of America* 96, S397–S409.

- Johnston, A.C., 1996. Seismic moment assessment of earthquakes in stable continental regions – II. New Madrid 1811–1812, Charleston 1886 and Lisbon 1755. *Geophysical Journal International* 126, 314–344.
- Jordan, T., Chen, Y., Gasparini, P., Madariaga, R., Main, I., Marzocchi, W., Papadopoulos, G., Sobolev, G., Yamaoka, K., Zschau, J., 2011. Operational earthquake forecasting: state of knowledge and guidelines for utilization. *Annals of Geophysics* 54 (4). <http://dx.doi.org/10.4401/ag-5350>.
- Kafka, A., 2007. Does seismicity delineate zones where future large earthquakes are likely to occur in intraplate environments? In: Stein, S., Mazzotti, S. (Eds.), *Continental intraplate earthquakes: science, hazard, and policy issues*, Special Paper, 425. GSA, Boulder CO, pp. 35–48.
- Kagan, Y.Y., 1996. Comment on “The Gutenberg–Richter or characteristic earthquake distribution, which is it?” by S. Wesnousky. *Bulletin of the Seismological Society of America* 86, 274–285.
- Kagan, Y.Y., Jackson, D.D., 1991. Seismic gap hypothesis: ten years after. *Journal of Geophysical Research* 96, 21,419–21,431.
- Kagan, Y.Y., Jackson, D.D., 1995. New seismic gap hypothesis: five years after. *Journal of Geophysical Research* 99, 3943–3959.
- Kagan, Y.Y., Jackson, D.D., 2012. Whole Earth high-resolution earthquake forecasts. *Geophysical Journal International* 190, 677–686.
- Kagan, Y.Y., Jackson, D.D., in press. Tohoku earthquake: a surprise. *Bulletin of the Seismological Society of America*.
- Kanamori, H., 1977a. Energy release in great earthquakes. *Journal of Geophysical Research* 82, 2981–2987.
- Kanamori, H., 1977b. Seismic and aseismic slip along subduction zones and their tectonic implications. *Island Arcs, Deep-sea Trenches and Back-arc Basins*. In: Talwani, M., Pitman III, W.C. (Eds.), *Maurice Ewing Ser., 1*. AGU, Washington, D.C., pp. 163–174.
- Kanamori, H., 2011. Prepare for the unexpected. *Nature* 473, 147.
- Kelsey, H.M., Nelson, A.R., Hemphill-Haley, E., Witter, R.C., 2005. Tsunami history of an Oregon coastal lake reveals a 4600 year record of great earthquakes on the Cascadia subduction zone. *Geological Society of America Bulletin* 117, 1009–1032.
- Kerr, R.A., 1993. Parkfield quakes skip a beat. *Science* 259, 1120–1122.
- Kerr, R.A., 2011. Seismic crystal ball proving mostly cloudy around the world. *Science* 332, 912–913.
- Kossobokov, V.G., Nekrasova, A.K., 2012. Global Seismic Hazard Assessment Program maps are erroneous. *Seismic instruments* 48, 162–170.
- Lee, Y., Turcotte, D.L., Holliday, J.R., Sachs, M.K., Rundle, J.B., Chen, C., Tiampo, K.F., 2011. Results of the Regional Earthquake Likelihood Models (RELM) test of earthquake forecasts in California. *PNAS* 108 (40), 16533–16538. <http://dx.doi.org/10.1073/pnas.1113481108>.
- Leonard, M., Robinson, D., Allen, T., Schneider, J., Clark, D., Dhu, T., Burbidge, D., 2007. Toward a better model of earthquake hazard in Australia. In: Stein, S., Mazzotti, S. (Eds.), *Continental Intraplate Earthquakes*, Special Paper, 425. GSA, Boulder, CO, pp. 263–283.
- Liu, M., Stein, S., Wang, H., 2011. 2000 years of migrating earthquakes in North China: How earthquakes in mid-continents differ from those at plate boundaries. *Lithosphere* 3. <http://dx.doi.org/10.1130/L129>.
- Lomnitz, C., 1989. Comment on “temporal and magnitude dependence in earthquake recurrence models” by C. A. Cornell and S. R. Winterstein. *Bulletin of the Seismological Society of America* 79, 1662.
- Loveless, J.P., Meade, B.J., 2010. Geodetic imaging of plate motions, slip rates, and partitioning of deformation in Japan. *Journal of Geophysical Research* 115. <http://dx.doi.org/10.1029/2008JB006248>.
- Lowenstein, R., 2000. *When Genius Failed: The Rise and Fall of Long-Term Capital Management*. Random House, New York.
- Luo, G., Liu, M., 2010. Stress evolution and fault interactions before and after the 2008 Great Wenchuan earthquake. *Tectonophysics* 491 (1–4), 127–140. <http://dx.doi.org/10.1016/j.tecto.2009.12.019>.
- Malservisi, R., Dixon, T.H., La Femina, P.C., Furlong, K.P., 2003. Holocene slip rate of the Wasatch fault zone, Utah, from geodetic data: earthquake cycle effects. *Geophysical Research Letters* 30, 1673. <http://dx.doi.org/10.1029/2003GL017408>.
- Manaker, D.M., Calais, E., Freed, A.M., Ali, S.T., Przybylski, P., Mattioli, G., Jansma, P., Prepetit, C., De Chabalie, J.B., 2008. Interseismic plate coupling and strain partitioning in the Northeastern Caribbean. *Geophysical Journal International* 174, 889–903.
- Marzocchi, W., Zechar, J.D., 2011. Earthquake forecasting and earthquake prediction: different approaches for obtaining the best model. *Seismological Research Letters* 82 (3), 442–448. <http://dx.doi.org/10.1785/gssrl.82.3.442>.
- McCaffrey, R., 2007. The next great earthquake. *Science* 315, 1675–1676.
- McCaffrey, R., 2008. Global frequency of magnitude 9 earthquakes. *Geology* 36, 263–266.
- McCloskey, J., Nalbant, S.S., Steacy, S., 2005. Indonesian earthquake: earthquake risk from co-seismic stress. *Nature* 434, 291.
- McGuire, R.K., 1995. Probabilistic seismic hazard analysis and design earthquakes: closing the loop. *Bulletin of the Seismological Society of America* 85, 1275–1284.
- Meghraoui, M., Delouis, B., Ferry, M., Giardini, D., Huggenberger, P., Spotteke, I., Granet, M., 2001. Active normal faulting in the upper Rhine Graben and the paleoseismic identification of the 1356 Basel earthquake. *Science* 293, 2070–2073.
- Meng, G., Ren, J., Wang, M., Gan, W., Wang, Q., Qiao, X., Yang, Y., 2008. Crustal deformation in western Sichuan region and implications for 12 May 2008 Ms 8.0 earthquake. *Geochemistry, Geophysics, Geosystems* 9. <http://dx.doi.org/10.1029/2008GC002144>.
- Merz, B., 2012. Role and responsibility of geoscientists for warning and mitigation of natural disasters. *European Geosciences Union General Assembly* 2012.
- Minoura, K., Imamura, F., Sugawa, D., Kono, Y., Iwashita, T., 2001. The 869 Jogan tsunami deposit and recurrence interval of large-scale tsunami on the Pacific coast of Northeast Japan. *Journal of Natural Disaster Science* 23, 83–88.
- Miyazawa, M., Mori, J., 2009. Test of seismic hazard map from 500 years of recorded intensity data in Japan. *Bulletin of the Seismological Society of America* 99, 3140–3149.
- Monecke, K., Finger, W., Klarer, D., Kongko, W., McAdoo, B., Moore, A., Sudrajat, S., 2008. A 1,000-year sediment record of tsunami recurrence in northern Sumatra. *Nature* 455, 1232–1234.
- Moseley, J.B., O'Malley, K., Petersen, N.J., Menke, T.J., Brody, B.A., Kuykendall, D.H., Hollingsworth, J.C., Ashton, C.M., Wray, N.P., 2002. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *The New England Journal of Medicine* 347, 81–88.
- Mucciarelli, M., Albarello, D., D'Amico, V., 2008. Comparison of probabilistic seismic hazard estimates in Italy. *Bulletin of the Seismological Society of America* 98, 2652–2664.
- Murphy, A.H., 1988. Skill scores based on the mean square error and their relation to the correlation coefficient. *Monthly Weather Review* 116, 2417–2424.
- Murphy, A.H., 1993. What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather and Forecasting* 8, 281–293.
- Nanayama, F., Satake, K., Furukawa, R., Shimokawa, K., Atwater, B., Shigeno, K., Yamaki, S., 2003. Unusually large earthquakes inferred from tsunami deposits along the Kuril trench. *Nature* 424, 660–663.
- Newman, A.V., 2011. Hidden Depths. *Nature* 474, 441–443.
- Newman, A., Stein, S., Weber, J., Engeln, J., Mao, A., Dixon, T., 1999. Slow deformation and lower seismic hazard at the New Madrid Seismic Zone. *Science* 284, 619–621.
- Newman, A., Stein, S., Schneider, J., Mendez, A., 2001. Uncertainties in seismic hazard maps for the New Madrid Seismic Zone. *Seismological Research Letters* 72, 653–667.
- Nishenko, S.P., Buland, R., 1987. A generic recurrence interval distribution for earthquake forecasting. *Bulletin of the Seismological Society of America* 77, 1382–1399.
- Nöggerath, J., Geller, R.J., Gusiakov, V.K., 2011. Fukushima: the myth of safety, the reality of geoscience. *Bulletin of the Atomic Scientists* 68, 37–46.
- Normile, D., 2012. One year after the devastation, Tohoku designs its renewal. *Science* 335, 1164–1166.
- Nuttli, O.W., 1973. The Mississippi Valley earthquakes of 1811 and 1812: intensities, ground motion, and magnitudes. *Bulletin of the Seismological Society of America* 63, 227–248.
- Okal, E.A., Synolakis, C., 2004. Source discriminants for near-field tsunamis. *Geophysical Journal International* 158, 899–912.
- Onishi, N., Seawalls offered little protection against tsunami's crushing waves, *New York Times*, March 13, 2011.
- Onishi, N., In Japan, seawall offered a false sense of security, *New York Times*, March 31, 2011.
- Onishi, N., Japan revives a sea barrier that failed to hold, *New York Times*, November 2, 2011.
- Oreskes, N., 2000. Why Predict? Historical Perspectives on Prediction in Earth Science. In: Sarewitz, D., Pielke Jr., R., Byerly Jr., R. (Eds.), *Prediction: science, decision making, and the future of Nature*. Island Press, Washington D.C.
- Overbye, D., They tried to outsmart Wall Street, *New York Times*, March 10, 2009.
- Panza, G.F., Irikura, K., Kouteva, M., Peresan, A., Wang, Z., Saragoni, R., 2010. Advanced seismic hazard assessment. *Pageoph* 168, 1–9.
- Parsons, T., 2008a. Monte Carlo method for determining earthquake recurrence parameters from short paleoseismic catalogs: example calculations for California. *Journal of Geophysical Research* 113. <http://dx.doi.org/10.1029/2007JB004998>.
- Parsons, T., 2008b. Earthquake recurrence on the south Hayward fault is most consistent with a time dependent, renewal process. *Geophysical Research Letters* 35. <http://dx.doi.org/10.1029/2008GL035887>.
- Pechmann, J.C., Arabasz, W., 1995. The problem of the random earthquake in seismic hazard analysis: Wasatch front region, Utah. *Environmental and Engineering Geology of the Wasatch Front Region*. In: Lunt, W.R. (Ed.), 1995 Symposium and Field Conference. Utah Geol. Assoc., pp. 77–93.
- Peterson, M.D., Cao, T., Campbell, K.W., Frankel, A.D., 2007. Time-independent and time-dependent seismic hazard assessment for the state of California: uniform California Earthquake rupture forecast model 1.0. *Seismological Research Letters* 78, 99–109.
- Pilkey, O.H., Pilkey-Jarvis, L., 2006. *Useless Arithmetic: Why Environmental Scientists Can't Predict the Future*. Columbia University Press, New York.
- Reyners, M., 2011. Lessons from the destructive Mw 6.3 Christchurch, New Zealand, earthquake. *Seismological Research Letters* 82, 371–372.
- Ruff, L., Kanamori, H., 1980. Seismicity and the subduction process. *Physics of the Earth and Planetary Interiors* 23, 240–252.
- Sagiya, T., 2011. Integrate all available data. *Nature* 473, 146–147.
- Sarewitz, D., Pielke Jr., R., Byerly Jr., R., 2000. *Prediction: science, decision making, and the future of Nature*. Island Press, Washington D.C.
- Satake, K., Atwater, B.F., 2007. Long-term perspectives on giant earthquakes and tsunamis at subduction zones. *Annual Review of Earth and Planetary Sciences* 35, 349–374.
- Savage, J.C., 1991. Criticism of some forecasts of the national earthquake prediction council. *Bulletin of the Seismological Society of America* 81, 862–881.
- Savage, J.C., 1992. The uncertainty in earthquake conditional probabilities. *Geophysical Research Letters* 19, 709–712.
- Savage, J.C., 1993. The Parkfield prediction fallacy. *Bulletin of the Seismological Society of America* 83, 1–6.
- Savage, J.C., 1994. Empirical earthquake probabilities from observed recurrence intervals. *Bulletin of the Seismological Society of America* 84, 219–221.
- Scholz, C.H., 2002. *The mechanics of earthquakes and faulting*. Cambridge University Press, Cambridge.



- Schorlemmer, D., Zecher, J.D., Werner, M., Jackson, D.D., Field, E.H., Jordan, T.H., the RELM Working Group, 2010. First results of the Regional Earthquake Likelihood Models Experiment. *Pure and Applied Geophysics (The Frank Evison Volume)* 167 (8/9), 859–876. <http://dx.doi.org/10.1007/s00024-010-0081-5>.
- Schwartz, D.P., Coppersmith, K.J., 1984. Fault behavior and characteristic earthquakes: examples from the Wasatch and San Andreas fault zones. *Journal of Geophysical Research* 89, 5681–5698.
- Searer, G., Freeman, S.A., Paret, T.F., 2007. Does it make sense from engineering and scientific perspectives to design for a 2475-year earthquake? In: Stein, S., Mazzotti, S. (Eds.), *Continental Intraplate Earthquakes*, Special Paper, 425. GSA, Boulder, CO, pp. 353–361.
- Sieh, K., Stuiver, M., Brillinger, D., 1989. A more precise chronology of earthquakes produced by the San Andreas fault in southern California. *Journal of Geophysical Research* 94, 603–624.
- Simons, M., Minson, S.E., Sladen, A., Ortega, F., Jiang, J., Owen, S.E., Meng, L., Ampuero, J.-P., Wei, S., Chu, R., Helmlinger, D.V., Kanamori, H., Hetland, E., Moore, A.W., Webb, F.H., 2011. The 2011 magnitude 9.0 Tohoku-Oki earthquake: mosaicking the megathrust from seconds to centuries. *Science* 332, 1421–1425.
- Stein, R., 1999. The role of stress transfer in earthquake occurrence. *Nature* 402, 605–609.
- Stein, S., 2010. *Disaster Deferred: How New Science Is Changing Our View Of Earthquake Hazards in the Midwest*. Columbia University Press, New York.
- Stein, J.L., 2012. *Stochastic Optimal Control and the U.S. Financial Debt Crisis*. Springer.
- Stein, S., Liu, M., 2009. Long aftershock sequences within continents and implications for earthquake hazard assessment. *Nature* 462, 87–89.
- Stein, S., Newman, A., 2004. Characteristic and uncharacteristic earthquakes as possible artifacts: applications to the New Madrid and Wabash seismic zones. *Seismological Research Letters* 75, 170–184.
- Stein, S., Okal, E.A., 2005. Speed and size of the Sumatra earthquake. *Nature* 434, 581–582.
- Stein, S., Okal, E.A., 2007. Ultralong period seismic study of the December 2004 Indian Ocean earthquake and implications for regional tectonics and the subduction process. *Bulletin of the Seismological Society of America* 87, S279–S295.
- Stein, S., Okal, E.A., 2011. The size of the 2011 Tohoku earthquake needn't have been a surprise. *Eos, Transactions, American Geophysical Union* 92, 227–228.
- Stein, J.L., Stein, S., 2012. Rebuilding Tohoku: a joint geophysical and economic framework for hazard mitigation. *GSA Today* 22, 28–30.
- Stein, S., Wyssession, M., 2003. *Introduction to Seismology, Earthquakes, and Earth Structure*. Blackwell, Oxford.
- Stein, R., Yates, R.S., 1989. Hidden earthquakes. *Scientific American* 260, 48–57 (January).
- Stein, S., Sleep, N.H., Geller, R.J., Wang, S.C., Kroeger, G.C., 1979. Earthquakes along the passive margin of eastern Canada. *Geophysical Research Letters* 6, 537–540.
- Stein, S., Engeln, J.F., DeMets, C., Gordon, R.G., Woods, D., Lundgren, P., Argus, D., Stein, C., Wiens, D.A., 1986. The Nazca-South America convergence rate and the recurrence of the great 1960 Chilean earthquake. *Geophysical Research Letters* 13, 713–716.
- Stein, S., Cloetingh, S., Sleep, N., Wortel, R., 1989. Passive margin earthquakes, stresses, and rheology. In: Gregerson, S., Basham, P. (Eds.), *Earthquakes at North-Atlantic Passive Margins: Neotectonics and Postglacial Rebound*. Kluwer, Dordrecht, pp. 231–260.
- Stein, S., Newman, A., Tomasello, J., 2003. Should Memphis build for California's earthquakes? *Eos, Transactions, American Geophysical Union* 84 (177), 184–185.
- Stein, S., Friedrich, A., Newman, A., 2005. Dependence of possible characteristic earthquakes on spatial sampling: illustration for the Wasatch seismic zone, Utah. *Seismological Research Letters* 76, 432–436.
- Stein, S., Liu, M., Calais, E., Li, Q., 2009. Midcontinent earthquakes as a complex system. *Seismological Research Letters* 80, 551–553.
- Stein, S., Geller, R.J., Liu, M., 2011. Bad assumptions or bad luck: why earthquake hazard maps need objective testing. *Seismological Research Letters* 82, 623–626.
- Stephenson, D., 2000. Use of the "Odds Ratio" for diagnosing forecast skill. *Weather and Forecasting* 15, 221–232.
- Stirling, M.W., Gerstenberger, M., 2010. Ground motion-based testing of seismic hazard models in New Zealand. *Bulletin of the Seismological Society of America* 100, 1407–1414.
- Swafford, L., Stein, S., 2007. Limitations of the short earthquake record for seismicity and seismic hazard studies. In: Stein, S., Mazzotti, S. (Eds.), *Continental Intraplate Earthquakes*, Special Paper, 425. GSA, Boulder, CO, pp. 49–58.
- Taleb, N.N., 2007. *The Black Swan: The Impact of the Highly Improbable*. Random House, New York.
- Toro, G.R., Abrahamson, N.A., Schneider, J.F., 1997. Model of strong ground motions from earthquakes in central and eastern North America: best estimates and uncertainties. *Seismological Research Letters* 68, 41–57.
- Toth, L., Gyori, E., Monus, P., Zsiros, T., 2004. Seismicity and seismic hazard in the Panonian basin. *Proc. NATO Ad. Res. Worksh.: The Adria microplate*, pp. 119–123.
- Valensise, G., Pantosti, D., 2001. Investigation of potential earthquake sources in peninsular Italy: a review. *Journal of Seismology* 5, 287–306.
- Wang, Z., 2011. Seismic hazard assessment: issues and alternatives. *Pure and Applied Geophysics* 168, 11–25.
- Wang, Z., Cobb, J., in press. A critique of probabilistic versus deterministic seismic hazard analysis with special reference to the New Madrid seismic zone, in: *Recent Advances in North American Paleoseismology and Neotectonics east of the Rockies*. GSA, Boulder, CO.
- Wang, H., Liu, M., Cao, J., Shen, X., Zhang, G., 2011. Slip rates and seismic moment deficits on major active faults in mainland China. *Journal of Geophysical Research* 116, B02405. <http://dx.doi.org/10.1029/2010JB007821>.
- Witze, A., 2009. The sleeping dragon. *Nature* 459, 153–157.
- Wyss, M., Nekraskova, A., Kossobokov, V., 2012. Errors in expected human losses due to incorrect seismic hazard estimates. *Natural Hazards* 62, 927–935.
- Yomogida, K., Yoshizawa, K., Koyama, J., Tsuzuki, M., 2011. Along-dip segmentation of the 2011 off the Pacific coast of Tohoku Earthquake and comparison with other megathrust earthquakes. *Earth, Planets and Space* 63, 697–701.
- Youngs, R.R., Coppersmith, K.J., 1985. Implications of fault slip rates and earthquake recurrence models to probabilistic seismic hazard estimates. *Bulletin of the Seismological Society of America* 75, 939–964.
- Zhang, P.-Z., Shen, Z., Wang, M., Gan, W., Burgmann, R., Molnar, P., Wang, Q., Niu, Z., Sun, J., Wu, J., Hanrong, S., Xinzhaoy, Y., 2004. Continuous deformation of the Tibetan Plateau from global positioning system data. *Geology* 32, 809–812.