PROBABILISTIC SEISMIC HAZARD ANALYSIS: A BEGINNER'S GUIDE

Thomas C. Hanks* and C. Allin Cornell**

ABSTRACT

This paper is the third and final iteration of a several year effort by the authors to reveal the essential simplicity of probabilistic seismic hazard analysis (PSHA) in more-or-less plain English, a simplicity customarily veiled by user-hostile notation, antonymous jargon, and proprietary software. As PSHA proponents press their case for primacy of PSHA over DSHA (deterministic seismic hazard analysis) in developing seismic design criteria for a broad band of structural types and functions, including nuclear facilities, some resistance has arisen from the defenders of "engineering determinism." In large part, this resistance is born of the inability of PHSA proponents to communicate clearly and directly to anyone but themselves just what it is that they are doing. Important recent developments in PSHA (or any other form of natural hazard analysis) include the consistent treatment of uncertainties and the use of expert judgment. The quantitative expression of both data/knowledge uncertainties and diverse expert opinion is likely to evolve for some time to come. Nevertheless, PSHA and DSHA have far more in common than they do in differences, and only one fundamental difference separates the two approaches: PSHA carries units of time and DSHA doesn't. Even so, it is generally possible to associate recurrence interval information with plausible deterministic earthquakes, and when this is the case they can always be found in hazard space.

INTRODUCTION

Every once in a while, something bad happens as a result of an earthquake, and probabilistic seismic hazard analysis (PSHA) is the basis on which one reckons how often bad happens at some place of interest. Similarly, bad things happen as a result of volcanoes, landslides, river floods, hurricanes, tornadoes, wildfires, and other natural events, and the hazards arising from them can be portrayed in the same format.

The essence of PSHA (or any other form of probabilistic hazards analysis) is that bad happens at a calculable rate. For scientists and engineers who deal with natural or manmade disasters on a regular basis, whether on the prevention or mitigation side of the

*U.S. Geological Survey, 345 Middlefield Road, Menlo Park, CA 94025
**Department of Civil Engineering, Stanford University, Stanford, CA 94305

fence, this proposition is little more than a truism. Ordinary folks, even well-educated ones, however, are only dimly aware that bad can and will happen to them, given enough time. The protective logic at work here is naive and wrong: Bad things don't happen to nice people. Bad things happen to someone else. Bad things don't happen in our engineering firm, they happen in yours (see, we got this logic, too). Bad things can't happen at NASA. And so on.

Nevertheless, there is some empirical basis to this "logic." Not counting various social realities such as poverty, child abuse, wife beating, and unenlightened superiors, bad things simply don't happen to most people. How many people do you know who got hurt in an earthquake, or who suffered losses in a hurricane, or who died in an airplane mishap, or who endured the 1993 floods in the Missouri/Mississippi River drainage? Already in our lifetimes, a couple of million Americans have died in automobile accidents. How many of these people do you know?

Our audience is large enough, we hope, that some of you will have had first-hand experience with one or more of the hazards above, and the two questions above are not intended to make light of any personal tragedy involved. The point is simply that the chances of any one American losing his life in an airplane or her house to a flood, or even of dying in an automobile are pretty small. Numbers like 10^{-6}/yr or 10^{-5}/yr or even 10^{-4}/yr are so small, in fact, that most people behave as if they were zero. While there is a certain practical reality in not wringing one's hands too hard over low-probability events, it is nevertheless a big mistake to equate "low" with zero in probability-land.

Would that life were this simple, even if incorrectly so. Do people regard the 10^{-6} to 10^{-7}/event chance of winning the California lottery as zero? No way, otherwise they never would have put their money down. Similarly, a radioactive release at Yucca Mountain, the proposed site of an underground repository for the nation's nuclear waste, may be down at the 10^{-7}/yr level, perhaps even smaller. Do most people consider this number zero? No way squared, in part because the health consequences of this hazard merit careful consideration.

In a technologically uninformed democracy such as ours, people are perfectly free to believe whatever they want to believe, whether or not their beliefs have a close association with objective reality. This presents an unusual challenge to scientists and engineers, because communicating hazard and risk to the American public will be a far more difficult task than calculating it. Communicating hazard and risk in plain language, however, requires an understanding of hazard and risk in plain language, and that is the point of this paper, for the special case of the hazard and risk attendant to earthquakes.

The heart of the PSHA format involves some very simple concepts, but as is often the case in science and engineering, the technical definition of a word or phrase may not make sense in terms of its household definition. So before we can really start, we need some vocabulary. First of all, things like earthquakes, volcanoes, and floods are not hazards, at least not in this business, so we will speak of them as **events**, natural events in these cases. Because of these events, things like strong ground motion, ashfalls, and river overbanks occur, but they aren't hazards either. **Hazard** is the **mean rate of exceedance** (MROE) of some chosen ground-motion amplitude, ashfall depth, or overbank height. MROE is almost always expressed numerically on a per-year basis. In the existing literature, one will mostly see hazard labeled **probability of exceedance** (hence PSHA) or less frequently, **frequency of exceedance**. But we will use MROE here because "frequency" and "probability" have other important usages in PSHA. Seismologists, for example, need frequencies (Hz) for ground-motion estimation purposes. And in the past decade, much effort has been directed to carrying uncertainty analysis along with the basic PSHA calculations, leaving novices in the field uncertain as to whether "probabilistic" refers to MROE, to the uncertainty in MROE, or to both. But the semantics are a side issue; the important matter here is that hazards are just small numbers with units of $(\text{year})^{-1}$.

Risk is the mean annual **loss** in dollars, property, or lives that results from the occurrence of the events, perhaps for a single structure at a single place or perhaps for a whole region that is seismically active. Risk has units of dollars/year or perhaps lives/year. While risk will always depend on hazard, risk is a very different thing from hazard, and each will be different functions of place. Seismic hazard can be high where the risk is low, on the Carrizo Plain adjacent to the San Andreas fault in California, and seismic risk can be high where the hazard is low, in Washington, D.C., for example.

Finally we should define **bad**, without getting too close to the precious gems of **hazard**, **risk**, and **loss**. So to keep things loose and sensible, we will use bad in this paper to mean one or more undesirable things, usually arising as conseqences of natural events. Similarly, we will continue to use PSHA as a convenient shorthand, but what we really mean is MROE seismic hazards analysis.

PSHA: THE BASICS AND SOME EXAMPLES

Figure 1 shows you everything you'll ever want to know about PSHA—which is precisely the problem. The only people who comprehend Figure 1 are the few hundred people in the world who know so much about PSHA that they don't need Figure 1 anyway. If Figure 1 makes you think you've been taken prisoner, intellectually speaking, that's O.K.; it's just the set-up for a plain-English treatment of PSHA.

Figure 2 is a back-of-the-envelope PSHA calculation that points to several issues implicit in the full-blown PSHA formalism. First, we need a seismic source zone (SSZ), fancy talk for a place where earthquakes occur. So Figure 2 begins with an SSZ quadrilateral of area $A = 5 \times 10^6$ km^2, which grossly approximates the United States east of the Rocky Mountains (EUS). Second, we need to know the seismicity rate for our EUS SSZ. Ordinarily, this is expressed as

$$\log N = a - b\mathbf{M} \ , \qquad (1)$$

the Gutenberg–Richter magnitude-distribution relation. N looks like it should denote the number of earthquakes, but it really stands for the number of earthquakes per unit time per unit area with moment magnitude (**M**) greater than or equal to **M**. This is far too busy for sensible people (and it gets a lot worse when one tries to account for truncation of equation (1) at some maximum magnitude, $\mathbf{M}_{max}$); to keep things simple, we will assume that earthquakes can occur anywhere within our EUS SSZ with equal likelihood and will do so at a mean rate of one $\mathbf{M} = 5$ event per year, one $\mathbf{M} = 6$ event per decade, and one $\mathbf{M} = 7$ event per century.

Finally we need what's known in the trade as a ground-motion attenuation relation. This technojive means that we need to determine horizontal distances from the earthquake source within which some ground motion measure—and we'll use peak ground acceleration (PGA) here—is achieved or exceeded for the earthquake sizes of interest. This is what is given in the matrix of numbers beneath "Horizontal Distances (km)..." in Figure 2. These numbers come from the WUS peak acceleration-magnitude relation of Joyner and Boore (1988), and they are **not** appropriate for hazard estimation in the EUS. The $R = 0$ entries at 0.4 g for $\mathbf{M} = 5$ and 6, however, underscore a point we shall return to later. It is worth emphasizing that within each distance R, the PGA will be **greater than or equal to** the given value of PGA, and that's why "exceedance" shows up in the prose.

We're just about home. We need to (1) convert each R to an area by forming πR^2; (2) divide each πR^2 by the total area $A = 5 \times 10^6$ km^2, this ratio being the likelihood that the place of interest will be in the action zone of some level of PGA given that an earthquake occurs somewhere in our EUS SSZ; and (3) multiply the elements of each column by the occurrence rate of each size earthquake. The results of this arithmetic are shown in the next matrix of numbers, the one entitled "MROE...", of course. Each entry is the number of times per year that one expects to be hit at any place in our EUS SSZ with PGA greater than or equal to 0.1 g, 0.2 g, and 0.4 g as a result of $\mathbf{M} = 5$, 6 or 7 earthquakes. The column labeled Σ is the addition of the results for $\mathbf{M} = 5$, 6, and 7. What we're really doing here, of course, is numerically integrating over our seismicity rate data (or the magnitude-distribution relation, equation (1)) with $\Delta\mathbf{M} = 1$.

We're talking pretty small numbers here (bottom of Figure 2), on the order of 10^{-4} to 10^{-6} per year. This is a source of endless confusion to PSHA beginners, including almost all Earth scientists who instinctively feel defining the 10^{-4}/yr to 10^{-6}/yr event requires having at least 10,000 to 1,000,000 years worth of data about such events. What makes these numbers small, however, is not the seismicity rate; it is the area ratio $\pi R^2/5 \times 10^6$ which is somewhere between 10^{-3} and 0 for the R's listed in Figure 2. The earthquakes, in fact, are occurring at the rate of 1/yr at $\mathbf{M} = 5$ to 10^{-2}/yr at $\mathbf{M} = 7$.

Figure 3 illustrates how important this matter of area is and also reminds us that some parts of the EUS are more seismically active than others, that is, our assumption in Figure 2 that our EUS SSZ is uniformly active with the given seismicity rate is pretty crude. How does this work? Figure 3 is a blow-up of our first pass at the EUS SSZ but includes two sub-regions that are more active than EUS SSZ as a whole. We have placed 0.5 of our seismicity rate data in Region 1 and 0.3 of it in Region 2 and assigned the remainder of our EUS SSZ the remaining 0.2 of our seismicity rate, so that as a whole our EUS SSZ has the same seismicity rate as in our back-of-the-envelope calculation (Figure 2). You also want to note that Regions 1 and 2 have areas 100 times smaller than the whole EUS SSZ. Thus, In Region 1 our seismicity rate is down by 0.5 but the $(\text{area})^{-1}$ term is up by a factor of 100 with respect to the numbers in Figure 2. Thus, for a site within Region 1 not close to the edges, all of the exceedance rates are up by a factor of 50 with respect to the Figure 2 calculations. What matters here is not so much the seismicity rate in Region 1 but its very small area. So you want to pay attention to how seismic source zones are drawn in the vicinity of your backyard.

Figure 4 shows a set of **hazard curves** for two different sites, one at the San Francisco abutment of the San Francisco–Oakland Bay Bridge (Fgure 4a) and the other for a site in Washington, D.C. (Figure 4b). They plot PGA on the abscissa and the MROE for it (the hazard) on the ordinate, referred to as "Annual Frequency" in Figure 4a and "Annual Exceedance Probability" in Figure 4b. In both cases, uncertainty bands are also associated with the mean hazard curves. At potentially damaging PGA's, say $\gtrsim 0.2$ g, the hazard is much lower in Washington, D.C., than at the San Francisco abutment, by factors of 100 to 1000. This is no surprise: earthquakes causing these levels of ground motion ($\mathbf{M} \gtrsim 5$) occur far more frequently in the Bay Area than they do in and around Washington. The uncertainty in these hazard estimates, however, is much greater for Washington than for San Francisco, a matter we shall return to in a later section, but this makes sense, too: much less is known about the causes and effects of seismogenesis in the vicinity of Washington, D.C., than for the active faults in the Bay Area. Finally, one should note the consequences of arithmetically averaging quantities that have highly skewed distribution functions. This is expressed in Figure 4a by a mean hazard curve that sits closer to the 95^{th} percentile curve than to the 5^{th} percentile curve and in Figure 4b by a mean hazard curve that sits above the median hazard curve.

For the fun of it, we plotted the hazard values we computed in Figure 2 as open circles in Figure 4b. Our estimates are a bit on the low side, but for folks horsing around on the back of an envelope, we didn't do bad at all, especially in view of our continental-scale seismicity rate approximation and our use of a WUS ground motion attenuation relation. So you don't have to be a statistiwizard to get in the right PSHA ballpark.

Nevertheless, there are other reasons, we think, why our back-of-the envelope calculations are low with respect to the seismic hazard curves in Figure 4b and why they decay more steeply with increasing peak acceleration. The essential matter here is that for a given magnitude and distance, you're not going to get the same value of PGA every time, you're going to get a range of them distributed about a median value. The data in Figure 2 tell us, for example, that for $\mathbf{M} = 6$ and $R = 12$ km, the anticipated or median value of PGA is 0.2 g. About 10 to 20% of the time, however, we expect to catch 0.4 or larger for $\mathbf{M} = 6$ and $R = 12$ km. This matters a lot when we want to estimate exceedance rates at 0.4 g. Figure 2 says that we only expect to get a contribution from $\mathbf{M} = 7$ events, but in fact we get important contributions from $\mathbf{M} = 6$ events. Figure 5 shows, in the case of the Northridge, California, earthquake (Jan. 17, 1994; $\mathbf{M} = 6.7$) that PGA's of $\geq 40\%$ g occurred routinely out to distances of 30 km, even though the median value at $R = 30$ km, as estimated by Boore *et al.* (1993) is only ~ 0.15 g. Figure 5 also shows us that the scatter of the results for any one earthquake is substantial and also that any one earthquake data set can stand well away from expected values.

To see how this dispersion of data about the median value affects hazard calculations, let's say that 15% of the time ground-motion amplitudes will exceed the median by a factor of 2 or more and that another 15% of the time they will be 1/2 or less of the median value. Returning to the MROE matrix at the bottom of Figure 2 we now want to take 15% of the numbers in the 0.1 g row and kick them up to the 0.2 g row. Similarly we want to take 15% of the numbers in the 0.2 g row and bounce them down to the 0.1 g row, and another 15% goes up to the 0.4 g row. Finally, we take 15% of the 0.4 g row and drop it down to the 0.2 g row. We could do this for every element of the MROE matrix, but for our purposes here, it suffices to work just with the Σ column, and the resulting modifications are shown in the Σ^σ column. The Σ^σ numbers are shown as solid circles in Figure 4b.

The hazard at 0.1 g has gone down a bit from 1.73 to 1.47×10^{-4}/yr, the artificial consequence in this exercise of gettng no "distribution contributions" from peak accelerations less than 0.1 g. The hazard at 0.2 g has increased by a factor of 2.3, and the hazard at 0.4 g has increased by more than a factor of 5. So these "distribution contributions," which is to say random variations in PGA data in this case, really matter, and they matter the most in the low-hazard, high PGA states. (It is worth noting the classic abuse of the English language that is going on here. Relatively speaking, ground motions that aren't really hazardous ($\sim$ 0.1 g) are said to have high hazard, while ground motions that can be very hazardous ($\sim$ 0.4 g) are said to have low hazard. You gotta keep your eye out for this kind of probabibabble. For us, of course, hazard is another way of saying MROE.)

A final matter worthy of note in Figure 2 is that the hazard computed for various ground-motion levels arises from a range of earthquake sizes, not just one. This will make for a problem when we try to associate a specific earthquake at a specific distance (a deterministic earthquake, for example) with a specific hazard level, a matter we will take up in a later section. But before we get too far adrift in the sea of PSHA details, we should return to more basic questions, things like: Why do people use PSHA, and how do they use it, and who are they?

FAILURE PROBABILITY AS AN EXAMPLE OF RISK

PSHA or any other imaginable form of seismic hazard analysis, is performed by one set of people because another set of people is worried about what earthquakes and attendant phenomena might do to something they want to build or have already built. Geologists, seismologists, and other Earth scientists figure prominently in the first set of people, but not exclusively so. Risk analysis and in particular the special case of PSHA are developments of the engineering sciences, not the Earth sciences. Most people imagine that the second set of people is entirely populated by engineers, but they just work for the people who really count, those very few—at the highest level of action—who have put or will put a billion dollars or so on the table to get something built, a nuclear reactor, say, or a Trans-Alaskan pipeline. We don't know people with money like that to spend, but it seems safe to say that one of the reasons these people have so much money is because they are pretty careful with what they do with it. At the same time, there is no way one can pile up that kind of money without taking some risks. So these people must know about risk, and they must be prepared to take risks. They just like to keep their risk really small, because when these people lose, they lose big. When it comes to building stuff, they hire engineers to keep the risk small for them.

More generally, of course, all of us face, both individually and collectively, a panoply of hazards and risks from both natural and human causes. In a sensible and very readable book, Lewis (1990) discusses a number of these, from air travel to toxic chemicals, helps you wade through them, and reminds us again that societal perceptions of hazards and risks may or may not be the same as the results of quantitative analysis.

But let's get back to earthquakes and buildings falling down. Everything starts with P_f, the failure probability per year or mean rate of failure of some structure of interest. (P_f is also known as the performance goal (DOE–STD–1020, 1994), although why anyone would want to identify a failure probability as a **performance goal** escapes us.) Anyway, P_f is set by economic considerations or life-safety concerns or perhaps political fiat. P_f could be anything, but let's agree it is 10^{-4}/year. What this means is that a one-in-ten-thousand chance per year of failure of the structure is tolerable to the people who fixed this number. This explicit tolerance of failure is why PSHA is now ascendant over what is known in the seismic-hazard business as deterministic seismic hazard analysis (DSHA), a subject we will take up in the next section; PSHA admits that bad can happen, and so do the bankers, but DSHA doesn't (see, for example, Hamburger (1996)).

How does P_f condition the design and construction of the structure? We need two things for this work, the first being a hazard curve $H(a)$ appropriate to the construction site, like those in Figure 4. The second thing we need is something called the fragility function, $F(a)$. $F(a)$ tells you about the probability of structural failure, given some acceleration a. The top of Figure 6 shows such a fragility function, together with a hazard curve at the bottom. At low levels of a, $F(a) = 0$, and the structure just goes along with the ground-motion ride. At larger a, $F(a)$ climbs from zero, which is to say that the probability of structural failure is increasing. At still larger a, $F(a)$ goes to 1 at which point you are pretty sure your structure is on the ground—or maybe at the angle of repose. Because $H(a)$ is the probability (per year) of getting a or greater, we multiply $F'(a)$, the derivative of $F(a)$, times $H(a)$ and integrate over all a to get P_f:

$$P_f = \int_0^\infty H(a) \cdot F'(a) da \tag{2}$$

$F(a)$ (and $F'(a)$, of course) depends on the type and amount of the materials you build with and how well you put the structure together. The better, or stronger, you build the structure, the more the $F'(a)$ bubble in the center of Figure 6 moves to the right, because such a structure can withstand stronger loads or forces, which arise in the structure because of the ground accelerations. The width of the $F'(a)$ bubble, however, depends on life's little disconcertainties (new word), that nominally the same concrete pours or steel reinforcing rods do not break/yield at exactly the same tensile force, and that the actual response of a column-beam connection will be different for different ground-motion records even if they all have the same PGA, and stuff like that.

So what are we going to do with Equation (1)? People don't know so much about $F'(a)$ as they would like, so it is idealized as a logarithmically normal distribution function, of course. $H(a)$ is generally pretty close to a power-law function, so the integral in equation (1) is hard to do for people who have better things to do than learn this much calculus. Besides, nice people aren't going to wallow in that mathematical slop when they can guess the right answer, and we can guess the right answer pretty accurately if that $F'(a)$ bubble is narrow enough and has unit area under the bubble, which it does.

The idea is that, if the $F'(a)$ bubble is narrow enough, the only contribution to the integral in equation (1) comes from the immediate vicinity of $H(a_{50})$, the hazard level corresponding to the center, a_{50}, of the bubble, so

$$P_f \simeq H(a_{50})\,. \tag{3}$$

While our evaluation of the integral in equation (2) is surely an approximation for any real structure, the important point is that performance goals as specified by P_f lead naturally to PSHA as specified by $H(a)$. There may be—and generally are—intermediate design considerations other than the shape of $F'(a)$, our favorite being R_r (DOE–STD–1020, 1994), the risk-reduction rascal which does nothing at all to reduce the risk as expressed by P_f. But the essence of this design process remains the same, that is, dealing with some hazard level and the ground motions that contribute to it. It would seem that we would be home free with such a performance-based approach to design, but we're not; it turns out that we are a prisoner of our own history.

DSHA AND THE KRINITZSKY FACTOR

The deterministic approach to seismic hazards analysis, DSHA for brevity, seems to be a very different animal from PSHA. Even though PSHA isn't exactly a puppy, its basics having been laid out almost 30 years ago (Cornell, 1968), most cogniscenti regard DSHA as the old-fashioned way of trying to make sense of one's exposure to real or imagined seismic hazards. It deals with fascinating things like "maximum credible earthquake" or MCE, "'safe-shutdown earthquake" or SSE, and "operating-basis earthquake" or OBE, terminology plied in the large-dams and nuclear-reactor trades. In the heyday of DSHA, back in the 1960's and 1970's when there were still acknowledged wise men in the Earth and Engineering Sciences related to earthquakes, one or maybe a few of these people would decide the MCE/SSE/OBE's and where they would be likely to occur. A little cook-book ground-motion estimation would then ensue and—bingo!—seismic design criteria. This doesn't sound like much for serious things like nuclear reactors, and people have been on the lookout for something better ever since. This, of course, is PSHA, essentially the only other game in town.

(SSE and OBE, just like DBE (Design Basis Earthquake) aren't really earthquakes, but rather some specified levels of ground motion or their response spectrum facsimiles. MCE used to be an earthquake until Clarence Allen pointed out that "maximum credible" is semantically synonymous with "minimum incredible." No one has wanted to touch this one since. Nevertheless, the acronym is presently being resurrected as the "maximum **considered** earthquake" (NEHRP, 1997)).

There has been considerable heat and smoke—but very little light—shed on the real and imagined differences between PSHA and DSHA. Because DSHA and PSHA both purport to perform seismic hazard analysis, they must have a lot of things in common, and they do. In both methodologies, one needs to know where earthquakes do and might occur and how to estimate ground motion across the frequency band of interest for one or more source-site combinations. Also common to both DSHA and PSHA are the same incomplete and inadequate seismicity and ground-motion data bases, with which practitioners of either art must work.

Nevertheless, there are some substantive differences between the two approaches. DSHA generally makes do with one or just a few earthquakes at specified locations, whereas PSHA integrates across a wide range of possible earthquake magnitudes and source-site distances. When it comes to presenting and justifying design basis ground motions (DBGM), DSHA has an important advantage in transparency in getting from deterministic earthquake(s) to DBGM. In PSHA land, this is a more sophisticated exercise, as we indicated earlier and will discuss in more detail in a following section, because a range of magnitudes and distances contribute to the calculated hazard. One benefit of the extra cost here, however, is having DBGM that can be tied to structural performance criteria, P_f.

A second important—but not fundamental—difference between DSHA and PSHA is the latter's use of uncertainty analysis and expert opinion, at least in the past ten years or so. DSHA commonly proceeds without uncertainty analysis and the systematic input of diverse expert opinion, but there is no reason why DSHA could not incorporate them.

The fundamental difference, both philosophically and practically we belive, is that PSHA carries units of time and DSHA does not. The important aspect of DSHA, then, is not that it is "deterministic" (whatever this means in view of all the guesswork) but that it is a time-independent statement of what you're dealing with. The idea is that, if the guesswork is good, you're safe now, you're safe later, and you're safe when the earthquake occurs. This feels good, no doubt about it. The essence of PSHA, on the other hand, has nothing to do with the inclusion of uncertainty and probability and all the distribution functions for the guesswork. What PSHA is really telling you is how often bad happens at place *per year*.

So, a very important aspect of PSHA is that bad happens, whereas DSHA is trying to tell us that bad can't happen, if we do our MCE, ground-motion, site-response, design and construction... homework right. A second important feature of PSHA is that bad happens at a calculable rate. As we noted in the Introduction, PSHA really is mean rate of exceedance seismic hazard analysis (MROESHA). If MROESHA goes on to include weighted multiple choices, distribution functions, and uncertainty analysis, we might want to call it MROESHA with uncertainty analysis (MROESHAWUA).

In more specific—and more gruesome—terms, let's go for something really bad, an earthquake right underneath—or within 10 km of—some EUS nuclear reactor, $\mathbf{M} = 7$, say, or whatever it takes to cause catastrophic failure of the facility, core melt, and massive radioactive release. Who has what to say about this? The DSHA people say this can't happen, even though it just did for the purpose of this paragraph. The MROESHA folks are more informative: sure, this can happen, but it has less than one in a million per year chance of happening at any one site (at $\mathbf{M} = 7$, even less if $\mathbf{M} > 7$ is required for this work). The MROESHAWUA folks tell us even more: while concurring with the best estimate of $< 10^{-6}$/year at $\mathbf{M} = 7$, they state it is uncertain to a factor of π^2, or thereabouts. Here, we think are the essential differences between DSHA and PSHA.

No discussion of the pros and cons of DSHA *vis à vis* PSHA would be complete without taking note of how Ellis Krinitzsky has enlivened the debate. Ellis works for the U.S. Army Corps of Engineers, out of the Waterways Experiment Station in Vicksburg, Mississippi, and has had a long and distinguished career in geotechnical matters pertaining to the seismic safety of dams, especially along the Mississippi–Missouri Rivers drainage. When he is not taking himself too seriously, Ellis is quite the droll fellow and a hellluva good conversation. But he gets very serious about PSHA (Krinitzsky, 1993a, b, and c).

These sermons, of considerable bulk in aggregate, make for fascinating, if not especially informative reading. Their basic message is that DSHA is the only true faith in the seismic hazard business, but there are only two matters of substance here: the use and abuse of expert opinion, the basic theme of Krinitzsky (1993a) and the reliability of b-values for estimating $\mathbf{M} \gtrsim 5$ earthquakes via equation (1), the principal concern of Krinitzsky (1993c).

The use of multiple-expert opinion is a fairly recent development in Earth Sciences circles and, at least at first encounter, a faintly repugnant one, suggesting as it does a sort of "science-through-consensus" solution to the problem at hand. Few scientists in this country are trained to think along these lines. Even worse, the aggregation of multiple-expert opinion implies the dilution of "right" information from the "right" experts with "wrong" information from the "wrong" experts. All too frequently, of course, it is not easy to identify who is "right" and who is "wrong," but such dilution must nevertheless occur. Our view is that this glass is half full: while the output of a well-executed aggregation of expert opinion is unlikely to be "more right" than the "right" experts, it is quite unlikely to be really wrong, as individual "experts" have, occasionally, been known to be.

These matters will be discussed in more detail in the following section, but it is worth noting here that the use of multiple-expert opinion is here to stay for two important reasons, Krinitzsky (1993a) notwithstanding. First, diverse expert opinion is an inescapable fact of life in the regulatory arena. If you don't deal with some expert/model/opinion in advance, you can bet that expert/model/opinion will show up somewhere else in the regulatory process, almost always, it seems, in places where they are harder to deal with, intravenor proceedings for example. Second, diverse expert opinion is in itself a measure of uncertainty, a measure of confidence (or lack thereof) in the intermediate and final answers. For large, low-seismcity areas like the EUS, the uncertainty in either DSHA or PSHA is considerable, and having quantitative measures of this uncertainty is essential to rational decision-making.

Krinitzsky(1993b) is a piece entitled "The Hazard in Using Probabilistic Seismic Hazard Analysis," positioned prominently in the November, 1993, issue of *Civil Engineering*. Over the ensuing several months, *Civil Engineering* published eight or ten letters to the editor, most of which approved and applauded Ellis' improbable stand against the probable establishment. While Krinitzsky (1993b) is mostly an extended abstract for Krinitzsky (1993c), it does make the classic mistake of PSHA beginners: inferring from performance goals (of DOE in this case) of 10^{-3}/yr to 10^{-5}/yr that "probabilistic estimates are required to hundreds of thousands of years." With these performance goals, all DOE is saying is that they want the mean rate of failure $P_f \leq 10^{-3}$/yr to 10^{-5}/yr, depending on the function of the facility. As we saw earlier, we can determine the corresponding $H(a_{50})$ if we know something about $F(a)$ for each class of facilities. And as we saw even earlier than that, you don't need 10,000 years of earthquake data and ground-motion records to define a hazard estimate of 10^{-4}/yr. This is because of the area-ratio matter we described earlier. More precisely, it is because N in equation (1) is not just some number of earthquakes; it is the number of earthquakes **per unit area** per unit time with magnitude $\geq$ **M**.

This brings us to Krinitzsky (1993c), the principal concern of which is the b-value in equation (1), the Gutenberg–Richter frequency-of-occurrence relation. In blunt-instrument prose, Krinitzsky (1993c) tells us "Gutenberg–Richter b-values are dysfunctional for site-specific applications in the engineering of critical structures," nuclear reactors in the EUS for example.

An enormous literature on seismicity statistics now exists on global, regional, and local scales; b-values of 1.0 ± 0.2 are very much the rule when sufficient data are available to determine b in the first place. Over the past 15 years, a number of theoretical studies have moreover argued that b should be equal to 1 (*e.g.,* Rundle (1989) and Hanks (1992) and numerous references therein). Even in the relatively aseismic EUS (east of the New Madrid seismic zone in this case), Seeber and Armbruster (1991) found, for the regions both east and west of the Appalachian crystalline front, that the b-values are identical at 1.05 ± 0.05. While it is certainly true that there are many smaller-scale regions in the EUS that are so aseismic that b is poorly determined, so is every other seismicity measure, including the deterministic earthquake. So while Krinitzsky (1993c) correctly notes that there are a number of exceptions to the rule, they hardly suffice to prove the anti-rule.

Lost in all of the DSHA–vs–PSHA posturing of the Krinitzsky chronicles, however, is the fact that DSHA and PSHA have much more in common than they do in differences, as we mentioned earlier in this section. Moreover, when deterministic earthquakes can be associated with recurrence intervals, no matter how uncertain, they can always be found in hazard space, and we shall illustrate how for Yucca Mountain in a later section. Finally, variations on both the DSHA and PSHA themes now exist. Scenario earthquake ground-motion modeling is a worthwhile exercise for significant earthquakes that can recur on time scales comparable to the expected lifetime of the facility of interest, perhaps 100 years or so in the case of the San Andreas fault and its principal branches and 10,000 years or so in the case of Yucca Mountain. Finally considerable attention is now being directed to time-dependent hazards in those areas where the near-term, impending hazard may be considerably larger than the long-term, mean hazard, in the coastal areas of northern California, Oregon, and Washington adjacent to the Cascadia subduction zone, for example.

UNCERTAINTY, DIVERSITY, AND EXPERTS

The matters of this section have arisen earlier in this paper, in several different places, and here we shall try to bring them together and to emphasize their importance in hazard analysis of any type. In the hazard assessment business, uncertainty is important for three reasons. The first is typical of any science. Scientific results are always produced with some uncertainty, and reporting that uncertainty is an essential feature of scientific knowledge. Second, because of the strong non-linearities inherent in the construction of hazard curves, the uncertainty distribution function describing a family of calculated hazard curves is highly skewed. As a result, the important mean hazard curve depends a lot on the few largest hazard curves—and the uncertainty they represent. Finally, careful tracking of uncertainties and their propagation throughout the hazard analysis provides a sharp focus on where further research can be most beneficial in improving hazard-analysis products.

Let's return to Figure 2, our back-of-the-envelope calculation to see how uncertain input information projects through to uncertainties in the resulting hazard numbers. If we were to double all the seismicity rate numbers, for example, it should be obvious that we will double all the hazard numbers at the bottom of the figure. More generally, rewriting equation (1) in exponential form

$$N = \lambda_0 10^{-b\mathbf{M}} \; , \tag{1'}$$

where $\lambda_0 = 10^a$, if we were to say that λ_0 is uncertain to $\pm$ a factor of 2, then we would say these hazard numbers are also uncertain to a factor of 2. There are also, of course, uncertainties in the choices of b and $\mathbf{M}_{\max}$, especially in the EUS, that will lead to further uncertainties in the resulting hazard numbers.

In our back-of-the-envelope calculation, we also used the ground-motion attenuation relation of Joyner and Boore (1988), explicitly noting that it is not appropriate for use in the EUS. How much uncertainty does this introduce? In their simplest forms, these relations are generally written as

$$ln\ PGA = c + d\mathbf{M} - e\, ln\, R \pm \epsilon \tag{4}$$

where c, d, and e are constants and ϵ represents random variability from the mean relationship. What we did earlier was to fix PGA (at 0.1, 0.2 and 0.4 g) and **M** (at 5.0, 6.0, and 7.0) and invert (4) for R, ignoring ϵ. Then we went on to explore the affects of ϵ, assuming that 15% of the time PGA's would be a factor of 2 greater than the median value and 15% of the PGA's would be a factor of 2 less.

A ground-motion attenuation relation different from Joyner and Boore (1988) will have different c, d, and e and thus will give rise to different average R's for the same fixed choices of PGA and **M**. This, in turn, will lead to different hazard numbers or, if you prefer as do we, uncertainty in the resulting hazard numbers.

One can keep track of all the different possibilities for λ_0, b, $\mathbf{M}_{\rm max}$, c, d, and e (and different choices of SSZ's as we indicated in Figure 3 and other stuff as well) in things called "logic trees" (*e.g.*, Power *et al.*, 1993; Wong *et al.*, 1996). Each choice for each essential property is entered with a relative weight, which allows you to calculate zillions of hazard curves if you want to. And from all of these you can calculate the mean, median, and 15th, 85th, 5th, and 95th percentile hazard curves shown in Figure 4, for example.

Although we have mentioned them before, two important messages of Figure 4 are worth repeating here. First, while the hazard in San Francisco is much greater than in Washington, D.C. for any PGA, the uncertainty in the hazard is much greater in Washington, D.C., than for San Francisco. This is true for most sites in the EUS relative to WUS and is a straightforward consequence of the sparse seismicity and ground-motion data bases available for EUS, which in turn are consequent to the relative aseismicity of the EUS. Second, insofar as the mean hazard curve is different from the median hazard curve, uncertainty analysis really does matter in fixing the mean hazard curve, the hazard curve of choice in cost/benefit and risk/loss analyses; this difference is a function of the level of uncertainty.

Nouveau uncertainty analysis involves the distinction between **aleatory uncertainty** and **epistemic uncertainty**. Epistemic uncertainty pertains to those things we don't know but are nevertheless knowable, the average earthquake stress drop in California, for example, or in EUS. As more and more determinations become available, we will know these quantities more and more accurately. Aleatory uncertainty pertains to the variability of random happenings, the natural variations in earthquake stress drops, for example, no matter how well we know the average value.

An individual hazard curve, like our back-of-the-envelope calculation, or our modification of it to account for the "distribution contributions" of variable PGA at any $\mathbf{M}$ and R, reflect aleatory uncertainty in the number, sizes, locations, and ground motions of future earthquakes. The epistemic uncertainty bands about the median or mean hazard curves in Figure 4 reflect the limited information we have about various parameters we need in the calculations, such as λ_0, b, and $\mathbf{M}_{\max}$.

In the ground-motion part of the problem, equation (4) for example, the $\pm\epsilon$ represents the aleatory uncertainty or variability in what future ln PGA values will be from a suite of events of the same $\mathbf{M}$ and R. Epistemic uncertainty is lurking around equation (4) in two ways. First, there is uncertainty in the model parameters c, d, and e, which fix the expected or average value of ln PGA, given $\mathbf{M}$ and R. Second, the model itself is incomplete; it makes no allowance for such systematic factors as local site response, crustal structure, or whole path anelastic attenuation of the form $e^{-\pi f R/Q\beta}$.

Simple, physically based models are more interesting than the empirical models insofar as they point more directly to what we do and do not know about the Earth. Hanks and McGuire (1981) developed the following relation for PGA on a single horizontal component at the surface of a halfspace of density ρ

$$\mathrm{PGA} = 0.32 \, \frac{\Delta\sigma}{\rho R} \sqrt{\frac{f_{\max}}{f_0}} \sqrt{2 \, ln \left(\frac{2 f_{\max}}{f_0} \right)} \, . \tag{5}$$

Here $\Delta\sigma$ is the earthquake stress drop, R is hypocentral distance (again), f_0 is the earthquake corner frequency fixed by $\mathbf{M}$ and $\Delta\sigma$, and $f_{\max}$ is the high-fequency band limitation of the record (Hanks, 1982). As for equation (4), equation (5) makes no provison for local site response, crustal structure, or whole-path anelastic attenuation. By limiting ourselves to hardrock sites at close distances, say $R \leq 60$ km, we can fix $f_{\max}$, thereby concentrating all uncertainty in PGA to uncertainty in $\Delta\sigma$, because we have included no ϵ term in equation (5). This is sort of a dumb thing to do; any site-to-site variation in PGA implying a site-to-site variation in $\Delta\sigma$ for the same earthquake would belie this simplicity, but we'll ignore that for the moment.

Our aleatory uncertainty in PGA, then, is a direct and nearly linear consequence of the real or imagined aleatory variations in $\Delta\sigma$. The epistemic uncertainty in PGA arises from epistemic uncertainty in the average earthquake stress drop $\overline{\Delta\sigma}$ and model imperfections, just as in the case for equation (4). But what if we were now to include the wave-propagation effects particular to local site response and crustal structure? Surely, we would find that both the epistemic and aleatory uncertainties attributed to $\Delta\sigma$ would decrease, being translated in part to the epistemic and aleatory uncertainties associated with local site response and crustal structure. And what if we were to find in the not-so-distant future that earthquake stress drops were different in compressional regimes than in extensional regimes, a result we anticipate even now? This would allow the population of earthquake stress drops to be partitioned into two groups (at least), each presumably with significantly smaller epistemic and aleatory uncertainties than exist for the population as a whole.

The point here is that epistemic and aleatory uncertainties are fixed neither in space (across a range of models existing in 1997, say) nor in time. What is aleatory uncertainty in one model can be epistemic uncertainty in another model, at least in part. And what appears to be aleatory uncertainty at the present time may be cast, at least in part, into epistemic uncertainty at a later date. As a matter of practical reality, the trick is to make sure that uncertainties are neither ignored nor double counted. The possibilities of doing so with parametrically complex models are large.

When data are plentiful, all this aleatory/epistemic jive and uncertainty decomposition is of not much consequence. Determining ground-motion amplitude mean values and uncertainties as a function of **M** and R in coastal California is a pretty straightforward matter, requiring fairly little in the way of models, experts, and classical-language skills. All too frequently, however, the available data are incomplete or nonexistent, as in the EUS. Then we have to compute the answers, both what we want and its likely uncertainty, which is why we need to keep track of all that stuff we were talking about in the previous paragraphs. To make life worse, we also know that there are several different ground-motion estimation models out there and that they are not the same, not in their physical bases, not in their computational apparatus, and not in the resulting answers. But these models are different only because different knowledge has been imparted to them by the experts responsible for them, but knowledge itself resides with the experts. Thus there must be uncertainty in our individual and collective knowledge. A distressing corollary of this is that some experts must be more knowledgeable than other experts, but we don't know which from which. So how does one deal with diverse expert knowledge?

Let's say, for example, that we need to know about PGA's for $\mathbf{M} = 6$ earthquakes at $R = 10$ km for one or more sites in the EUS, both the mean value under these conditions as well as the distribution of PGA's about the mean. There are no instrumental recordings of earthquake ground motions under these conditions, so we assemble a team of experts to provide it for us. In this case, diverse expert opinion will become surrogate data, and such a mapping of experts into data is a novel and still suspect notion in this country. Most scientists prefer their data from well calibrated instruments, not from poorly calibrated other scientists who are easily identified as such because they never seem to have read your papers.

But there are still no data, and we still need the answer, so we press on, asking each member of our team for his or her best estimate of the mean PGA, $\overline{a}$. Each expert may use one or more ground-motion models to calculate the answer; or perhaps read up on what other EUS ground-motion data are available; or go check out foreign data sets in similar circumstances; or do a back-of-the-envelope calculation—or maybe all of the above and more, too. We also ask each expert for his or her estimate of the epistemic uncertainty in his or her estimate of $\overline{a}$, as well as their estimates of the aleatory distribution of PGA's about $\overline{a}$. And if we're really cool—and one of us is—we'll ask them for their estimate of the uncertainty in their aleatory-uncertainty estimate. From all this, we can construct a team distribution of PGA for $\mathbf{M} = 6$ and $R = 10$ km with mean value $\overline{A}$ and standard deviation $\sigma_{\overline{A}}$, reflecting both the uncertainties in every expert's estimates and expert-to-expert variability or diversity. What is this worth?

Scientifically, we would say that the team distribution function is robust if it represents the body and range of informed scientific opinion, in this case on what peak accelerations at $R = 10$ km for $\mathbf{M} = 6$ earthquakes should be in the EUS. It may not be true to the

Earth, but in the absence of any data to the contrary (or any data at all, in this case), there is no way to falsify it, if it indeed represents the full range of differing opinion. The cost of this robustness is a larger uncertainty for the team distribution than any one expert would alone supply. (Experts almost always think they know their stuff with greater accuracy than they really do, at least if other, equally well-qualified experts are judges of the matter.) But this is a small price to pay for hedging your bets against being out on one tail or another of the team distribution, where any one expert can always be. Put another way, experts are not resolving, let alone reducing uncertainty in an exercise of this sort; indeed, in their differing assessments, they are the source of it.

Defining the body and range of informed opinion in explicit ways amenable to quantitative analysis from a team or teams of experts holding diverse views is the trick, of course, and this has been a rapidly evolving subject in the 1990's. Interested readers may wish to explore the appropriate sections of the report of the Senior Seismic Hazard Analysis Committee (SSHAC, 1997), the review of the SSHAC Report by the National Research Council Panel of the National Academy of Sciences (Panel, 1997), and the probabilistic volcanic hazard analysis (PVHA) at Yucca Mountain recently conducted for DoE (Geomatrix/TRW, 1996).

SSHAC (1997) finds that strong and extensive but directed and facilitated interaction among the experts is an essential feature of eliciting the most meaningful information from them. Experts often do not understand each other's work and model in sufficient detail for the work at hand, and thorough discussion of such misunderstandings is an essential prelude to the elicitation and aggregation processes. Experts often bring very different experiences to their common field of expertise, and differences in opinion naturally arise from these different experiences. All of the relevant data and information must be made uniformly available to all of the experts. Experts are commonly distracted by interesting scientific questions that are irrelevant to the problem at hand. Perhaps most importantly, the experts need to know that they have been convened to represent and quantify the current body and range of scientific opinion on the matter at hand, not to decide who is "right" and who is "wrong" about this matter. Using experts successively in the roles of proponents, evaluators, and integrators of models/knowledge shows promise for achieving this goal. The SSHAC (1997) recommendations on the principles and procedures of expert elicitation and aggregation may be found in Appendix J (64 pages). Practical application and implementation of these principles and procedures to ground-motion estimation in the EUS may be found in Appendix A (72 pages) and Appendix B (511 pages).

FROM HAZARD TO GROUND MOTION: BARRY BONDS, DE-AGGREGATION, AND UHS

Four sections ago, we noted how a range of earthquake sizes contributed to the hazard, which is to say the MROE, of 0.2 g (Figure 2). To fix this idea in mind, we offer the following analogy, born in part of our baseball deprivation in the summer and fall of 1994. While we realize that some of our audience may not appreciate the mysteries of baseball, it just has to be true that most Americans know a lot more about baseball than PSHA. So here we go. There are two men on base when Barry Bonds comes to the plate. You slip off to the fridge for a beer, and when you return, two runs have scored. What magnitude of hit (single, double, or triple) did Barry deliver? (We know it is not a home run because then three runs would have scored.)

A little more information goes a long way. Let's say the runners were on second and third, "close distance" to home. In this case, we can be pretty sure that Barry hit a single. This is because Barry, like every other ballplayer, hits a lot more singles than doubles and a lot more doubles than triples, sort of like earthquakes. But a double or a triple will do the job, too, even if their rates of occurrence, or probability, are less and even less respectively.

But what if runners were on first and second, "far distance" from home? For this situation, we need a larger magnitude hit, because a runner on first will only rarely score on a single. A double does the job some of the time but not all of the time. If we could reduce the "uncertainty" here, by knowing how fast the first-base runner was, for example, or whether there were two out, we could get a better fix on whether a double would have sufficed to drive in the runner from first. Even if only half of Barry's doubles drive in a runner from first, however, that is still a lot more than the number of triples he hits. So we're pretty sure Barry scorched out a double. But we're not perfectly sure: the infrequent triple certainly does the job, and very infrequently a single does it, too.

So far, so good with seismological baseball: to get the same action at home plate, or the same ground-motion level in your backyard, Barry needs a larger magnitude hit when the runners are at far distance than when they are at close distances. Not counting home runs, so characteristic of Barry, his distribution of singles, doubles, and triples reminds us of the distribution of small, medium, and large earthquakes. But what if we hadn't known whether the runners were on first and second or on second and third? There are probabilities associated with each of these configurations as well; it is considerably more likely that the runners are on first and second than on second and third.

Working through all the combinations of earthquake-magnitude and epicentral-distance ranges that contribute to a chosen hazard level is what is known as de-aggregation. To see how this stuff works, we return to Figure 2, yet again, to de-aggregate the hazard of 0.18×10^{-4}/yr for 0.2 g, working with the original numbers unmodified by the dispersion considerations giving rise to the Σ^{σ} column. The contribution of 0.06 (leaving the 10^{-4}/yr as implicit for the remainder of this paragraph) for $\mathbf{M} = 5$ and $R \leq 3.2$ km is easy enough to deal with; it says that 1/3 of the total hazard comes from $\mathbf{M} = 5$ earthquakes at $R \leq 3.2$ km. The $\mathbf{M} = 6$ contribution of 0.09 for $R \leq 12$ km is a little trickier because part of this ($0.09 \times 3.2^2/12^2$) comes from $R \leq 3.2$ km and the rest from $3.2 < R \leq 12$ km. Similarly, the contribution of 0.03 from $\mathbf{M} = 7$ earthquakes must be divided into 3 parts, $0.03 \times 3.2^2/22^2$ for $R \leq 3.2$ km, $0.03(12^2/22^2$ - $3.2^2/22^2)$ for $3.2 < R \leq 12$ km, and the remainder for $12 < R \leq 2.2$ km. If we now divide all of these numbers by the total hazard of 0.18, we can plot the fractional contribution to the total hazard for each pair of magnitude and distance ranges.

Rather than do this for the back-of-the-envelope data, we show two, more formal de-aggreations in Figure 7 for a site in South Carolina. (De-aggregation of the seismic hazard at numerous EUS sites may be found at http://gldage.cr.usgs.gov/eq) Figure 7a is for the case of a hazard of 5×10^{-5}/year for a PGA of 0.23 g, and Figure 7b is for the case of a hazard of 1×10^{-4}/yr for a 1 Hz spectral acceleration of 18.8 cm/sec. In the former case, the most likely combination to cause PGA ≥ 0.23 g, is a very close magnitude 5 to 5.5. This would be a logical DBE for a PGA-sensitive structure if the performance criteria called for a P_f in the 5×10^{-5}/yr range. On the other hand, a magnitude 6 to 6.5 at 100 to 150 km is more appropriate—but hardly uniquely so—for a longer period structure (1 Hz) with a 10^{-4}/yr performance goal. There is certainly no single (deterministic) earthquake (*i.e.*, magnitude-distance pair) that captures all these factors.

These two figures reveal an additional seismological feature for which Barry Bonds does not provide us with an analogy, namely that both the source-excitation and distance-attentuation of ground motion are strong functions of frequency. To deal with these period-dependent phenomena, PSHA cogniscenti work with what's called the uniform hazard spectrum (UHS) which is an ensemble of response spectrum ordinates as a function of period, each calculated for the same MROE, that is to say, hazard. Just as was the case in Figure 7, UHS (Figure 8a) at short periods depends on the more frequent, small earthquakes at close distances, but at longer periods, the greater source excitation of less frequent, large earthquakes even at considerable distance wins out. Figure 8b shows how UHS varies according to choice of MROE.

The problem here, which should come as no surprise at this point, is that UHS does not correspond to a single ground-motion time history arising from a single earthquake at a specific distance, a deterministic earthquake, say. But from a first-order design point of view, say for a single-degree-of-freedom elastic oscillator, what difference does it make whether the UHS (or its ground-motion, time-history equivalent) corresponds to a $M \simeq 5$ earthquake at close distance that is enriched in long-period motion, or to a $M \simeq 7$ earthquake at far distance that is enriched in high-frequency motion, or to a $M \simeq 6$ earthquake at intermediate distance that is slightly enriched in both? With respect to second-order issues, of course, UHS—or any other form of elastic response spectra—have many limitations when applied to real structures, especially in the post-elastic regime.

In the EUS, the great difficulty in making sense of real and imagined seismic hazards relates to the fact that, with the possible exception of the New Madrid seismic zone, throughgoing, seismogenically active, crustal fault zones evidently do not exist. While this inference arises from an earthquake history that is, perhaps, as yet too short to draw it, it is nevertheless strongly reinforced by the nearly complete absence of surface faulting in the EUS. While SSZ's of greater and lesser seismicity can be grossly defined, their smaller dimension is rarely less than 100 km. Only in very restricted areas of the EUS can one say with confidence that a magnitude x, y, or z earthquake will occur here, where "here" has a location uncertainty of $\lesssim$ 100 km (if not more), almost always because one or more earthquakes occurred at "here" in the historical record. And even these very few positive identifications (assuming they are correct) hardly preclude the occurrence of $M = x$, y, or z earthquakes elsewhere. Thus, for most of the EUS, the deterministic earthquake is poorly defined with respect to either size or location. But no matter how arbitrary the size and location of the deterministic earthquake may be, we can always find it in hazard space, so long as we know something about its frequency of occurrence or recurrence interval.

In the WUS, we are on less shaky ground, at least with respect to defining deterministic earthquakes. Not only is the earthquake data base far richer, if even shorter in years than for the EUS, but in many areas paleoseismic investigations have extended the earthquake history back many thousands of years. And in numerous places, recurrent earthquakes of about the same size have occurred along the same fault segment. From such observations was born the concept of characteristic earthquakes (Schwartz and Coppersmith, 1984), that the same fault segment is visited by the same maximum-size earthquakes that dominate that segment's slip rate. And, just as importantly, these same paleoseismic investigations often as not provide estimates of their recurrence intervals. Characteristic earthquakes with known recurrence intervals will figure prominently as deterministic earthquakes at Yucca Mountain, the next great playing field for PSHA *vs.* DSHA. But here they are one and the same: one simply reckons the hazard of any level of ground motion for any deterministic earthquake of interest.

YUCCA MOUNTAIN AND ELLIS IN WONDERLAND

Yucca Mountain is the site proposed for an underground repository for the nation's nuclear waste. It straddles the western boundary of the Nevada Test Site 100 or so miles northwest of Las Vegas, in the southern Basin and Range province. When the sun stands tall on Yucca Mountain, a numbing whiteness, the output of our principal source of nuclear energy, dilutes the place. The sky turns the thinnest of blues and the sagebrush the palest of greens. The landscapes lose form as if one of the spatial dimensions got lost, and time just bumps along, with no clear direction in mind. You feel flat and blank and abandoned, and you keep an eye on your shadow which would just as soon evaporate on you. All in all, not a bad spot for a nuclear waste repository, so long as this is O.K. with the desert tortoises.

In the cool of the morning, though, or when the afternoon stretches into evening, the shapes and shades slip back in with the shadows. Apart from a few young cinder cones off to the west in Crater Flat and the recent incision of Forty-Mile Wash, the impression that lasts is the smoothness of the landforms and their topography. The rocks exposed on Yucca Mountain are late-Tertiary volcanic ashes and tuffs, uplifted along the Solitario Canyon fault and back-tilted to the east. The surface expression of the Solitario Canyon fault, which sits at the base of the western flank of Yucca Mountain, is almost invisible except when the shadows are long in the early morning, and the ridges between the modern drainages are all evenly rounded off. Even the cliff-formers on the exposed western face of Yucca Mountain are smooth in their second spatial derivatives. Maybe there has been a meter or so of displacement along the Solitario Canyon fault in the past 50,000 to 100,000 years, but no more than that, which is pretty much the same story that the trenches across the fault tell.

There are, however, a number of Quaternary faults within walking distance of the proposed site for handling radioactive waste on the east flank of Yucca Mountain (Figure 9). With the exception of the Ghost Dance fault, all of these have ruptured 2 to 3 times in the last 100,000 years or so, with individual events having displacements as large as 1.0 to 1.2 m on the Solitario Canyon and Paintbrush Canyon faults (Menges *et al.*, 1994). Altogether, these faults have produced four or five $\mathbf{M} \simeq 6^1\!/_2$ earthquakes in the past 100,000 years, one or two of which may have been a bit bigger, if the Paintbrush Canyon/Stagecoach Road faults ruptured simultaneously. $\mathbf{M} \simeq 6^1\!/_2$ earthquakes at such close distances make for a helluva ground-motion ride, and peak accelerations in excess of $^1\!/_2$ g are a common occurrence in these circumstances (Figure 5). It is easy to see how these earthquakes and their ground motions could figure into DSHA for Yucca Mountain, should one wish to proceed on this basis.

But these earthquakes and the ground motion that arise from them are very well defined in hazard space as well. Without the benefit of a formal analysis, we can hazard a guess that the hazard or MROE of $^1\!/_2$ g or greater ground acceleration will be something like 3 or 4 or 5×10^{-5}/yr for any site in or around Yucca Mountain. As usual, larger MROE's will be associated with lower PGA's and smaller MROE's will be associated with higher PGA's. The choice of design basis ground motions then comes down to the choice of performance goal, what we were calling P_f earlier, and attendant seismic design criteria. But it's not hard to figure out where the action will be on the DSHA–vs–PSHA playing field. At a design hazard level of 2×10^{-5}/yr, the deterministic earthquakes will figure prominently in fixing the design ground motions and dominate the de-aggregation at this hazard level. For a design hazard level of 2×10^{-4}/yr, the deterministic earthquakes will contribute hardly at all. This is just what happened (Wong *et al.*, 1996).

Hazard space, then, is a fascinating seismological Wonderland of earthquake sizes, epicentral distances, recurrence intervals, ground-motion attenuation relationships (often as a function of period), and the hazard numbers themselves. For proponents of DSHA and PSHA alike, the important matter is that any deterministic earthquake, when specified by its size, location, and recurrence interval, and the ground motion arising from it can always be found in Wonderland. In the case of Yucca Mountain, this is easy to see. In the EUS, specifying a deterministic earthquake with respect to size, place, and recurrence interval is, in general, problematic and arbitrary. Even worse—or even better, depending on your position on DSHA—any deterministic earthquake has a zero probability of occurrence when cast in the framework of areally distributed seismicity. But this mathematical glitch can be overcome by choosing magnitude and distance **ranges** for our deterministic event(s). Indeed, Figure 7, which we used earlier to illustrate de-aggregation, can also be viewed as a suite of deterministic earthquake choices, specified in terms of half-magnitude units and the indicated distance ranges. And, as we also saw earlier, we know the hazard contribution for each and every one of these choices. So we know you're out there somewhere, Ellis, and we can calculate your hazard anytime we want to.

ACKNOWLEDGMENTS AND OTHER THINGS

This manuscript evolved fitfully over a period of almost four years and supersedes an earlier version (Hanks and Cornell, 1994). Much of what is written here arose in connection with meetings of the Senior Seismic Hazard Analysis Committee and the report (SSHAC,

1997) ultimately produced from these deliberations. One of us (CAC) served on that committee, together with R. J. Budnitz (chair), G. Apostalakis, D. M. Boore, L. S. Cluff, K. J. Coppersmith, and P. A. Morris. TCH attended many of the SSHAC meetings as part of his responsibilities to the Panel on Seismic Hazard Evaluation. This Panel (Carl Kisslinger, chair) was constituted by the Committee on Seismology of the National Research Council of the National Academy of Sciences, at the request of the United States Nuclear Regulatory Commission to review and evaluate the SSHAC Report.

Ellis Krinitzsky figures prominently in this essay. We suspect that we have enduring differences of opinion with him on matters we address here and that he might choose words differently than we have in addressing yet other matters. Nevertheless, Ellis has, we think, forced our thinking to places it might not have gotten under its own momentum, and for this we are indeed appreciative. D. M. Boore, J. R. Kimball, and J. W. Whitney provided us information in advance of publication (Figures 5, 7, and 9, respectively) and joined us in several useful conversations. D. M. Boore, W. B. Joyner, and R. K. McGuire offered many useful suggestions in reviewing this manuscript.

If this manuscript has any value at all, it will reside with those readers who, much like TCH, have come to realize that PSHA is more informative and more interesting than it might otherwise seem to be. Those interested in the historical development, other points of view, or just more information on PSHA may wish to peruse Cornell and Merz (1975), Panel on Seismic Hazard Analysis (1988), McGuire and Arabasz (1989), Reiter (1990), and/or Kramer (1996), in addition to the references cited in the text. This work was supported in part by the U.S. Nuclear Regulatory Commission.

REFERENCES

Boore, D. M., W. B. Joyner, and T. E. Fumal, 1993, Estimation of response spectra and peak accelerations from western North American earthquakes: an interim report, *U.S. Geological Survey Open-File Report 93–509*, 72 pp.

Cornell, C. A., 1968, Engineering seismic risk analysis, *Bulletin of the Seismological Society of America 58*, 1583–1606.

Cornell, C. A., and H. A. Merz, 1975, Seismic risk analysis of Boston, *Journal of Structures Division 101*, 2027–2034, American Society of Civil Engineers.

DOE STD 1020, 1994, Natural phenomena hazards and design and evaluation criteria for Department of Energy facilities, *DOE–STD–1020–94*, U.S. Dept. of Energy, Washington, D.C.

Geomatrix/TRW, 1996, Probabilistic volcanic hazard analysis for Yucca Mountain, Nevada, *BA0000000–01717–2200–00082, Rev. 0*, U.S. Department of Energy.

Hamburger, R. O., 1996, Implementing probability-based design in structural engineering practice, *Proceedings of 11th World Conference of Earthquake Engineering*, Acapulco, Mexico.

Hanks, T. C., 1982, $f_{\max}$, *Bulletin of the Seismological Society of America 72*, 1867–1879.

Hanks, T. C., 1992, Small earthquakes, tectonic forces, *Science 256*, 1430–1432.

Hanks, T. C., and C. A. Cornell, 1994, Probabilistic Seismic Hazard Analysis: A Beginner's Guide, *in* Proceedings of the Fifth Symposium on Current Issues Related to Nuclear Power Plant Structures, Equipment and Piping, I/1–1 to I/1–17, North Carolina State University, Raleigh.

Hanks, T. C., and R. McGuire, 1981, The character of high-frequency strong ground motion, *Bulletin of the Seismological Society of America 71*, 2071–2095.

Joyner, W. B., and D. M. Boore, 1988, Measurement, characterization, and prediction of strong ground motion, *Proceedings of Earthquake Engineering & Structural Dynamics*

II, Geotechnical Division, American Society of Civil Engineers, Park City, Utah, 27–30 June, 1988, 43–102.

Kimball, J. K., and A. Bieniawski, 1994, Use of probabilistic earthquake models for design and evaluation, *Proceedings of 26th Joint Meeting, U.S.–Japan Cooperative Program in Natural Resources Panel on Wind and Seismic Effects*, National Institute of Standards and Technology, Gaithersburg, MD.

Kramer, S. L., 1996, *Geotechnical Earthquake Engineering*, Prentice Hall, Upper Saddle River, New Jersey.

Krinitzsky, E. L., 1993a, Earthquake probability in engineering–part 1: the use and misuse of expert opinion, *Engineering Geology 33*, 257–288.

Krinitzsky, E. L., 1993b, The hazard in using probabilistic seismic hazard analysis, *Civil Engineering* (Nov.), 60–61.

Krinitzsky, E. L., 1993c, Earthquake probability in engineering–part 2: earthquake recurrence and limitations of Gutenberg–Richter *b*-values for the engineering of critical structures, *Engineering Geology 36*, 1–52.

Lewis, H. W., 1990, *Technological Risk*, W. W. Norton & Company, New York.

McGuire, R. K., and W. J. Arabasz, 1989, An introduction to probabilistic seismic hazard analysis, *Society of Exploration Geophysics*, Special Publication on Environmental Geophysics, S. H. Ward, Editor, Feb.

Menges, C. M., J. R. Wesling, J. W. Whitney, F. H. Swan, J. A. Coe, A. P. Thomas, and J. A. Oswald, 1994, Preliminary results of paleoseismic investigations of Quaternary faults on eastern Yucca Mountain, Nye County, Nevada, *Proceedings of Fifth International Conference, High-Level Radioactive Waste Management IV* 2373–2390.

Panel on Seismic Hazard Analysis, 1988, *Probabilistic seismic hazard analysis*, National Academy Press, Washington, D.C.

Panel on Seismic Hazard Evaluation, 1997, *Review of recommendations for probabilistic seismic hazard analysis, guidance on uncertainty and use of experts*, National Academy Press, Washington, D.C.

Power, M. S., R. P. Youngs, C.–Y. Chang, K. Sadigh, K. J. Coppersmith, C. L. Taylor, J. Penzien, W. S. Tseng, N. Abrahamson, and J. H. Gates, 1993, Development of seismic ground motions for San Francisco Bay bridges, *Proceedings of First U.S. Seminar, Seismic Evaluation and Retrofit of Steel Bridges*, San Francisco.

Reiter, L., 1990, *Earthquake Hazard Analysis; Issues and Insights*, Columbia University Press, New York.

Rundle, J. B., 1989, Derivation of the complete Gutenberg–Richter magnitude-frequency relation using the principle of scale invariance, *Journal of Geophysical Research 94*, 12,337–12,342.

Schwartz, D. P., and K. L. Coppersmith, 1984, Fault behavior and characteristic earthquakes: examples from the Wasatch and San Andreas fault zones, *Journal of Geophysical Research 89*, 5681–5698.

Seeber, L., and J. G. Armbruster, 1991, The NCEER–91 earthquake catalog: improved intensity-based magnitudes and recurrence relations for U.S. earthquakes east of New Madrid, *NCEER–91–0021.*

SSHAC, 1997, Senior Seismic Hazard Analysis Committee, Recommendations of proabilistic seismic hazard analysis: guidance on uncertainty and use of experts, *U.S. Nuclear Regulatory Commission*, in press.

Wong, I. G., S. K. Pezzopane, C. M. Menges, R. K. Green, and R. C. Quittmeyer, 1996, Probabilistic seismic hazard analysis of the exploratory studies facilities at Yucca Mountain, Nevada, *Proceedings, Methods of Seismic Hazards Evaluation, Focus '95*, American Nuclear Society.

FIGURE CAPTIONS

Figure 1. PSHA: Snippets of the fur.

Figure 2. PSHA: A back-of-the-envelope calculation.

Figure 3. PSHA: Inhomogeneous seismicity.

Figure 4. PSHA: Seismic hazard curves for (a) a site in San Francisco (Power *et al.*, 1993) and (b) a site in Washington, D.C. (SSHAC, 1997). See text for explanation of open and solid circles, which overlap at 0.1 g.

Figure 5. Peak acceleration for the Northridge, California, earthquake (Jan. 17, 1994; M = 6.7) with the regression curves (median and $\pm\sigma$) of Boore *et al.* (1993). Courtesy of D. M. Boore.

Figure 6. Forming P_f: The fragility and hazard curves.

Figure 7a. De-aggregation: peak acceleration (Kimball and Bieniawski, 1994).

Figure 7b. De-aggregation: 1 Hz spectral acceleration (Kimball and Bieniawski, 1994).

Figure 8. The Uniform Hazard Spectrum, from Reiter (1990). (a) as a function of period for a hazard of 10^{-3}/yr (b) as a function of hazard level and period for the Vogtle site.

Figure 9. Principal faults of the Yucca Mountain area. Courtesy of J. W. Whitney.

Weight for sth seismicity expert

A discussion of the background for evaluating a single weight for each seismicity expert is given in Section C.3.4. The weight for the sth seismicity expert, W_s, is the weighted average of the self weights in the four regions, i.e.

$$W_s = \sum_{w=1}^{4} W_{sw} \hat{P}_s(A = A_w)$$

where $\hat{P}_s(A = A_w)$ is the estimate, based on the sth expert's best estimate inputs, of the probability that the maximum PGA at the site is due to an earthquake originating in a zone in the wth region, which is the normalized value of

$$\hat{P}_s(A = A_w) = \{\sum_{a_i} [\prod_{w' \neq w} \hat{P}_s(A_{w'} \leq a_i)][\hat{P}_s(A_w \leq a_{i+1}) - \hat{P}_s(A_w \leq a_i)]\} / \hat{P}_s(A > a_i) \quad (C.4)$$

where $\hat{P}_s(A_w \leq a_{I+1}) = 1$

o a categorical correction, i.e., a more extensive catagorization of site soil types

$$\lambda_a = \lambda_0 \int_{M_0}^{M_U} \int_{r>0} P(A > a|m,r) f_R(r) dr\, dF_M(m|M_0,M_U)$$

Finally, under the assumption that events between zones are independent, the seismic hazard in t years at a site can be evaluated by

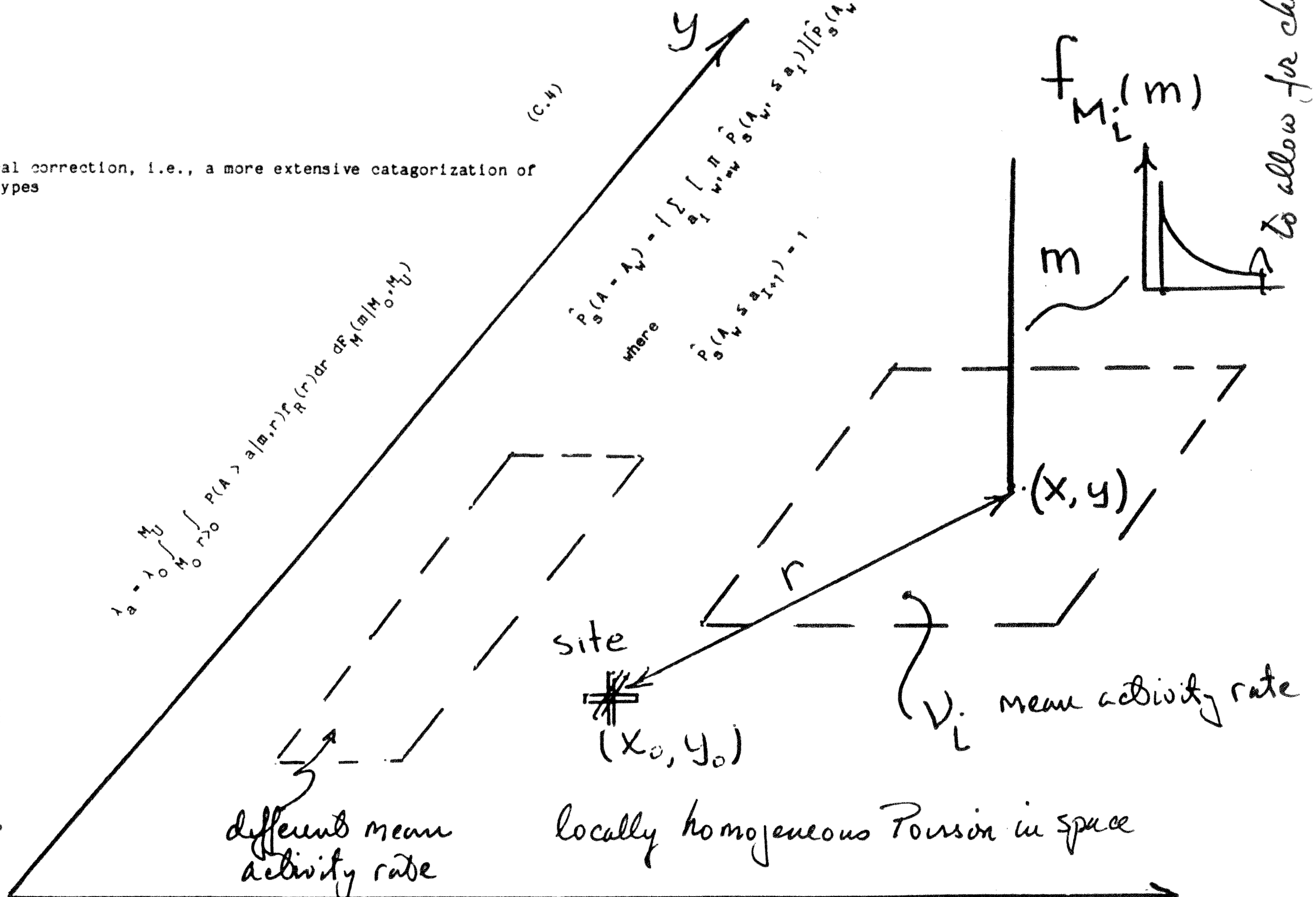
$$P(A_t > a) = 1 - \Pi_q[1 - P_q(A_t > a)]$$

$$= 1 - \Pi_q\{\exp[-t \sum_{j=1}^{J} \lambda_q(m_j) \sum_{k=1}^{K} \Pi_q(r_k) P(A > a|m_j, r_k)]\} \quad (C.9)$$

Best estimate hazard at the site due to events over all zones in the best estimate map

$$\hat{P}_{su}(A_t > a) = 1 - \prod_q \exp[-t \sum_{j=1}^{J} \hat{\lambda}_{sq}(m_j) \sum_{k=1}^{K} \pi_{sq}(r_k) \hat{P}_u(A > a | m_j, r_k)],$$

for $a = a_1, a_2, \ldots, a_I$



$$r = \sqrt{(x-x_0)^2 + (y-y_0)^2}$$

$$p = f(m, r)$$

$$= 1 - \Phi\left(\frac{\ln a - \ln(g(m,r))}{\sigma_{\ln a}}\right)$$

← ground-motion attenuation law to give expected value

↳ standard deviation about the expected value

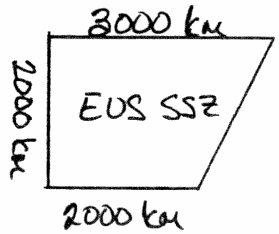
2. Truncated Exponential Model

A second method for adjusting the exponential magnitude-recurrence model in Eq. C.12 is based on assuming the distribution of magnitudes, conditional on $M_0 < m < M_U$, to be a truncated exponential distribution. That is,

$$P(M > m|M_o, M_U) = \frac{e^{-\beta(m-M_0)}[1 - e^{-\beta(M_U-m)}]}{[1 - e^{-\beta(M_U-M_0)}]} \quad (C.14)$$

Figure 1

Eastern United States Seismic Source Zone



Seismicity Rate for EUS SSZ

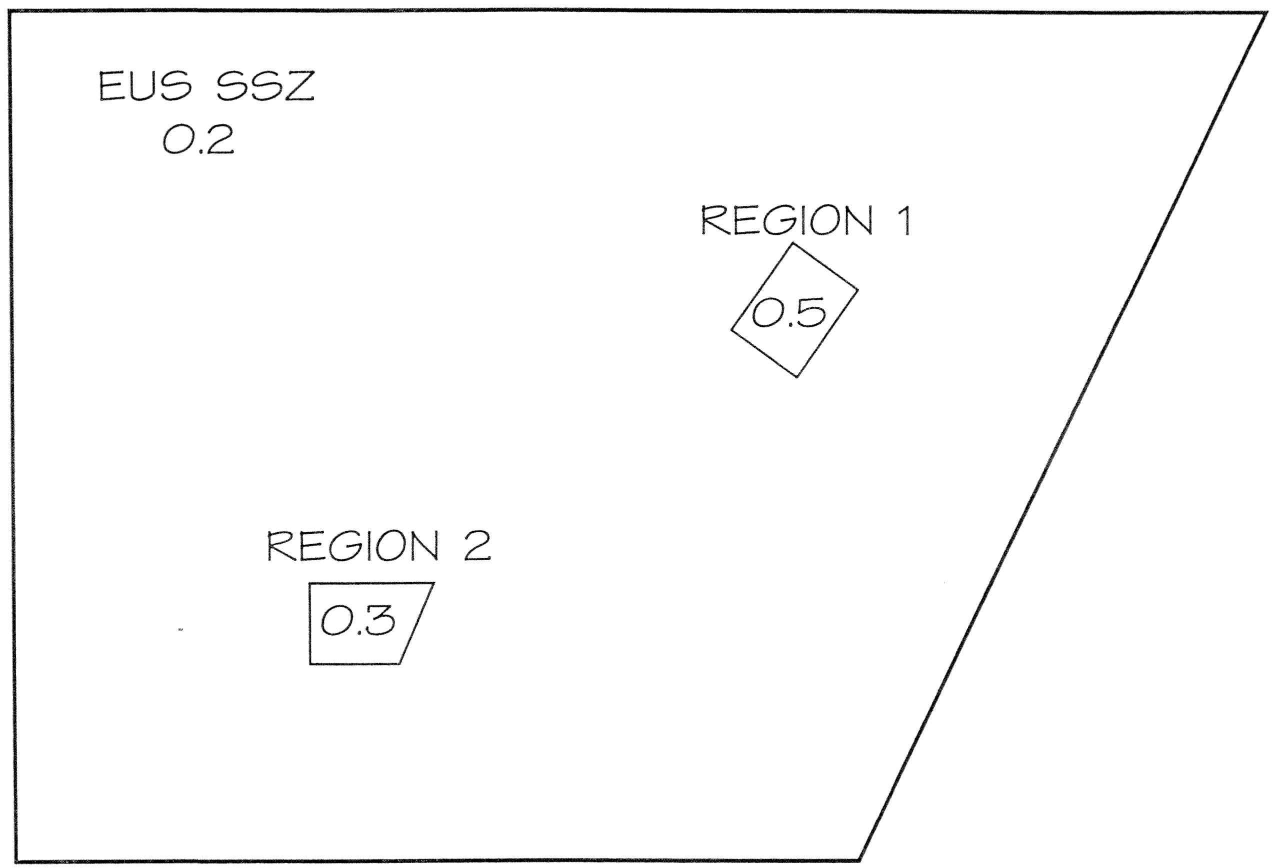
one M=5 per year
one M=6 per decade
one M=7 per century

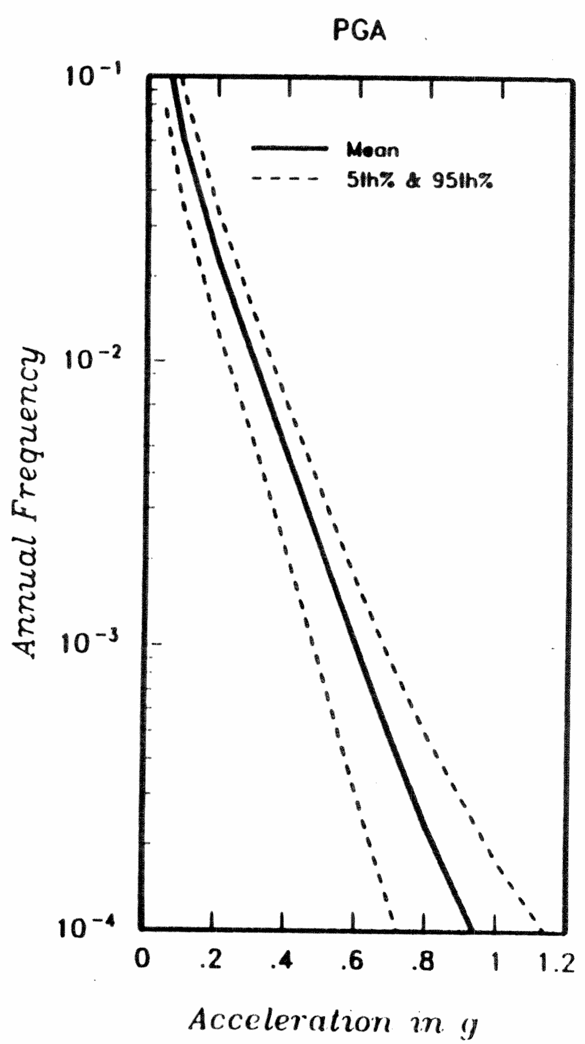
Horizontal distances (km) within which the given PGA's are achieved or exceeded for the given magnitudes

	M=5	M=6	M=7
0.1g	14	25	41
0.2g	3.2	12	22
0.4g	0	0	10

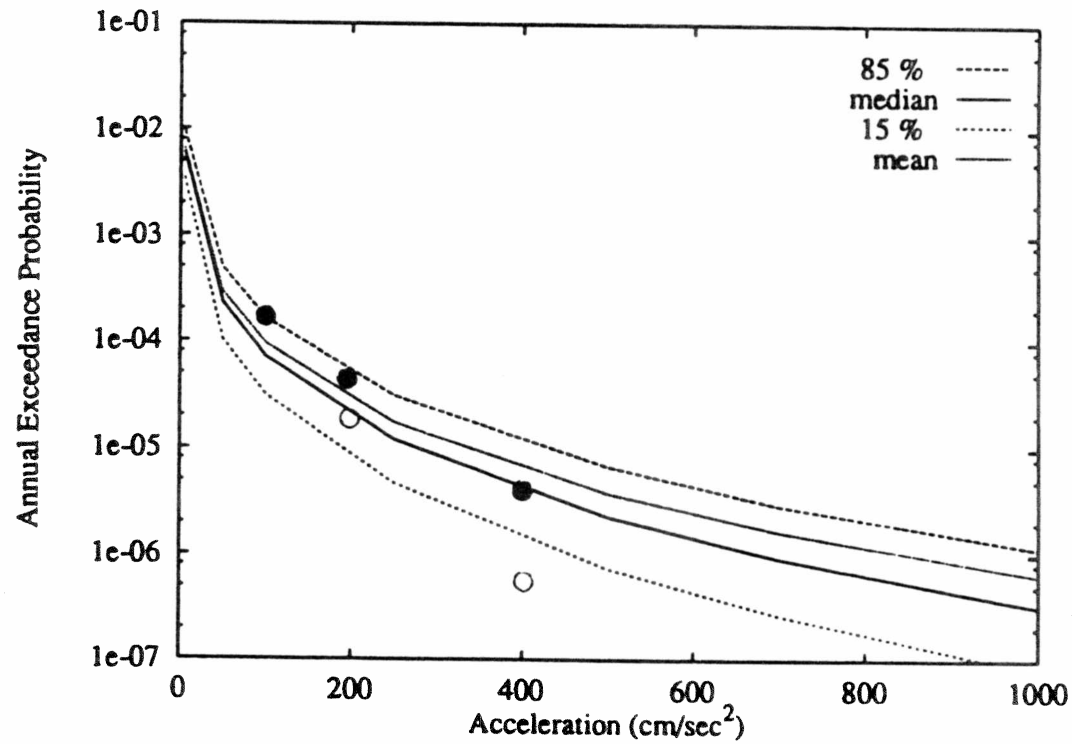
Mean rate of exceedance (MROE), $\times 10^{-4}$ per year, for the given PGA's for the given magnitude earthquakes

	M=5	M=6	M=7	Σ	Σ^{σ}
0.1g	1.23	0.39	0.11	1.73	1.47
0.2g	0.06	0.09	0.03	0.18	0.41
0.4g	0	0	0.006	0.006	0.034

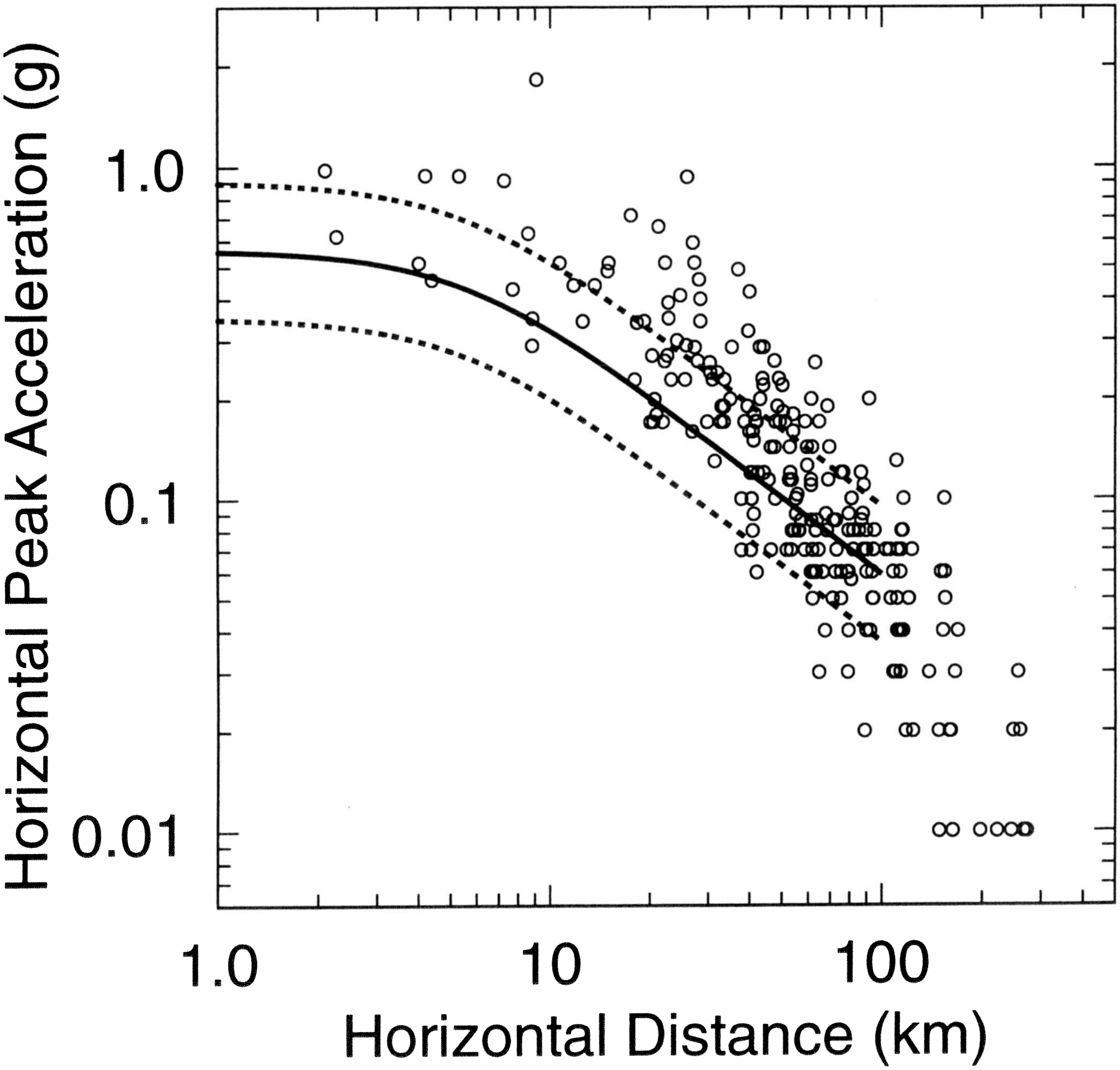

EUS SSZ
0.2
REGION 1
0.5
REGION 2
0.3


PGA
Mean
5th% & 95th%
Annual Frequency
10^{-1}
10^{-2}
10^{-3}
10^{-4}
0
.2
.4
.6
.8
1
1.2
Acceleration in g

(a)

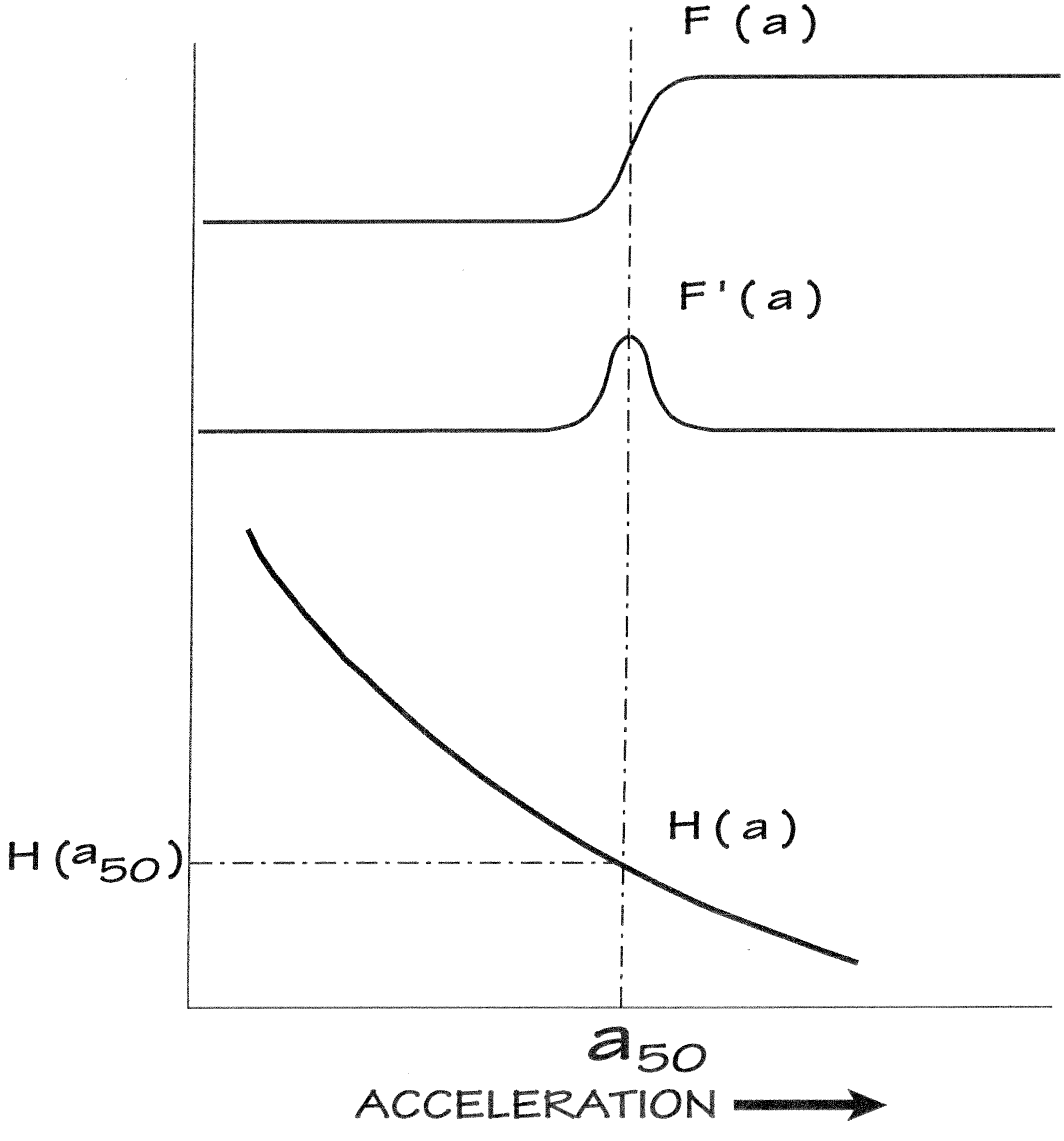

85 %
median
15 %
mean
Annual Exceedance Probability
1e-01
1e-02
1e-03
1e-04
1e-05
1e-06
1e-07
0
200
400
600
800
1000
Acceleration (cm/sec^2)

(b)



Horizontal Peak Acceleration (g)
1.0
0.1
0.01
1.0
10
100
Horizontal Distance (km)

Figure 5



F (a)
F'(a)
H(a)
H(a50)
a50
ACCELERATION

EPRI MEAN PROBABILISTIC SEISMIC HAZARD
PEAK GROUND ACCELERATION

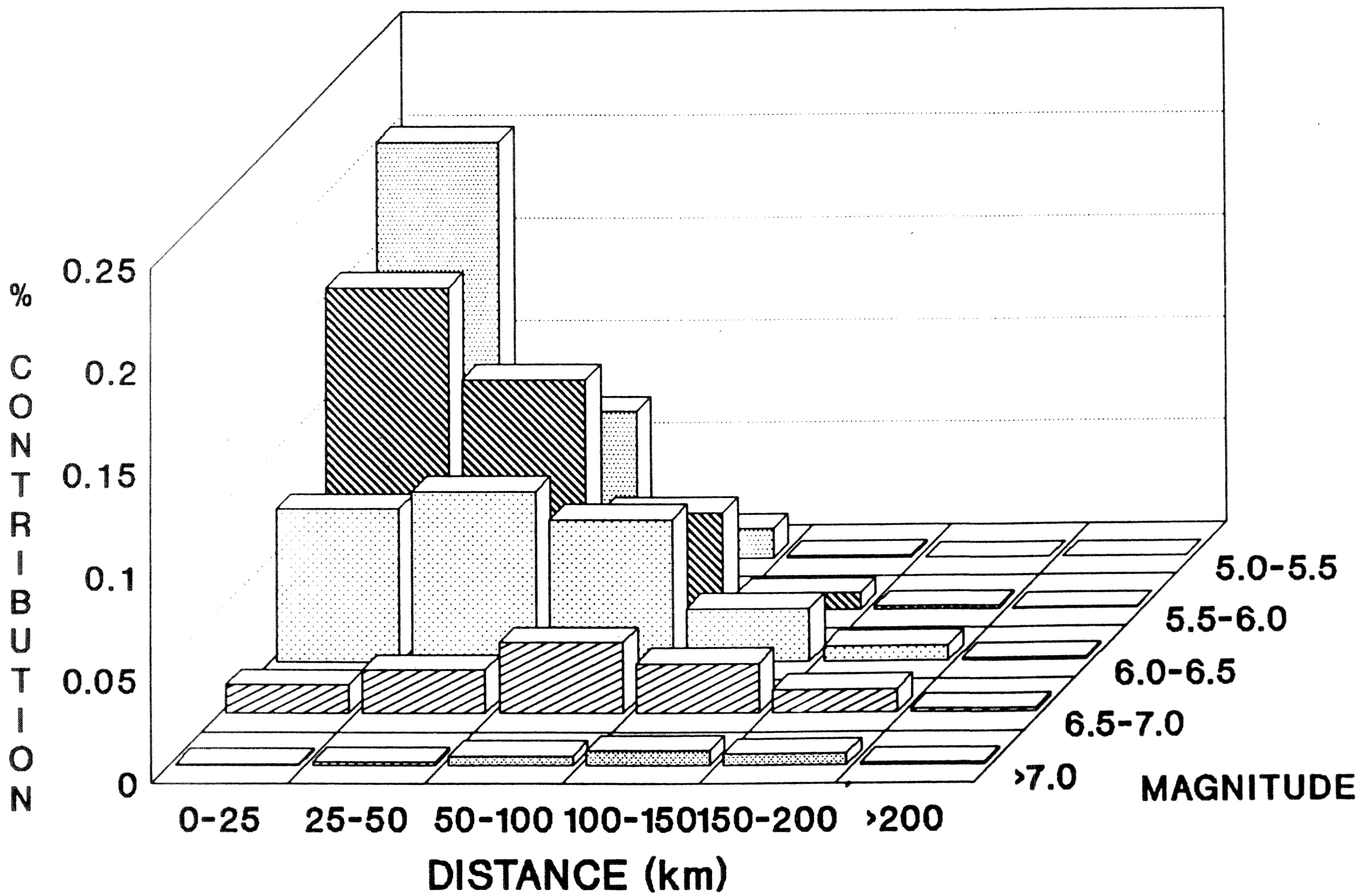
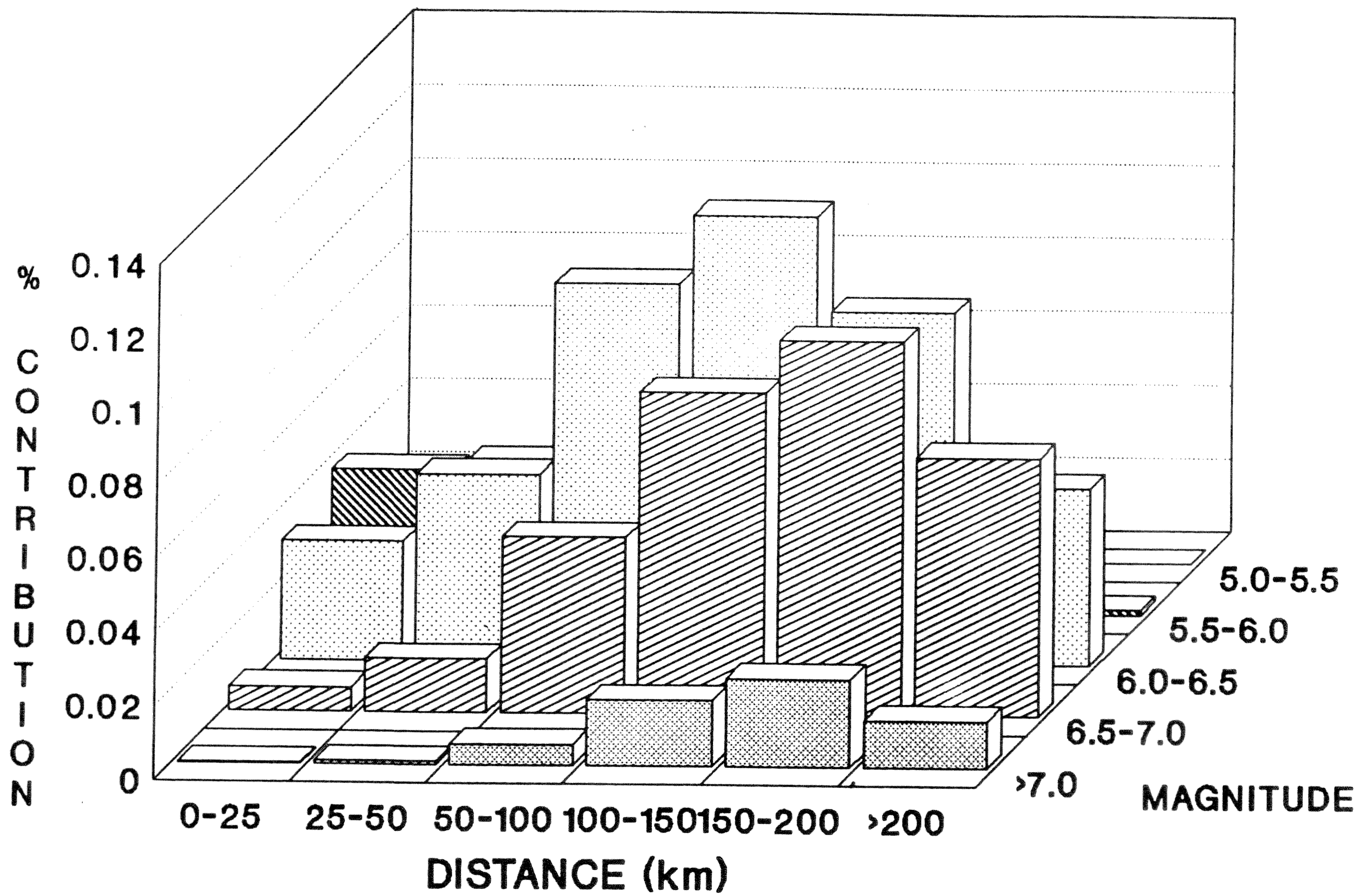
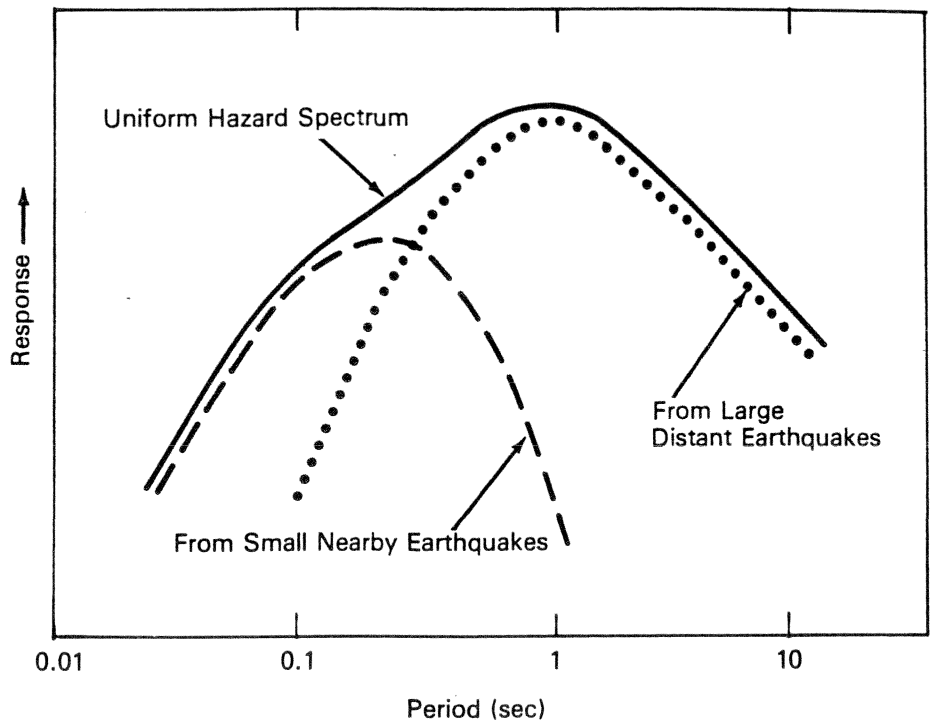


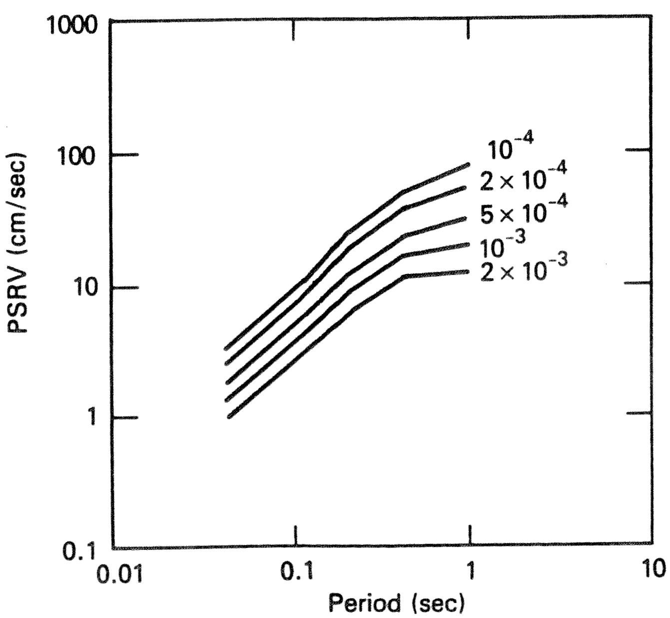
Figure 7a

EPRI MEAN PROBABILISTIC SEISMIC HAZARD
1 HERTZ SPECTRAL ACCELERATION




Uniform Hazard Spectrum
Response
From Large
Distant Earthquakes
From Small Nearby Earthquakes
0.01
0.1
1
10
Period (sec)

(a)


1000
100
10
1
0.1
PSRV (cm/sec)
10^{-4}
2×10^{-4}
5×10^{-4}
10^{-3}
2×10^{-3}
0.01
0.1
1
10
Period (sec)

(b)

Figure 8

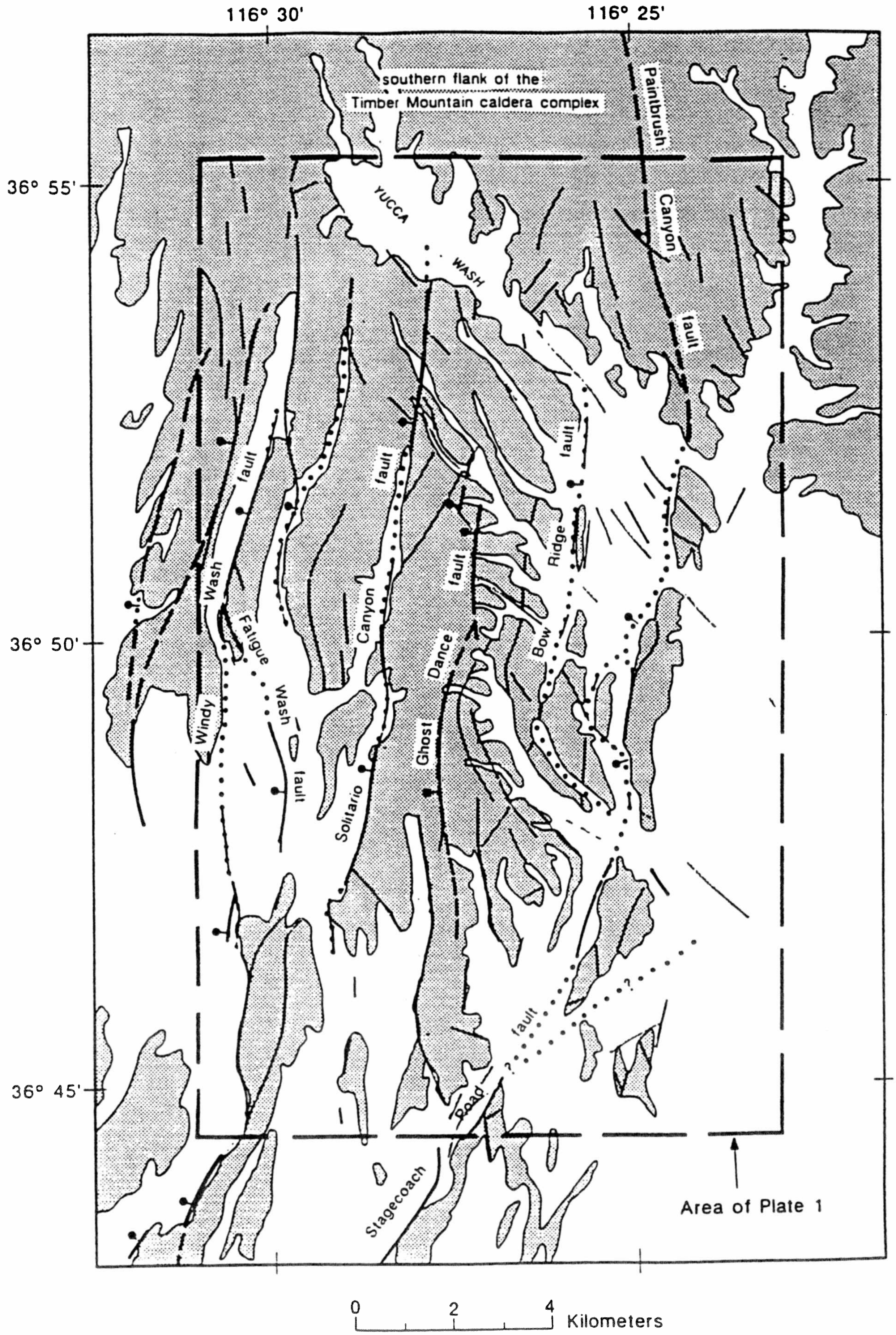

116° 30'
116° 25'
36° 55'
36° 50'
36° 45'
southern flank of the
Timber Mountain caldera complex
Paintbrush
Canyon
fault
YUCCA
WASH
Windy
Wash
fault
Fatigue
Wash
fault
Canyon
fault
Solitario
Ghost
Dance
fault
Bow
Ridge
fault
Stagecoach
Road
fault
Area of Plate 1
0
2
4
Kilometers

Figure 9