EARTHQUAKE ENGINEERING AND STRUCTURAL DYNAMICS
Earthquake Engng Struct. Dyn. 2008; **37**:329–338
Published online 19 October 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/eqe.765

REVIEW

Probabilistic seismic hazard analysis: Early history

Robin K. McGuire*,†

Risk Engineering, Inc., 4155 Darley Av., Suite A, Boulder, CO 80305, U.S.A.

SUMMARY

Probabilistic seismic hazard analysis (PSHA) is the evaluation of annual frequencies of exceedence of ground motion levels (typically designated by peak ground acceleration or by spectral accelerations) at a site. The result of a PSHA is a seismic hazard curve (annual frequency of exceedence *vs* ground motion amplitude) or a uniform hazard spectrum (spectral amplitude *vs* structural period, for a fixed annual frequency of exceedence). Analyses of this type were first conceived in the 1960s and have become the basis for the seismic design of engineered facilities ranging from common buildings designed according to building codes to critical facilities such as nuclear power plants. This *Historical Note* traces the early history of PSHA. Copyright © 2007 John Wiley & Sons, Ltd.

Received 16 August 2007; Accepted 31 August 2007

KEY WORDS: PSHA; seismic; hazard; history

INTRODUCTION

Probabilistic seismic hazard analysis (PSHA) integrates over all possible earthquake ground motions at a site to develop a composite representation of the spectral amplitudes and hazards (annual frequencies of exceedence) at that site. The analysis has a strong basis in earth sciences and earthquake engineering, and allows decisions on seismic design levels for a facility to be made in the context of the earthquake magnitudes, locations, and ground motions (including the effects of local site conditions on amplitudes of strong shaking) that may occur. The use of PSHA is common throughout the world for determining seismic design levels. The collaboration of two researchers in the 1960s resulted in the fundamental concepts of PSHA.

EARLY DEVELOPMENT OF PSHA

The seeds of PSHA were sown in the early 1960s in the form of two efforts that came together in 1966. One effort was the 1964 doctoral dissertation of Allin Cornell at Stanford titled ‘Stochastic

*Correspondence to: Robin K. McGuire, Risk Engineering, Inc., 4155 Darley Av., Suite A, Boulder, CO 80305, U.S.A.
†E-mail: mcguire@riskeng.com

Processes in Civil Engineering,' [1] which studied probability distributions of factors affecting engineering decisions. A key concept in this thesis was 'derived distributions,' in which the probability distribution of a complicated, dependent variable is derived given its relationship with other independent variables whose probability distributions are known or can be assumed. The second effort consisted of studies at the Universidad Nacional Autonoma de Mexico (UNAM) by PhD student Luis Esteva, Prof. Emilio Rosenblueth, and co-workers, who were studying earthquake ground motions, their dependence on magnitude and distance, and the relationship between the frequency of occurrence of earthquakes and the frequency of occurrence of ground motions at a site. An underlying concept in these studies was that the optimal design of buildings for earthquakes could be achieved by accounting for the probabilities of earthquake occurrences, of the associated ground motions, and of the resulting engineering failures.

These efforts fortuitously came together in the summer of 1966, during which Allin Cornell taught his MIT undergraduate probability class at UNAM and interacted with Luis Esteva. Both parties gained from this exchange of ideas. The UNAM group was using Bayesian updating to estimate distributions of earthquake occurrences in regions, and was estimating distributions of ground motion intensity at a site [2]. This group was also studying the dependence of peak ground acceleration, velocity, and displacement (PGA, PGV, and PGD, respectively) on earthquake magnitude and distance. These studies had allowed Luis Esteva [3] to publish the first seismic zone maps that included modified Mercalli intensity (MMI), associated PGA value, and return period. These maps were constructed on a simple basis, combining the recurrence rates of magnitudes in broad zones and the resulting ground motions at sites and equating the return periods of site intensity levels to the recurrence intervals of the causative earthquakes.

The UNAM effort also concentrated on design decisions, estimating probabilities of future structural failure caused by earthquake ground motions, and the costs of those failures to derive optimal design levels. These concepts included the (time-discounted) credits and costs of future benefits and failures, and the idea that not all credits and costs are measurable objectively in financial terms (examples are discredited reputation or loss of human lives and works of art).

Allin Cornell recognized the importance of these concepts and saw that their application for any site required a well-founded ground motion hazard curve (ground motion amplitude *vs* annual frequency of exceedence). This prompted discussions about how to derive what is now termed 'probabilistic seismic hazard' from the relationships among earthquake magnitudes, rate of occurrence of those magnitudes, locations of events, and the resulting ground motions at a site. Cornell saw this as a classic problem in derived distributions, and convinced Esteva and others at UNAM of this approach.

C. Allin Cornell (1938–)

Luis Esteva (1935–)

DOI: 10.1002/eqe

BASIC FORMULATION OF PSHA

When Allin Cornell returned to MIT in the fall of 1966, another fortuitous event occurred. He was asked to participate in a consulting project funded by the government of Turkey to help determine an appropriate earthquake design ground motion for the Alibey Dam located north of Istanbul, Turkey, and affected by earthquakes on the nearby North Anatolian fault. This was a simple case in which the site was influenced by a linear fault whose magnitude distribution could be estimated. It led Cornell to the derivation of a simple equation for probability of exceedence of MMI intensity level i, given an earthquake on a fault of length l:

$$P[I \geqslant i] = l^{-1} C G \exp(-\beta i / c_2), \quad i \geqslant i' \tag{1}$$

where C, G, and c_2 are constants related to the ground motion dependence on magnitude and to the geometry of the fault, β is ln(10) times the Richter b-value for the fault, and i' is a lower bound on intensity. Earthquake occurrences were taken into account by assuming that they occur as a Poisson process and by recognizing that ground motion intensities $I \geqslant i$ were a Poisson process under random selection. For this Poisson process, the distribution of the largest intensity $i_{\max}$ in time period t could be calculated as the probability that exactly zero events with $I \geqslant i_{\max}$ occurred in time 0 to t:

$$F_{I_{\max}} = \exp[-v C G \exp(-\beta i / c_2)], \quad i \geqslant i' \tag{2}$$

where v is the rate of earthquake occurrence on the fault. Equations (1) and (2) allowed the distribution of earthquake ground motions to be derived from the distribution of earthquake occurrences, the attenuation of motion, and the geometry of the fault with respect to the site. In the 1960s, the application of computers to solve engineering problems was still in its infancy, and standard practice was to find closed-form solutions to integral equations. In this respect the most challenging part of applying equations (1) and (2) was to derive constant G, which relates the distribution of source-to-site distance to the geometry of the fault with respect to the site. Specifically, G is the expected value of distance when raised to the power ρ (see Cornell and Vanmarcke [4]), where ρ is equal to $-\beta b_3 / b_2$ and b_3 and b_2 are given in Equation (4). (As Cornell and Vanmarcke [4] point out, this is not the same as raising the expected distance to the power ρ.) This led Cornell to include a graphical solution in his 1968 paper to allow the evaluation of constant G for some simple cases. Cornell and Vanmarcke [4] also published graphical evaluations of constant G for point, fault, and area sources.

The formulation in Equation (2) has became known as PSHA. Several parts of the formulation benefited from the earlier collaboration with UNAM colleagues. Specifically, the earthquake magnitude distribution was assumed to be exponential, which came from early work on earthquake magnitude statistics by Richter [5]. Also, MMI and instrumental measures of ground motion (PGA, PGV, and PGD) were assumed to depend on magnitude M and distance R in the following ways:

$$\text{MMI} = c_1 + c_2 M + c_3 \ln(R) \tag{3}$$

$$Y = b_1 \exp(b_2 M) R^{-b_3} \tag{4}$$

where Y is PGA, PGV, or PGD and where the b's and c's are constants. Equation (3) is, of course, just a logarithmic form of Equation (4), but the difference was important. The distribution of maximum MMI at a site in time t was determined to be a double exponential distribution (Equation

DOI: 10.1002/eqe

(2)), which is the form of a Type I asymptotic extreme value distribution (or Gumbel distribution). The distribution of maximum ground motion Y at a site in time t was determined to be the form of a Type II asymptotic extreme value distribution:

$$F_{Y_{\max}} = \exp[-\nu C G y^{\beta/b_2}], \quad y \geqslant y' \tag{5}$$

Again the background of Allin Cornell in developing derived distributions was instrumental in these derivations.

An important distinction was made by Cornell that these extreme value distributions were developed based on the probability distributions of the underlying variables, without reliance on arguments related to the asymptotic nature of extreme values, which was more common practice. In fact, such a reliance had been used in earlier publications (e.g. Milne and Davenport [6]) to represent earthquake ground motions, by extending 'observations' (which actually were *estimates* of ground motions from historical earthquake catalogs) to low probabilities. Extreme value distributions had been used to analyze wind and flood levels, where for example 30 years of observations of annual maxima could be used to predict the 100-year wind velocity or flood elevation. For earthquake ground motions, very few sites had 30 years of actual recorded data, and the derivation by Cornell was an important step in recognizing that earthquake ground motions conform to distributions that are similar to other natural phenomena.

The above formulation of seismic hazard analysis was published by Allin Cornell in his 1968 paper, 'Engineering Seismic Risk Analysis.' The solution for a linear fault was extended to source areas, and Cornell showed how to model an arbitrarily complex region as a set of faults or a set of annular sources that contribute to hazard. He also showed that, for multiple earthquake sources and low probabilities of exceedence, the total hazard (probability of exceedence) was the sum of hazards from the contributing sources.

Figure 1 is a reproduction of a figure from Cornell [7] showing a point source, two line sources, and a large area source, and showing the critical geometries of these sources for seismic hazard calculations. Seismic hazard analyses conducted today use similar geometries to determine the source-to-site distances of earthquakes in large zones, and of earthquakes occurring on recognized faults.

Concurrent with this mathematical derivation of PSHA, Luis Esteva was developing and publishing similar concepts. The emphasis of Esteva, however, was on the overall engineering decision process for earthquakes, not just on the extreme ground motion distribution. These concepts were published by Esteva in [8] and in his PhD thesis at UNAM [9]. They included the Bayesian updating of seismic activity rates, ground motion equations for PGA, PGV, and PGD, the derivation of design spectra from these values, and seismic hazard curves for spectral ordinates that were scaled from PGA, PGV, and PGD. This was the first publication of hazard curves for spectral ordinates.

Two important aspects of Esteva's work were particularly notable. First, ground motions were recognized to have significant 'aleatory uncertainty' associated with their occurrences. This uncertainty was quantified, and introduction of the term 'aleatory uncertainty' was 20 years ahead of its widely accepted use. Second, decision rules for seismic design were developed using structural failure as the decision variable. These decision rules accounted for discounting of future costs and benefits by a monetary discount rate, and took into account whether structures were to be rebuilt following earthquake-induced failure or were to be replaced. The decision rules also accounted for different possible levels of damage during earthquake motions, and noted that design decisions depend only on the expected rate of structural failure, not on its uncertainty.

DOI: 10.1002/eqe

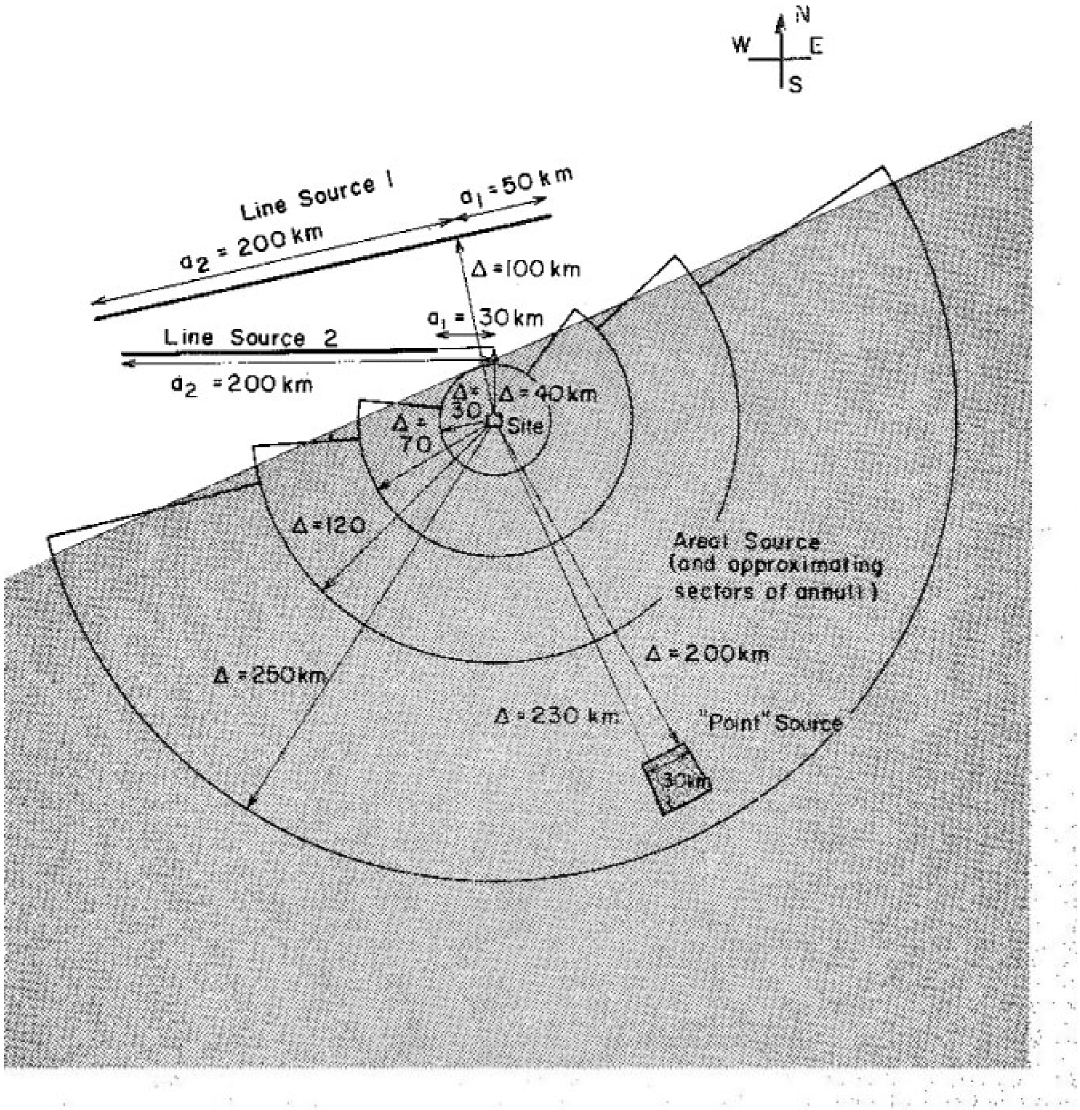

Figure 1. Geometry of point source, two line sources, and a large area source (from Cornell [7]).

EXTENSIONS OF ORIGINAL PSHA

The basic formulation of PSHA was generalized in the 1970s using the 'total probability theorem':

$$P(Y > y) \simeq \sum v_i \iint P[Y > y|M, R] f_{M,R}(m, r)\,\mathrm{d}m\,\mathrm{d}r \tag{6}$$

DOI: 10.1002/eqe

where Y and y are earthquake ground motions, M and m are magnitudes, R and r are distance, v_i is the rate of earthquakes for source i, and the summation is over all sources.

Several important aspects of this generalization came from further applications. First, Cornell recognized that earthquake magnitude distributions are not unlimited, as was assumed in the original formulation. As a result, the curve representing probability of exceedence *vs* ground motion amplitude is not a straight line on a log–log plot, as was illustrated in Cornell [7] and Esteva [9]. This extension was published in Cornell and Vanmarcke [4]. The latter publication was also ahead of its time in recognizing that '…only the closer, smaller, more frequent earthquakes are significant in contributing to the risk of large ground accelerations at a site. For ground velocities, with their typically smaller attenuation constants,…distant sources can be more important.' It took more than two decades for this conclusion to be quantified, when the deaggregation of seismic hazard was developed as a tool for understanding the contributions to seismic hazard [10].

Second, the aleatory uncertainty in ground motion was recognized to be an important contributor to probabilities of exceedence. This was recognized by Esteva [9, 11, 12], who published equations to calculate the rate of exceedence of Y (see Equation (6)) by first calculating the rate of exceedence using the ground motion equation without aleatory uncertainty and then integrating over the ground motion uncertainty. Cornell [13] derived closed-form solutions to the seismic hazard integral (Equation (6)) when the earthquake magnitude distribution is exponential and the mean ground motion equation is of a simple form (see Equations (3) and (4)). This derivation was for the case where the magnitude distribution is bounded by a maximum magnitude and where the ground motion uncertainty is included in the hazard calculation.

The inclusion of ground motion uncertainty in PSHA was important but unfortunately was published in more obscure references than Cornell's original paper. Cornell [13] documented his results in the proceedings of a conference held in Swansea, U.K., in July 1970. Esteva's formulation was published in his PhD thesis [9] at UNAM, in the proceedings of an MIT symposium [11], and in a paper at the 4th World Conference on Earthquake Engineering [12] that predominately addressed seismicity distributions.

The fact that aleatory uncertainty in ground motion was not included in the original Cornell [7] formulation of PSHA has been the subject of recent discussion. Bommer and Abrahamson [14] for example ascribe increased estimates of seismic hazard in studies in the 1990s to the inclusion of this uncertainty in the calculations, compared with earlier studies that ignored ground motion uncertainty. These authors attribute the exclusion of aleatory uncertainty in earlier studies to its exclusion in the original Cornell formulation of seismic hazard and to its subsequent exclusion in US Geological Survey (USGS) seismic hazard maps (see below).

Following the lead of Esteva [9] in using seismic hazard results to make decisions on optimal structural design, Cornell [13] included results on computing structural damage in the form of linear and non-linear (elasto-plastic) response of simple structural systems to earthquake motions. Cornell's results were in the form of the mean and variance of damage. Cornell and Esteva continued to interact in 1969 when Esteva spent a sabbatical semester at MIT, and in 1971, when Cornell spent a sabbatical semester at UC Berkeley and Esteva visited.

The development of quantitative ground motion equations paralleled the advance of PSHA. The first ground motion equations for peak parameters (PGA and PGV) based on least-squares regression analysis of data were published by Esteva and Rosenblueth [15], using California data. This paper included an analysis of uncertainties in the data, thereby quantifying the aleatory uncertainty. An important part of this development, also documented by Esteva [9], was that ground motions depend on the geologic conditions at the recording site. Virtually all quantitative ground

DOI: 10.1002/eqe

motion equations based on empirical data have quantified the scatter in observations about the predictive equations.

SEISMIC HAZARD MAPS

A logical application of PSHA is to determine seismic design levels for building codes that reflect consistent annual probabilities of exceedence (or equivalently the return period of a specified level of ground motion). As noted above, the first published seismic zone map that included levels of ground motion and associated return periods was Esteva [3]. This was based on estimates of recurrence intervals of large magnitudes in seismic zones, attenuating the ground motion to estimate intensity in various zones, and equating the ground motion return periods to the earthquake recurrence intervals.

Cornell [13] included a 'preliminary' seismic hazard map for southern California based on his formulation of PSHA, to demonstrate the result of moving the site sequentially throughout a region, calculating PSHA at each site, and summarizing the results as a contour map of PGA for a fixed return period. Esteva [16] calculated and presented the first PGA and PGV maps for all of Mexico for return periods of 50, 100, and 500 years, thereby establishing the first national seismic hazard maps in 1970. An example is shown in Figure 2. The first national seismic hazard maps in

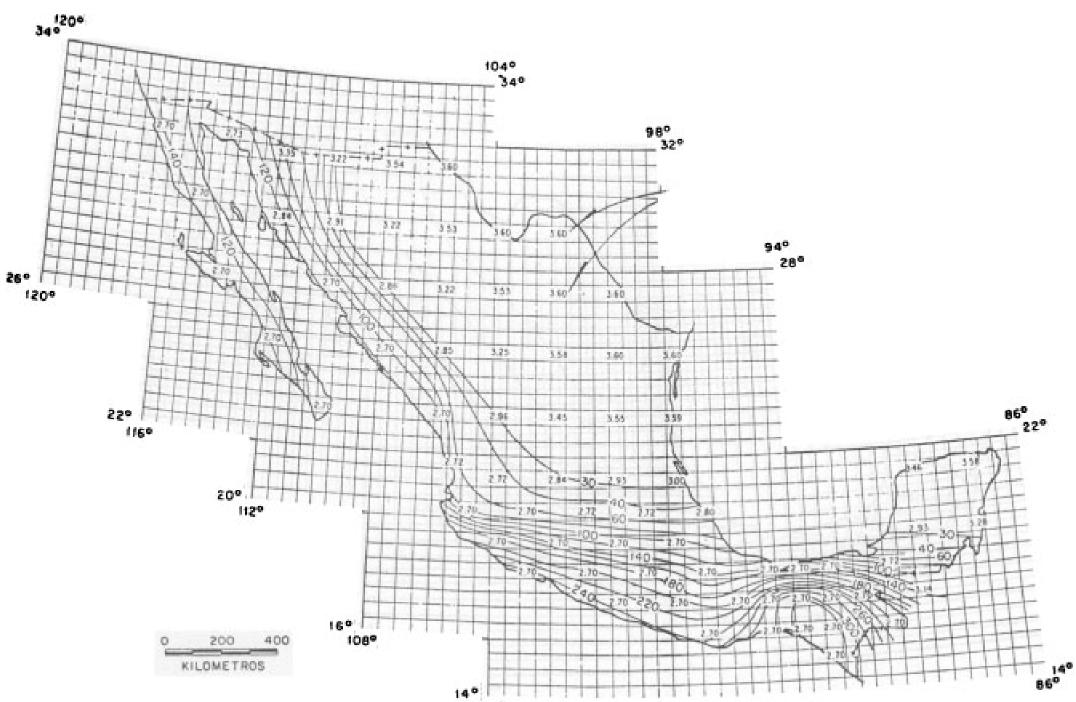

Figure 2. National seismic hazard map for Mexico showing PGA with a 500-year return period, published by Esteva [16] in 1970.

DOI: 10.1002/eqe

the U.S. were not published until 6 years later [17], in 1976. Prior to that, U.S. seismic codes were based on maps that showed some variant of maximum expected ground motion. The 1976 USGS hazard maps (as well as updated maps published in 1982) were based on a formulation that excluded aleatory uncertainty in ground motion estimates, thereby contributing to the delay in incorporating this uncertainty into standard PSHA calculations. It was not until 1990 that the USGS published hazard maps that included aleatory uncertainty in ground motion.

ADDITIONAL DEVELOPMENTS

Some additional developments were made by other researchers to the basic formulation of PSHA as it is used today. First, earthquakes were recognized to rupture a finite segment of the causative fault, thus becoming a source of energy with finite dimensions rather than a point source as assumed by Cornell [7]. Der Kiureghian and Ang [18] described this effect and recommended that the distance from the site to the closest point of rupture was the best distance measure to use for ground motion estimation. This effect is important only for large magnitude earthquakes, of course, but these are often the events that dominate the seismic hazard in plate margin areas. The result was that seismic hazard results were more accurate at near-fault sites affected by large magnitude earthquakes, and that seismic hazard maps for regions with major faults were more realistic.

A development that had a large impact on seismic hazard calculations was the recognition that ground motion equations and seismic hazard curves could be developed directly on spectral response [19], leading to the concept of a uniform hazard spectrum. Prior to that, seismic hazard curves were developed for PGA, PGV, and perhaps PGD, and the spectrum was constructed by amplifying these peak motions measures (e.g. Newmark and Hall [20]).

Finally, the proliferation of digital computers in the 1970s meant that PSHA calculations (particularly of the constants in Equations (1) and (2)) could be made with software, thus avoiding the necessity of achieving closed-form solutions to the seismic hazard integral (Equation (6)). Using solution techniques based on numerical integration of the seismic hazard integral meant that arbitrarily complex ground motion equations, rather than just the simple forms represented in Equations (3) and (4), could be used to calculate seismic hazard, with little penalty in terms of calculational effort. Among the software programs developed in the 1970s, the EQRISK and FRISK programs [21, 22] were made available as public domain programs. These programs calculated seismic hazard for area sources and faults, respectively, they were applied (and are still being applied) to PSHA problems around the world, and they contributed to the wide acceptance of PSHA as a tool for earthquake engineering decision-making.

Since the 1970s, many advances in PSHA have allowed this technology to gain acceptance as the preferred method of specifying seismic design levels for engineered facilities. Quantification of epistemic uncertainty allows multiple competing hypotheses on models and parameters to be incorporated into the analysis. Deaggregation of seismic hazard [10] promotes understanding of the major contributors to hazard and allows multiple design spectra (not just the uniform hazard spectrum) to be used in design. This means that, for the common case of a site where high-frequency hazard is dominated by frequent local earthquakes and low-frequency hazard is dominated by large, distant earthquakes, each type of ground motion can be modeled separately, to accurately determine the response of the soil column or of the structure. The alternative of using a broad-banded spectrum would represent a ground motion that is not realizable and might over-drive

DOI: 10.1002/eqe

the soil column or the structure, resulting in incorrect estimation of non-linear soil behavior or structural response. Understanding of what motions are critical to engineering failures allows risk-consistent spectra (which have, as a goal, constant probability of failure across all structural frequencies) to be derived from the basic hazard curves. All of these advances depend, at their core, on the early formulation of seismic hazard analysis by Cornell and Esteva.

CLOSURE

The experience of Allin Cornell in (among other things) applying the concept of derived distributions in probability analysis combined fortuitously with the interest of Luis Esteva at UNAM in solving earthquake design problems. Cornell applied the mathematical rigor of probabilistic analysis, and Esteva provided the underlying distributions of earthquake magnitudes, locations, and ground motion attenuation. The result was a mathematical formulation that has become the basis for choosing seismic design levels around the world. This demonstrates the value of cross-fertilization in engineering fields, and illustrates that even minor interactions among colleagues (a summer sabbatical to teach a summer course) can pay major dividends.

ACKNOWLEDGEMENTS

This effort to document the early history of PSHA has benefited from discussions with, and reviews by, Allin Cornell and Luis Esteva, whose memories of interactions some 40 years ago are amazingly consistent. Early drafts of this manuscript were improved as a result of reviews by Julian Bommer and Tom Hanks. Finally, I thank Anil Chopra for suggesting this topic and for his patience in reviewing its progress.

REFERENCES

1. Cornell CA. Stochastic processes in civil engineering. *Ph.D. Thesis*, Department of Civil Engineering, Stanford University, 1964.
2. Rosenblueth E, Esteva L. On seismicity. *Seminar on Applications of Statistics in Structural Mechanics*, *Civil and Mechanical Engineering*, University of Pennsylvania, May 1966.
3. Esteva L. Regionalizacion sismica de la Republica Mexicana. *Revista Sociedad Mexicana do Ingenieria Sismica* 1963; **1**(1):31–35.
4. Cornell CA, Vanmarcke EH. The major influences on seismic risk. *Proceedings of the 4th World Conference on Earthquake Engineering*, Santiago, Chile, 1969.
5. Richter CF. *Elementary Seismology*. W.H. Freeman and Co.: San Francisco, 1958.
6. Milne WG, Davenport AG. Statistical parameters applied to seismic regionalization. *Third World Conference on Earthquake Engineering*, Auckland, New Zealand, 1965.
7. Cornell CA. Engineering seismic risk analysis. *Bulletin of the Seismological Society of America* 1968; **58**(5): 1583–1606.
8. Esteva L. Criterios para la construcción de espectros para diseño sísmico. *Proceedings of XII Jornadas Sudamericanas de Ingeniería Estructural y III Simposio Panamericano de Estructuras*, Caracas, 1967. Published later in *Boletín del Instituto de Materiales y Modelos Estructurales*, Universidad Central de Venezuela, No. 19, 1967.
9. Esteva L. Bases para la formulacion de decisiones de diseño sismico. *Ph.D. Thesis and Report 182*, Universidad Autonoma Nacional de México, August 1968.
10. McGuire RK. Probabilistic seismic hazard analysis and design earthquakes: closing the loop. *Bulletin of the Seismological Society of America* 1995; **85**(5):1275–1284.
11. Esteva L. Seismic risk and seismic design decisions. *Proceedings of MIT Symposium on Seismic Design of Nuclear Power Plants*, Cambridge, MA, 1969.

DOI: 10.1002/eqe

12. Esteva L. Seismicity prediction: a Bayesian approach. *Proceedings of the 4th World Conference on Earthquake Engineering*, Santiago, Chile, 1969.
13. Cornell CA. Probabilistic analysis of damage to structures under seismic loads. In *Dynamic Waves in Civil Engineering*, Chapter 27, Howells DA, Haigh IP, Taylor C (eds). Wiley-Interscience: New York, 1970; *Proceedings of Conference on Earthquake and Civil Engineering Dynamics*, University College at Swansea, 7–9 July 1970.
14. Bommer JJ, Abrahamson NA. Why do modern probabilistic seismic-hazard analyses often lead to increased hazard estimates? *Bulletin of the Seismological Society of America* 2006; **96**(6):1967–1977.
15. Esteva L, Rosenblueth E. *Espectros de temblores a distancias moderadas y grandes*. Sociedad Mexicana de Ingenieria Sismica, Publ. No. 1, vol. II, 1–18 March 1964.
16. Esteva L. Regionalizacion sismica de Mexico para fines de ingenieria. *Report 246*, Universidad Nacional Autonoma de Mexico, April 1970.
17. Algermissen ST, Perkins DM. A probabilistic estimate of maximum acceleration in rock in the contiguous United States. *U.S. Geological Survey Open-File Report*, 1976; 76–416.
18. Der Kiureghian A, Ang AH. A fault-rupture model for seismic risk analysis. *Bulletin of the Seismological Society of America* 1977; **67**(4):1173–1194.
19. McGuire RK. Seismic structural response risk analysis incorporating peak response regressions on earthquake magnitude and distance. *Publ. 399*, M.I.T. Department of Civil Engineering, 1974; 371.
20. Newmark NM, Hall WJ. Earthquake spectra and design. *Monograph MNO-3*, Earthquake Engineering Research Institute, 1982; 103.
21. McGuire RK. Fortran program for seismic risk analysis. *U.S. Geological Survey Open-File Report* 76–67, 1976.
22. McGuire RK. FRISK: computer program for seismic risk analysis using faults as earthquake sources. *U.S. Geological Survey Open-File Report*, 1978; 78–1007.

DOI: 10.1002/eqe