MISCELLANEOUS PAPER S-73-1

STATE-OF-THE-ART FOR ASSESSING EARTHQUAKE HAZARDS IN THE UNITED STATES

Report 18

ERRORS IN PROBABILISTIC SEISMIC HAZARD ANALYSIS

by

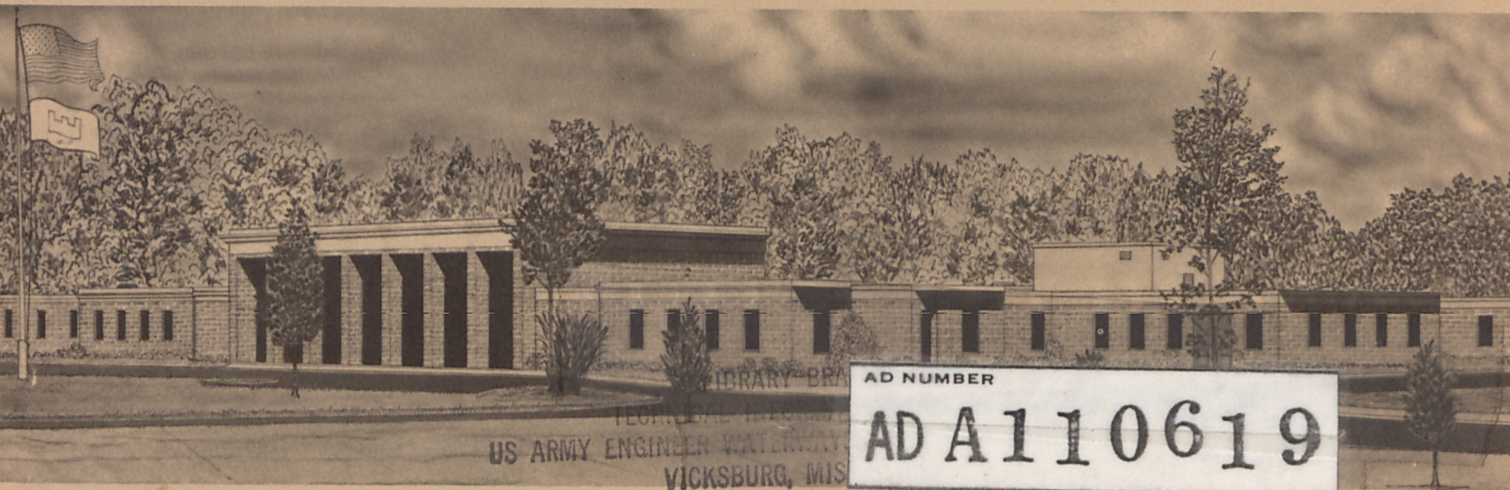
Daniele Veneziano

Department of Civil Engineering
Massachusetts Institute of Technology
Cambridge, Mass. 02139

January 1982

Report 18 of a Series

Approved For Public Release; Distribution Unlimited



Prepared for Office, Chief of Engineers, U. S. Army
Washington, D. C. 20314

Under Contract No. DACW39-80-M-2381

Monitored by Geotechnical Laboratory
U. S. Army Engineer Waterways Experiment Station
P. O. Box 631, Vicksburg, Miss. 39180

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER Miscellaneous Paper S-73-1	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) STATE-OF-THE-ART FOR ASSESSING EARTHQUAKE HAZARDS IN THE UNITED STATES; Report 18, ERRORS IN PROBABILISTIC SEISMIC HAZARD ANALYSIS	5. TYPE OF REPORT & PERIOD COVERED Report 18 of a series
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Daniele Veneziano	8. CONTRACT OR GRANT NUMBER(s) Contract No. DACW39-80-M-2381
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Civil Engineering Massachusetts Institute of Technology Cambridge, Mass. 02139	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office, Chief of Engineers, U. S. Army Washington, D. C. 20314	12. REPORT DATE January 1982
	13. NUMBER OF PAGES 132
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) U. S. Army Engineer Waterways Experiment Station Geotechnical Laboratory P. O. Box 631, Vicksburg, Miss. 39180	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Available from National Technical Information Service, 5285 Port Royal Road, Springfield, Va. 22151.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Earthquake engineering
Earthquake hazards
Earthquakes
Seismic risks

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

After a period of fast development, procedures for the probabilistic quantification of seismic hazard have reached a state of relative maturity. Standard methods usually differ only by secondary characteristics. In spite of this operational stability, estimates of seismic hazard that are produced by different algorithms depend on and are often very sensitive to the form of the mathematical model and the choice of some of the parameters. It is important

(Continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT (Continued).

to recognize this sensitivity in order to avoid placing excessive confidence in numerical estimates of hazard, especially in the case of rare events. Errors may originate from conscious simplification of the mathematical representation (e.g., from the use of a Poisson model of earthquake occurrences as opposed to a physically more attractive point process with spatial and temporal memory), from limited statistical information, or from faulty judgment (expert errors). The purpose of this report is to review different types and sources of errors and to assess their impact on calculated seismic hazard. In the available literature, no generally accepted procedure exists to systematically deal with uncertainty on seismic hazard and, most frequently in practice, elusive arguments of conservatism are used to avoid direct confrontation with the problem.

Scientific studies are sparse and fragmentary; they typically focus on narrow aspects of the problem, use different statistical techniques, and report results in different formats. Accordingly, an effort is made here to coordinate these results and provide a unified picture.

A first and fundamental problem, which in the past has received only minimal attention, is to give a definition of objective or "true" hazard against which various estimates can be evaluated. One such definition is proposed here, based on the notion of hazard as relative frequency in time. Three classes of errors are then introduced and separately analyzed: 1. Errors due to uncertainty of model parameters, such as the geometrical configuration of the earthquake sources, the parameters of the magnitude-frequency relationship (including the upper bound), and those of the attenuation function. A recently proposed analytical procedure for the analysis of such errors is reviewed and results from sensitivity studies are presented. 2. Errors due to the use of a wrong form of model. These errors result in biases (hence, their statistical magnitude does not depend on the amount of information available) and may be very large, especially for rare events. 3. Finally, errors of interpretation and modeling by seismologist experts. Case studies are reviewed and mathematical models for errors of this last type are proposed.

PREFACE

This report was prepared by Dr. Daniele Veneziano, Department of Civil Engineering, Massachusetts Institute of Technology, under Contract No. DACW39-80-M-2381. It is part of ongoing work at the U. S. Army Engineer Waterways Experiment Station (WES) in Civil Works Investigation, "Seismic Effects of Reservoir Loading and Fluid Injection," sponsored by the Office, Chief of Engineers, U. S. Army.

Preparation of this report was under the direction of Dr. E. L. Krinitzsky, Engineering Geology and Rock Mechanics Division (EG&RMD), Geotechnical Laboratory (GL). General direction was by Dr. D. C. Banks, Chief, EG&RMD, and Dr. W. F. Marcuson, Chief, GL.

COL Nelson P. Conover, CE, and COL Tilford C. Creel, CE, were Commanders and Directors of WES during the period of this study. Mr. F. R. Brown was Technical Director.

CONTENTS

1. INTRODUCTION

For centuries, engineering design has been guided by the continuity of successful practice and professional judgement. A first element of objective rationality has been provided by the principles of mechanics, the application of which has become easier through the use of fast computing devices. A second movement toward rationalization, based on the formal treatment of quantifying uncertainty, is now under way. By itself, none of the three elements--mechanics, computers, and quantitative analysis of uncertainty--can produce intelligent engineering decisions, but their combination seems to be both necessary and sufficient. Benefits from probabilistic treatment of uncertainty are especially large in problem areas, such as earthquake engineering, where solutions by traditional means require a great deal of experienced judgment and intuition.

1.1 The Problem of Accuracy in Seismic Hazard Analysis

Wherever a mathematical model, deterministic or probabilistic, is used to represent a complicated physical phenomenon, the problem of determining its accuracy arises. It is clear what accuracy means in the case of deterministic models (one should just compare deterministic predictions with actual observations); this is not so for probabilistic models. For example, there are statisticians who consider probability in seismic hazard statements as quantified personal uncertainty about future earthquake events. From this premise they conclude that if an individual faithfully quantifies his own state of uncertainty, then calculated hazard is necessarily exact, at least for that individual. This argument is built around the so-called subjective, degree-of-belief (d.o.b.) interpretation of probability. This interpretation will be discussed at length in

Sec. 2 and conclusions will be drawn that it is inappropriate or at least insufficient in the context of seismic hazard. One should notice, however, that the issue of accuracy is legitimate within the d.o.b. interpretation for two reasons: (1) because it is typical to use simple mathematical approximations of complicated states of uncertainty and (2) because there is ample evidence that humans make mistakes in quantifying their own state of uncertainty.

It will be proposed in Sec. 2.1 that probability be given a frequentist, time-average interpretation in the quantification of seismic hazard; e.g., that the probability that the peak ground acceleration at a given site during the next 50 years will exceed 0.2 g be interpreted as the relative frequency with which the event occurs over periods of 50 years. Notice that in this interpretation, the <u>next</u> time interval is replaced with the <u>generic</u> time interval. The same interpretation has two major advantages over the d.o.b. interpretation: (1) it is <u>objective</u> and (2) it applies if the physical earthquake process is regarded as <u>deterministic</u>. It also makes clear what is meant by an accurate model: an accurate model is one which produces probabilities of seismic events close to the actual relative frequencies in time.

Accuracy of a procedure for the estimation of seismic hazard is not by itself a sufficient measure of engineering value. ("Seismic hazard" is used here to denote a suitable probabilistic characterization of future ground motions at the site, often simply in terms of the mean rate at which peak motion amplitudes at the site exceed given values.) What is of greater interest for engineering purposes is "seismic risk," i.e., the probabilistic characterization of consequences from exposure of engineering facilities to seismic hazard. For example, it may be of interest to know the yearly probability of overtopping of a dam due to earthquake events. If the dam is completely unsafe with respect to very severe ground motions, then inaccuracy of seismic hazard in that range is inconsequential.

A second reason why accuracy is not sufficient is that simpler procedures are often preferred to more precise but complicated methods. Decision-theoretic approaches to engineering model selection have been proposed (e.g., Grigoriu, Veneziano, and Cornell, 1979) and they account for model errors, their consequences, and the cost of developing and using the model. It would be interesting to examine the accuracy of seismic hazard methods from this wider decision-theoretic perspective, but studies on this subject are almost nonexistent. For some types of engineering systems, information is available on the dependence of risk on seismic hazard and also on the sensitivity of risk to errors on seismic hazard. This information is reviewed in Sec. 3, as it is useful to evaluate the practical relevance of hazard estimation errors.

1.2 Seismic Hazard Procedures and Sources of Errors

Seismic hazard procedures or models, M, are sequences of analytical or numerical operations that produce estimates of earthquake hazard from suitable seismological information. Denote by z the available information and by $\hat{H}$ the hazard estimate from z and from a given procedure M. The form of $\hat{H}$ varies depending on the purpose of seismic hazard analysis: for example, $\hat{H}$ might consist of a random sequence of earthquake times and intensities at the site. In other cases $\hat{H}$ is simply a "hazard function," e.g., a function $\hat{\Lambda}(a)$ which, for each given peak ground acceleration a, gives the mean rate at which a is exceeded at the site. In other cases still, $\hat{H}$ may characterize seismic threat at several or even at all points of a geographical region. Estimates of this last type are needed to characterize the vulnerability of spatially extended systems (networks, lifelines, groups of facilities), to establish insurance risk, or to study the socio-economic impact of earthquakes.

Although procedures of different types have been proposed, a certain degree of standardization has been achieved in recent years in favor of models, the elements of which are: a set of "seismic sources" near the site of interest, a random process model of time, location, and size of earthquakes from each source, and an attenuation function that relates ground motion amplitude at the site to earthquake size and distance from the energy source (Cornell, 1968; Der Kiureghian and Ang, 1977). In the present study, only procedures of this standard type are considered and $\hat{H}$ is assumed to be a hazard function, e.g., of the type previously denoted by $\hat{\Lambda}(a)$.

In order to understand in which ways $\hat{H}$ may be in error, it is useful to distinguish among the following:

M_T = <u>true</u> (actual) <u>but unknown model</u>. This is a model which produces probabilities of hazard-related events equal to the associated relative frequencies in time. The associated true hazard is denoted by H_T.

$\mathfrak{M}$ = class of (simple) seismic hazard models to which, for the purpose of obtaining $\hat{H}$, the true model M_T is assumed to belong. For example, $\mathfrak{M} = \{M(\lambda, \underline{\theta})\}$ = class of Poisson occurrence models with constant mean activity rate λ in space and time and independent identically distributed (I.I.D.) earthquake sizes $Y_1, Y_2, \ldots$. The Y_i have a given probability distribution function $F_Y(y, \underline{\theta})$ with $\underline{\theta}$ an unknown vector of parameters. Hence, in this case $\mathfrak{M}$ is a family of Poisson models parameterized by λ and $\underline{\theta}$.

$\hat{M}$ = estimate of M_T obtained from given information, under the assumption that M_T belongs to $\mathfrak{M}$. For example, $\hat{M} = M(\hat{\lambda}, \hat{\underline{\theta}})$ in which $\hat{\lambda}$ and $\hat{\underline{\theta}}$ are point estimates of λ and $\underline{\theta}$. The hazard estimate $\hat{H}$ corresponds to the model $\hat{M}$.

The empirical validation of $\hat{H}$ or the evaluation of estimation errors by comparing predictions from $\hat{H}$ with observations from H_T is impractical because seismic events of engineering interest are too rare. More realistic means of accuracy evaluation are as follows:

<u>If indeed</u> M_T <u>belongs to</u> $\mathfrak{M}$, then one can study properties of the random estimation error $\hat{H}$-H_T by means of Bayesian analysis or by methods of classical statistics. In the former case, the properties of the error depend on the distribution of M_T in $\mathfrak{M}$; in the second case, the same properties depend on which of the elements of $\mathfrak{M}$ is the true model. Of course, in both cases the error distribution depends on the type and amount of information available.

A simpler but less precise method for this type of error analysis consists of calculating the sensitivity of $\hat{H}$ to the choice of $\mathfrak{M}$; for example, sensitivity with respect to the form of the attenuation and size-frequency functions.

<u>If</u>, contrary to the assumption in the calculation of $\hat{H}$, M_T <u>does not belong to</u> $\mathfrak{M}$ (e.g., if the actual sequence of times, locations, and earthquake sizes has memory, and Poisson processes with independent sizes are the only models in $\mathfrak{M}$), then severe bias in hazard estimation may result. If bias is small over a range of erroneous modeling assumptions, then the hazard procedure is said to be <u>robust</u> over that range. Robustness is an important property because in practice, M_T is always known not to belong to $\mathfrak{M}$.

Hazard estimation errors when M_T is an uncertain model in $\mathfrak{M}$ are considered in Sec. 4; those when M_T does not belong to $\mathfrak{M}$ are studied in Sec. 5.

Reliance on statistical analysis only with historical data treated as a random sample may lead to exceedingly large uncertainty on some critical parameters (typically source geometry and upper bound earthquake size) and hence on calculated hazard. One way to reduce this variability is to use parameter and uncertainty estimates by seismologists based on statistical as well as geophysical and geological information. Expert opinion is especially valuable in regions of low seismicity for which historical samples are small. Whenever statistical analysis is possible, experts should be provided with results from it. Thus, use of expert opinion is not to be viewed as an alternative to statistical analysis, but rather as a complementing technique.

Expert opinion is not itself infallible, as demonstrated by the variability of estimates from different experts. It is therefore important to understand any model errors from using this approach and to devise methods to reduce these errors by combining estimates from different experts. This is the subject of Sec. 6.

Unfortunately, errors in seismic hazard analyses have not been studied systematically; contributions are sparse and fragmentary, each focussing on a narrow aspect of the problem. A variety of techniques is used to express results in different forms. In spite of the effort to give unity to the subject, voids and inconsistencies still remain. There is no accepted procedure to rationally and systematically deal with uncertainty on seismic hazard and, frequently in practice, elusive arguments of "conservatism" are used to avoid direct confrontation with the problem.

Additional work is needed on each topic in this report and a special effort should be made to unify the language and coordinate the results. Practical guidelines on how to deal with non-frequentist uncertainty would be of much value. One possibility, which is advocated by many, is to regulate safety in terms of events, the probability of which is high enough to be accurately estimated by available procedures. For example, one might base seismic design on the earthquake intensity with a return period of 10^3 years, as opposed to a return period of 10^6 years. Of course, higher protection against damage should be provided under the less intense motion.

It is hoped that considerations in this study, especially those on the interpretation of seismic hazard, the classification of estimation errors, and the treatment of expert opinion, will prove useful to future endeavors.

2. FREQUENTIST SEISMIC HAZARD AND UNCERTAINTY DUE TO ERRORS

The argument that uncertainty expressed by seismic hazard is the product of ignorance and that seismic hazard probability therefore should be regarded as a measure of strength of belief is unsatisfactory in many respects because degrees of belief (1) do not necessarily satisfy the axioms of probability theory, (2) are vaguely defined and variable from individual to individual, and (3) are difficult to quantify. Due to (1), the mathematical theory of probability does not strictly apply. Difficulty (2) makes it questionable whether important engineering decisions should be based on essentially subjective hazard assessments, and (3) indicates that implementation faces problems of inference.

In Sec. 2.1, rephrasing of seismic hazard statements is proposed, thus making it possible to interpret probability as relative frequency in time. Such an interpretation overcomes the above difficulties and, contrary to the ensemble-frequentist definition, does not refer to a statistical population, which in the present case obviously does not exist.

Not all uncertainties that contribute to seismic hazard are of the same type or admit the time-average frequentist interpretation; for example, uncertainty about seismicity parameters does not. In this case, how should one combine probabilities with different meanings and what interpretation should one give to the results? We shall address these questions in Sec. 2.2 by analyzing three different ways of dealing with multiple interpretations: (1) avoid combination, (2) combine probabilities and accept a weaker interpretation of the results, and (3) avoid explicit treatment of non-frequentist uncertainty by using so-called predictive analysis. The second alternative relies on the hierarchical order that exists among some probability interpretations. Much of the contents of this section is from current research by Veneziano and Chung (1980).

2.1 The Time-Average (t.a.) Interpretation of Probability

In the case of seismic hazard, the interpretation of probability as degree of belief can be avoided. Consider, for example, the sequence of times at which strong earthquakes occur. It is typically found that these earthquakes take place without significant regularity or clustering in time, in a way that has been effectively described by Poisson point processes (the specific form of the process has no relevance in the present argument). Although the Poisson model--as any random point process--cannot describe the deterministic physical sequence exactly, there are at least two ways in which it may be correct:

1. It may be correct in the sense of degree of belief (d.o.b. correct) if it exactly quantifies uncertainty about (not intrinsic in) actual seismicity or, more important here,
2. It may be correct in the frequentist sense that probabilities calculated from the Poisson model coincide numerically with limits of relative frequency in time. In this case, the Poisson model is t.a. correct. As the definition makes obvious, t.a. models have the form of stationary random processes.

For the second interpretation to be possible, the physical sequence of earthquakes must possess some degree of regularity, so that limits of relative frequency in time exist. This condition does not imply time-invariance of the process that generates the natural sequence; geophysical laws, earthquake mechanism, tectonic stress or strain rates might vary in time and cause clustering of the earthquake events, and yet a stationary t.a. process may exist. Of course, in this case the process must have memory and not be one of the Poisson type. Limits of relative frequency in time satisfy the axioms of probability theory.

The t.a. interpretation of probability is not new; it was one of the first interpretations proposed in the physical sciences (Hofstetter, 1964). For example, it was used to represent the erratic variation of voltage at the terminals of a resistor and of velocity at a given point in turbulent air flow. Its spatial analogue (probability as limit of relative frequency in space) was exploited by Matheron (1971) to justify certain geostatistical procedures of mineral exploration and estimation.

Let x(t) denote the physical function of interest, e.g., x(t) = average shear stress on the plane of a fault at time t . It may not be easy to prove that for a given function x(t) an associated t.a. process X(t) exists, but this is of little consequence here. In application, x(t) is not known and assumptions are made directly on the random process X(t).

In this context, one faces the opposite problem, i.e., to establish whether a given random function X(t) is a legitimate t.a. process (whether or not a function x(t) exists that has X(t) as its associated t.a. process).

It is clear that t.a. processes do exist. For example, all processes X(t) that are ergodic in the sense that mean squares are legitimate t.a. processes.

It is also clear that there are infinitely many deterministic functions that are compatible with any legitimate t.a. process X(t); let x(t) be one such function. Then, translates of x(t) and functions that differ from x(t) only inside a finite interval of time still have X(t) as their t.a. process. More importantly, and as shown by the theory of ergodic processes, the family of deterministic functions that share the same t.a. process includes functions of most diverse form. Nor should one only think of deterministic functions that are realizations of stationary processes. Because of time-averaging, functions with nonstationary character also belong to this family. This indeterminacy of $x(t)$, given $X(t)$, is the mechanism through which t.a. probability represents uncertainty.

The point should be stressed that, although there are mathematical links between the theory of ergodic processes and the partial characterization of an unknown function through its t.a. process, there are also important differences. In the first case, one views the physical function itself as a random process, whereas in the second case, one deals with the t.a. properties of a class of deterministic functions.

The t.a. interpretation requires some modification of seismic hazard statements. Suppose that seismic hazard is expressed in terms of the probability distribution of A_t = peak ground acceleration at a given site during the next t years. The ensemble-frequentist interpretation is not applicable in this case because a statistical population of seismic histories at the site cannot be defined. In t.a. analysis, the definition of A_t must be changed and $F_{A_t}(a) = P[A_t \leq a]$ interpreted as the <u>relative frequency in time with which</u> a <u>is exceeded during intervals of</u> t <u>years</u>. Although specific reference to the next time interval is lost due to time-averaging, one may argue that this is a very reasonable interpretation of seismic hazard, if degrees of belief are to be avoided. Reference to selected intervals of time can still be made by way of <u>conditional</u> t.a. analysis, as shown next.

Making seismic hazard predictions on the basis of the stationary process $X(t)$ alone may not be appropriate if one has specific information on the present state of the physical system (e.g., from knowledge of the recent seismic history at the site). Information of this type can be incorporated into t.a. hazard statements by way of conditional probabilities. With A and B as any two events, the conditional t.a. probability $P[A|B]$ is defined as the limit of relative frequency in time with which event A occurs, given that event B occurs. It satisfies $P[A|B] = P[A \cap B]/P[B]$, in which all probabilities are t.a. probabilities. In application to seismic hazard, B represents information on the physical function $x(t)$ other than that already contained in the random process $X(t)$. It often happens that event B occurs with zero frequency in time, so that $P[B] = P[A \cap B] = 0$, hence the previous equation for $P[A|B]$ cannot be used. In this case one must replace probabilities with probability densities that have the t.a. meaning of densities of relative frequency in time.

In the estimation of seismic hazard, all available information on the physical function $x(t)$ should be used, not only to infer properties of the associated t.a. process $X(t)$, but also to update seismic hazard by calculating conditional t.a. probabilities. This is the mechanism through which t.a. analysis can express time dependence. It is only when no information on $x(t)$ is available--other than knowledge of the process $X(t)$--that seismic hazard predictions are independent of current time.

The t.a. interpretation makes it possible to give frequentist meaning to probabilistic statements of seismic hazard and suggests itself as a measure of rational d.o.b., whenever applicable. This is an extension of the notion that whenever the ensemble relative-frequency interpretation is possible, the beliefs of a rational individual should be consistent with frequentist probabilities.

2.2 Combination of Frequentist and Non-Frequentist Uncertainty

The frequentist interpretation of seismic hazard is conditional on a given t.a. process representation of seismicity, symbolized here by the random process X(t). However, X(t) is not always known. For example, one may know that X(t) is a homogeneous Poisson impulse process with random event sizes Y_i at the times t_i of earthquake occurrences,

$$X(t) = \sum_i Y_i \; \delta(t-t_i)$$

but parameters of the process (the mean rate of events and the marginal distribution of the I.I.D. variables Y_i) may not be exactly known. In other cases, the form of the process X(t) may also be unknown.

Uncertainty about the form and the parameters of the t.a. representation of seismicity cannot be measured by probabilities with frequentist interpretation. This fact causes some difficulty; if probability is to be used, then it becomes necessary to allow for subjective interpretations and the problem arises whether or not different interpretations "mix," i.e., whether or not in the course of analysis, d.o.b. probabilities can be meaningfully combined with frequentist probabilities. How should results be interpreted? This problem, which has important implications on the procedure of seismic hazard analysis and on the format and numerical value of hazard results, is currently the source of much controversy.

Advocates of the d.o.b. interpretation observe that measures of strength of belief are numerically compatible with frequentist probabilities whenever a frequentist interpretation is possible. Therefore, all probabilities--including those that result from mixing operations--can be interpreted in the same d.o.b. sense.

Other statisticians disagree, either because they object to quantifying the d.o.b. through probability, or because they cannot attribute any meaning to mixtures. At most, they accept statements of the type

$$\text{"The d.o.b. that } \Lambda(a) \le \lambda \text{ is } P_{dob}(\lambda)\text{"} \quad (a)$$

$$\text{"The d.o.b. that } \left\{P_{fr}[A_t \le a] \le p\right\} \text{ is } P_{dob}(p)\text{"} \quad (b) \qquad (2.1)$$

In the first equation, $\Lambda(a)$ is the mean rate at which peak ground acceleration at the site exceeds the value a. Subscripts fr and d.o.b. denote frequentist (for example, t.a.) and d.o.b. interpretations, respectively. In case (b), $\left\{P_{fr}[A_t \le a] \le p\right\}$ is viewed as an event to which a d.o.b. probability is assigned. Hence in this case--that of multiple interpretations without mixing--it is possible to talk of the "probability of a probability," in the sense of the d.o.b. in the relative frequency" of an event.

With reference to Eq. 2.1b, practical interest is in $P[A_t \le a]$. If one decides that mixing is possible, e.g., because frequentist probabilities also have the meaning of d.o.b., then $P[A_t \le a]$ can be found from the Total Probability Theorem, which gives

$$P[A_t \le a] = \int_0^1 p \, dP_{dob}(p) \qquad (2.2)$$

We have therefore identified three different attitudes toward the use and mixing of probability interpretations:

1. Different interpretations are allowed and probabilities can be mixed (Eq. 2.2).
2. Different interpretations may coexist without mixing (Eq. 2.1).
3. Only frequentist interpretations are allowed.

In order to focus on conceptual issues, we have been purposely schematic in classifying probability interpretations as either "frequentist" or d.o.b." Other interpretations exist and many (sometimes subtle) variants have been proposed with the frequentist and d.o.b. frameworks (Fine, 1973; Lucas, 1970). In the case of more than two interpretations, generalization of Eq. 2.1 (especially, the second statement) produces very confusing language, and option 2 should be avoided.

Option 1 (mixing) seems to be available only for probabilities with identical interpretations, and since a number of interpretations may be acceptable for the same probability, one should first determine whether it is possible to have a single common interpretation. A hierarchical order exists for some probabilities; for example, frequentist interpretations are stronger than d.o.b. interpretations, since for rational individuals the former imply the latter, whereas the reverse is not true. Similarly, the rational d.o.b. interpretation is stronger than the personal d.o.b. interpretation. Ordering is not possible in all cases, e.g., there is no ordering between the ensemble and the t.a. frequentist interpretations.

It may be that common interpretations are too weak to be useful. For example, one might interpret all probabilities as personal d.o.b., thus losing objectivity and having to face problems mentioned earlier in this section. Before mixing probabilities, one should therefore look for <u>the strongest possible common interpretation</u>. Also, in the context of seismic hazard, many seismologists and engineers seem to prefer statements of the type illustrated by Eq. 2.1 to the alternative of mixing probabilities and producing results with weaker d.o.b. interpretation.

Since mixing (Eq. 2.2) reduces the strength of interpretation and multiplicity of meanings (Eq. 2.1) produces unappealing statements, it would be valuable to develop procedures that combine uncertainties of different types and still retain a strong interpretation of the results (Option 3). This would seem to be an impossible task in the light of what has been just said. However, some such procedures are available which, for <u>specific problems</u>, account for parameter uncertainty using frequentist analysis exclusively. Results are in the form of <u>prediction regions</u> of given probability content and are obtained by the method exemplified next.

Consider for simplicity a stationary, independent sequence of variables X_i (i=1,2,...), the marginal CDF of which, $F_X(x)$, depends on an unknown vector of parameters $\underline{\theta}$. This might be the sequence of earthquake intensities at the site. Statistical information on $\underline{\theta}$ is obtained by observing n variables of the sequence, and we want to characterize uncertainty on the next unobserved variable X_{n+1}. One can do this by calculating "prediction" sets on the real line with given probability content. A prediction set of P-content is a <u>random</u> set $D = D(X_1, \ldots, X_n)$ such that for any given $\underline{\theta}$, D has probability content P, i.e.,

$$
\underset{X_1,\ldots,X_n}{E} P[D(X_1,\ldots,X_n)|\underline{\theta}] = \int_{\text{all } x_1,\ldots,x_n} \prod_{i=1}^{n} dF_{X|\underline{\theta}}(x_i) \cdot \int_{D(x_1,\ldots,x_n)} dF_{X|\underline{\theta}}(x) = P \tag{2.3}
$$

Notice that Eq. 2.3 treats the quantities $X_1,\ldots,X_n$ as random variables and that uncertainty in $\underline{\theta}$ is not being quantified. In particular, one may constrain D to be a one-sided interval of the type $(-\infty, \xi(X_1,\ldots,X_n,P)]$. Then $\xi(X_1,\ldots,X_n,P)$ is the random P-fractile of the so-called predictive distribution of X_{n+1}. As the sample size n increases, uncertainty in $\underline{\theta}$ decreases and prediction regions tend to become smaller. It should be emphasized that frequentist prediction sets may not exist (the problem is analogous to that of finding confidence intervals for unknown parameters).

The literature is rich in results on frequentist prediction regions for univariate and multivariate sequences and for simple and multiple (simultaneous) prediction (Proshan, 1953; Aitchison, 1964; Aitchison and Sculthorpe, 1965; Aitchison and Dunsmore, 1975; Chew, 1968; Guttman, 1970). For a review, see Veneziano (1974). In the case of multiple prediction (Chew, 1968), $D(X_1,\ldots,X_n)$ is a random set in the sample space of m unobserved variables from the sequence, $X_{n+1}, \ldots, X_{n+m}$.

Frequentist prediction regions are often numerically identical or very close to Bayesian prediction regions if, for the latter, $\underline{\theta}$ has (essentially) noninformative prior distribution.

The definition of the frequentist prediction region in Eq. 2.3 applies also to probabilities with t.a. frequentist meaning.

2.3 Representation of Seismic Hazard

It is proposed that an objective probabilistic measure of seismic hazard be defined as the frequency in time of relevant earthquake events. The purpose of hazard analysis is to estimate this objective probability from given information.

If information is limited, then estimation is imperfect and, whenever possible, one should obtain "predictive" frequentist statements of hazard which incorporate uncertainty on seismicity parameters without recourse to d.o.b. interpretations. If frequentist predictive statements are impossible, then an alternative is to quantify estimation uncertainty through d.o.b. probability on true hazard. That is, hazard probabilities are themselves considered as uncertain quantities with non-frequentist distributions. Bayesian analysis produces probabilities that correspond to the mean value of these distributions. The mean value is not necessarily the best estimate of seismic hazard; the very fact that many engineers do not make decisions only on the basis of Bayesian hazard indicates that other characteristics of the non-frequentist distribution of hazard are also important. When viewed from a "no-mixing" perspective, the Bayesian approach is correct but incomplete; it is still useful to quantify the bias of procedures that neglect estimation errors. In this sense we shall use Bayesian analysis in Sec. 3.2.2. Knowledge of the non-frequentist distribution of hazard is important to assess the impact of future research and data-gathering efforts aimed at improving knowledge of seismicity and procedures of hazard analysis.

3. RISK CONSEQUENCES OF ERRORS

Typical seismic hazard procedures generate single (point) estimates of frequentist hazard. Possibly wrong assumptions (e.g., spatial and temporal independence of earthquake events) and uncertainty on the value of parameters (e.g., of the magnitude-frequency and intensity-attenuation functions) are sources of estimation errors. Before quantifying these errors in Secs. 4, 5, and 6, we shall look at the influence that uncertainty on seismic hazard has on engineering risk. Should one find that risk is insensitive to this uncertainty, the importance of errors in hazard analysis would be much reduced.

The present task is complicated by two facts: (1) non-frequentist uncertainty on hazard depends on the seismic province and site in question, and (2) seismic risk depends on seismic resistance, which is itself uncertain and variable from facility to facility. Nevertheless, some information about typical uncertainty on hazard and variability of seismic resistance is available, thus allowing exploratory treatment to the problem. This can be done by:

1. Fitting parametric seismic hazard models and resistance distributions to available data (Sec. 3.1).
2. Calculating seismic risk with parameters either fixed to their mean values (thus ignoring estimation errors, Sec. 3.2.1) or treated as random variables (thus including the effect of uncertain estimation errors, Sec. 3.2.2).

3.1 Simple Models of Seismic Hazard and Seismic Resistance

For the purpose of calculating engineering risk (e.g., the probability or the mean rate of "failure" events), seismic hazard and resistance should be described in terms of the same ground motion intensity parameter, Y. For example,

$$
\begin{aligned}
&Y = \ell n\ a, \text{ where } a = \text{peak ground acceleration, or} \\
&Y = \ell n\ v, \text{ where } v = \text{peak ground velocity, or} \\
&Y = \ell n\ Q(\omega, \xi), \text{ where } Q(\omega, \xi) = \text{maximum pseudo-velocity of an} \\
&\qquad \text{oscillator with natural frequency } \omega \text{ and damping ratio } \xi, \text{ or} \\
&Y = I, \text{ where } I = \text{Modified Mercalli (MM) intensity}
\end{aligned}
\tag{3.1}
$$

<u>Models of Seismic Hazard</u>. Much experience has been accumulated on hazard functions in terms of the above intensity measures. Some such functions obtained for Boston under different modeling assumptions are shown in Fig. 3.1 as relationships between MM intensity I and the mean rate $\Lambda(I)$ at which events of larger local size occur (Cornell and Merz, 1975).

When plotted on semilog paper ($\ell n\,\Lambda$ versus Y), hazard functions are typically well approximated by straight lines in the low-to-medium intensity range but decay faster at higher intensities. Let R denote the resistance of the statistical population of facilities under consideration, expressed in the same units as Y, and denote by m_R and σ_R^2 the mean value and variance of R, respectively. Then failure occurs if $Y > R$ or equivalently if $Y' > R'$, in terms of the normalized variables

$$
\begin{aligned}
R' &= (R-m_R)/\sigma_R \\
Y' &= (Y-m_R)/\sigma_R
\end{aligned}
\tag{3.2}
$$

A few simple analytical hazard functions of the type $\Lambda'(y')$ = mean rate at which normalized site intensity y' is exceeded are given in Table 3.1 and plotted in Fig. 3.2 for selected parameter values. The associated hazard

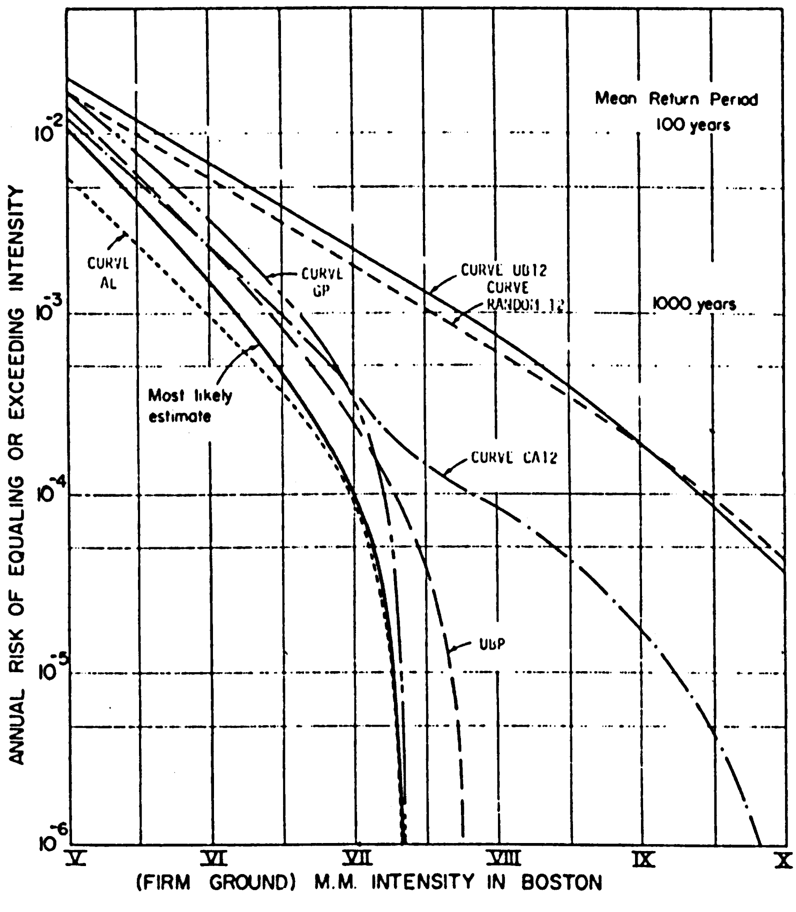


FIGURE 3.1 Influence of different assumptions on seismic hazard in Boston (from Cornell and Merz, 1975).

CASE	HAZARD MODEL	MEAN EXCEEDANCE RATE $\Lambda'(y')$	PARAMETERS
a	Linear	$\lambda' \; e^{-\beta' y'}$	λ', β'
b	Truncated linear	$\begin{cases} \lambda' \; e^{-\beta' y'}, & y' \leq y_1' \\ 0, & y' > y_1' \end{cases}$	λ', β', y_1'
c	Truncated linear $d\Lambda'(y')/dy'$	$\begin{cases} \lambda'(e^{-\beta' y'} - e^{-\beta' y_1'}), & y' \leq y_1' \\ 0, & y' > y_1' \end{cases}$	λ', β', y_1'
d	Truncated linear $d^2\Lambda'(y')/dy'^2$	$\begin{cases} \lambda'[(e^{-\beta' y'} - e^{-\beta' y_1'}) - \beta'(y_1' - y')e^{-\beta' y_1'}], & y' \leq y_1' \\ 0, & y' > y_1' \end{cases}$	λ', β', y_1'
e	Quadratic	$\lambda' \; e^{-\alpha' y'^2 - \beta' y'}$	$\lambda', \alpha', \beta'$

TABLE 3.1 Simple normalized hazard models.

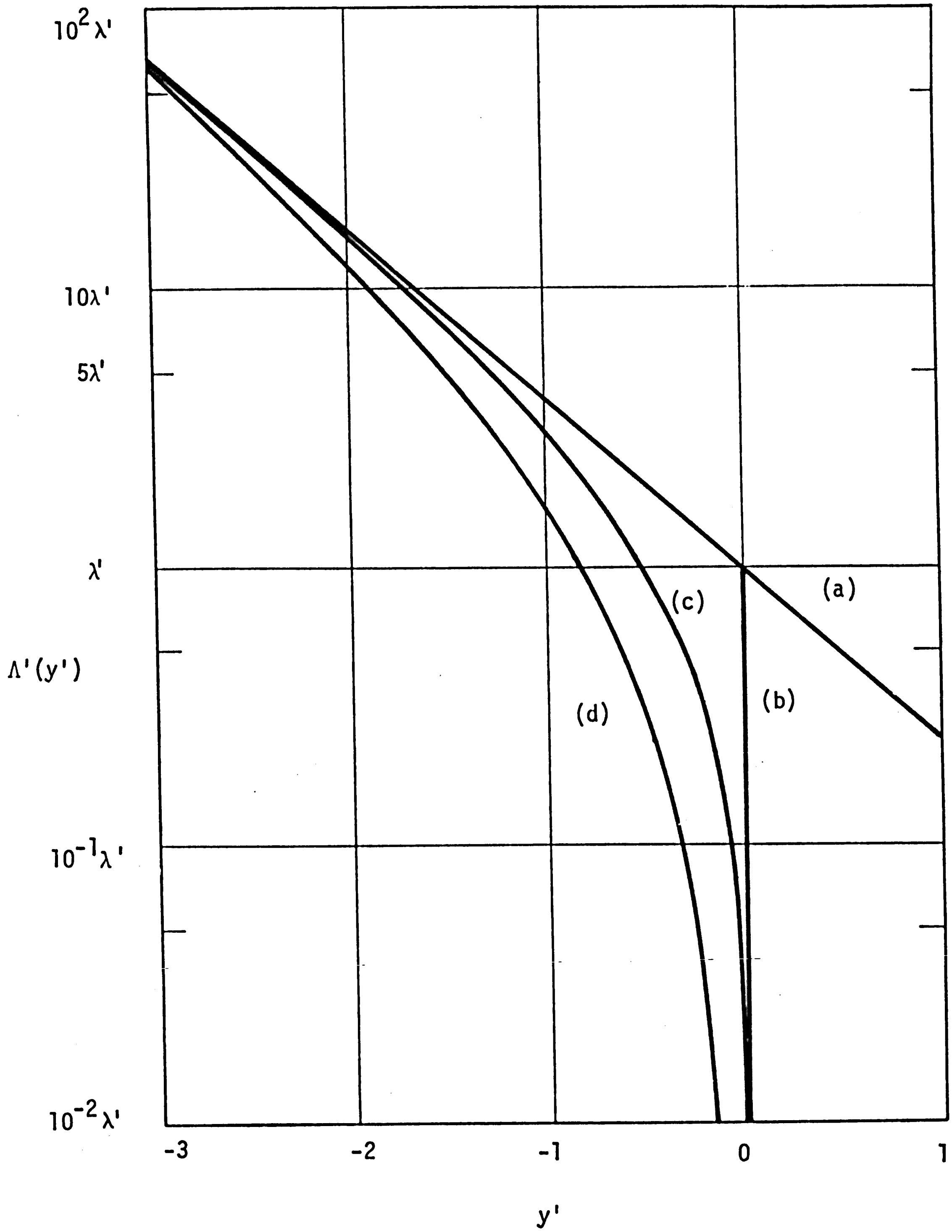


FIGURE 3.2 Untruncated and truncated "linear" hazard models in Table 3.1. In cases b, c, and d, y_1' is assumed zero.

functions in terms of abnormalized intensity, $\Lambda(y)$, can be found from

$$\Lambda(y) = \Lambda'((y-m_R)/\sigma_R) \tag{3.3}$$

For example, in the case of the first model in Table 3.1, Eq. 3.3 gives

$$\Lambda(y) = \lambda\ e^{-\beta y} \tag{3.4}$$

in which $\lambda = \lambda' e^{\beta' m_R}$ and $\beta = \beta'/\sigma_R$.

The hazard functions of Table 3.1 are flexible enough to accurately fit nearly all practical cases. For example, the hazard curves in Fig. 3.1 are well approximated by "linear" models with slope parameter $\beta = 1.1$ (curves UB 12 and RANDOM 12) or $\beta = 1.77$ (curve CA 12), or by "truncated linear" models with β between 1.7 and 2.0 (remaining curves). More sophisticated models in Table 3.1 would provide even better fits. They would also be conceptually more satisfactory because, as observed by many authors, there should be a physical upper limit to intensity; however, the truncated linear model unrealistically associates a nonzero mean rate (of magnitude $\lambda' e^{-\beta' y_1'}$) with this limit.

Models of Seismic Resistance. Errors in the hazard function may have substantially different consequences on risk, depending on the resistance characteristics of the exposed facility. Safety engineers have been especially concerned with estimation accuracy at high levels of intensity, hence for return periods $1/\Lambda$ that require extrapolation beyond the duration of historical records. This concern is justified by the increase of uncertainty on $\ell n\, \Lambda$ given y (on y given Λ) with increasing y (with decreasing Λ).

However, rare high intensity events may not be of determining importance in the calculation of risk.

Consider a population of engineering facilities (e.g., rockfill dams or high-rise buildings) the seismic resistance R of which has probability density function $f_R(y)$. The associated cumulative distribution function, $F_R(y) = \int_{-\infty}^{y} f_R(r)dr$, is sometimes referred to as the "fragility curve." It gives the probability of failure (the fraction of population that fails) if an earthquake occurs with site intensity y. Also, denote by $\lambda(y) = d\Lambda(y)/dy$ the mean rate <u>density</u> of events with site intensity exactly y, and let seismic risk be measured by the mean failure rate λ_f. For example, a value $\lambda_f = 10^{-2}$ failure/year means that if facilities from the population were exposed to probabilistically independent but otherwise identical seismic environments, the average life of the generic facility would be $E[T] = 1/\lambda_f = 100$ years. Under the assumption of Poisson events, the failure probability of the generic facility in t years would be $P_f(t) = 1-e^{-\lambda_f t}$.

The definition of R as the resistance of the generic facility of a statistical population makes risk results meaningful in a frequentist sense. If, however, only one facility exists, then f_R expresses non-frequentist uncertainty on the resistance of that facility and calculated risk can have only d.o.b. interpretation. For example, $\lambda_f = 10^{-2}$ failures/year must be taken to mean that the d.o.b. distribution of time to failure has a mean value of 100 years. In order to avoid statements of this type, which are often obscure, uncertainty on seismic resistance will always be given frequentist interpretation here.

The mean failure rate λ_f can be calculated by any of the following expressions:

$$\begin{aligned}
\lambda_f &= \int_{-\infty}^{\infty} f_R(y)\Lambda(y)dy && \text{(a)} \\
&= \int_{-\infty}^{\infty} F_R(y)\lambda(y)dy && \text{(b)} \\
&= \int_{-\infty}^{\infty} f_{R'}(y')\Lambda'(y')dy' && \text{(c)} \\
&= \int_{-\infty}^{\infty} F_{R'}(y')\lambda'(y')dy' && \text{(d)}
\end{aligned} \qquad (3.5)$$

in which R' is the standardized resistance in Eq. 3.2 and λ' is the derivative of Λ'.

Errors may be present not only in the estimated hazard (functions Λ and λ) but also in the estimated resistance distribution (functions F and f). According to Eqs. 3.5a and 3.5b, λ_f is a weighted average of F_R, f_R, Λ, or λ with weight functions λ, Λ, f_R, and F_R, respectively. Hence, for example, in order to evaluate the seismic risk consequences of errors in $\Lambda(y)$ for different intensity values y, one should look at the corresponding weight function, f_R: where the latter function is maximum (at the mode of R) errors in Λ are most critical. This is actually true for errors of fixed magnitude, but in the case of Λ, errors tend to be larger for smaller intensities (notice that interest here is in Λ, not in $\ln \Lambda$). One concludes, then, that the most critical region for accuracy of Λ is <u>below</u> the mode of R.

It is also of interest to look at the whole integrands of Eqs. 3.5a and 3.5b, as they vary with y. The integrand of the first equation, $f_R(y)\Lambda(y)$, gives the contribution to the failure rate <u>from facilities with different resistance</u>, whereas the integrand of the second equation,

$F_R(y)\lambda(y)$, gives the contribution to λ_f from earthquakes of different intensity. The first function is maximum below the mode of R; this is usually true also for the second function.

The early belief that uncertainty on seismic resistance contributes little to overall risk and that even models with deterministic resistance produce valuable results has been seriously questioned. Analysis in Sec. 3.2 also indicates that variability of R may be critical, especially when site intensity has an upper bound or statistical uncertainty on resistance distribution parameters is included. It appears that for some engineering facilities a substantial fraction of risk comes from medium-size earthquakes that, although associated with small failure probabilities, are much more frequent than large and more destructive events.

At the present time, the probabilistic prediction of seismic resistance by explicit analytical means (e.g., random vibration) is unreliable for all but the simplest systems. Repeated deterministic analyses, using simulated ground motions conditional on site intensity and simulated system behavior and resistance parameters, are more general but also very expensive procedures. For some types of facilities such as earth dams, building frames, and major structural components of nuclear power plants, simplified analytical procedures (Banon and Veneziano, 1981; Hasselman and Simonian, 1980) that may prove feasible and sufficiently accurate are now under development. In many cases, however, the best source of information is historical seismic performance. Much of this information has been collected in terms of the damage ratio (DR), which is the cost of repair divided by the cost of replacement. Data on wooden frame buildings and on ordinary or reinforced masonry construction has been available for some time. In the last few years, damage statistics on high-rise buildings, large dams, and important engineering facilities have also been gathered, in part (e.g., in the nuclear industry) through large-scale testing (see

Proceedings of SMiRT Conferences: Berlin, 1971, 1973; London, 1975; San Francisco, 1977). For high-rise buildings, reliable information has been collected from the 1971 San Fernando earthquake. The most extensive survey (Whitman, 1973) documents 368 buildings of five stories or more, classified by age, structural material, and height. Fig. 3.3 displays mean damage ratio (MDR) data on high-rise buildings from the San Fernando as well as other earthquakes. The only classification here is by UBC zone where all heights and construction materials (steel and concrete) are lumped together. Based in part on statistical data, in part on professional judgement, curves relating MDR to MM intensity have been proposed by several authors. Those in Fig. 3.4 for high-rise buildings are from Whitman (1973). Other estimates can be found in Mann (1974), Whitman and Hong (1973), and Benjamin (1974). For most cases and over a wide range of intensities, the expected log damage ratio varies almost linearly with I and can be represented as

$$E[\ell n\ DR|I] = a_d + b_D I \tag{3.6}$$

with constants a_D and b_D that depend on the type of facility.

Some information is also available on the dispersion of the conditional damage ratio, DR|I. Benjamin (1974) found that for broad classes of buildings, the damage data for given intensity are fitted well by either lognormal or gamma distributions. If the lognormal model is used, then (ℓn DR|I) has normal distribution. The same author found that the variance of (ℓn DR|I) is approximately constant with I. Damage statistics in Whitman (1973) confirm that $\sigma_{\ell nDR|I}$ is not sensitive to I and is compatible with a normal distribution of (ℓn DR|I).

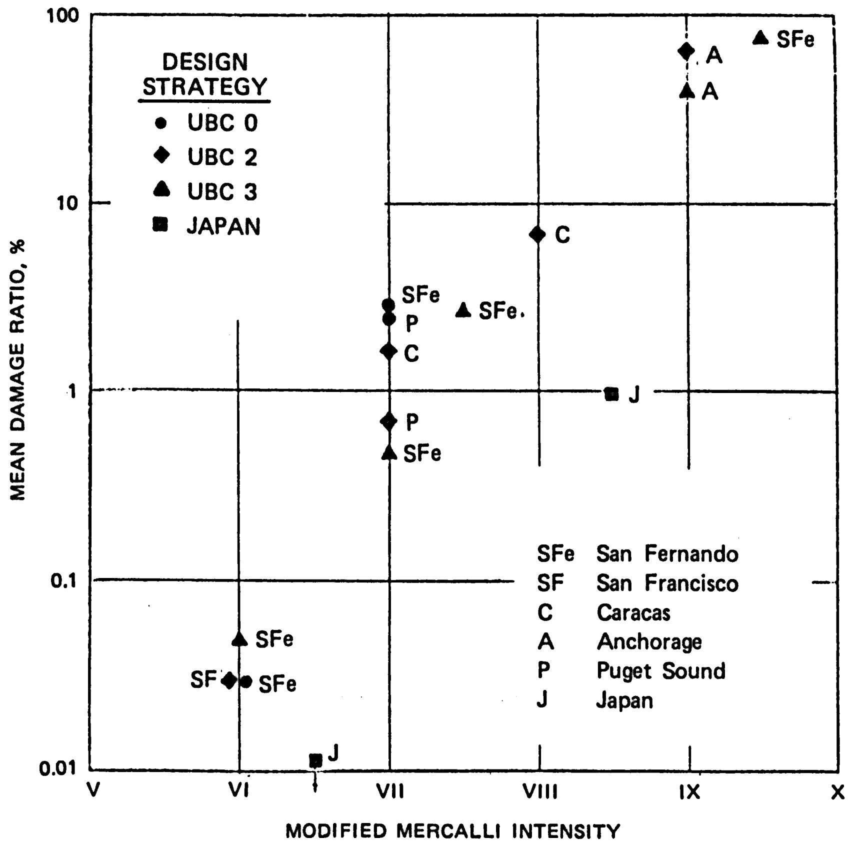


FIGURE 3.3 Data on mean damage ratio for high-rise buildings (from Whitman, 1973).

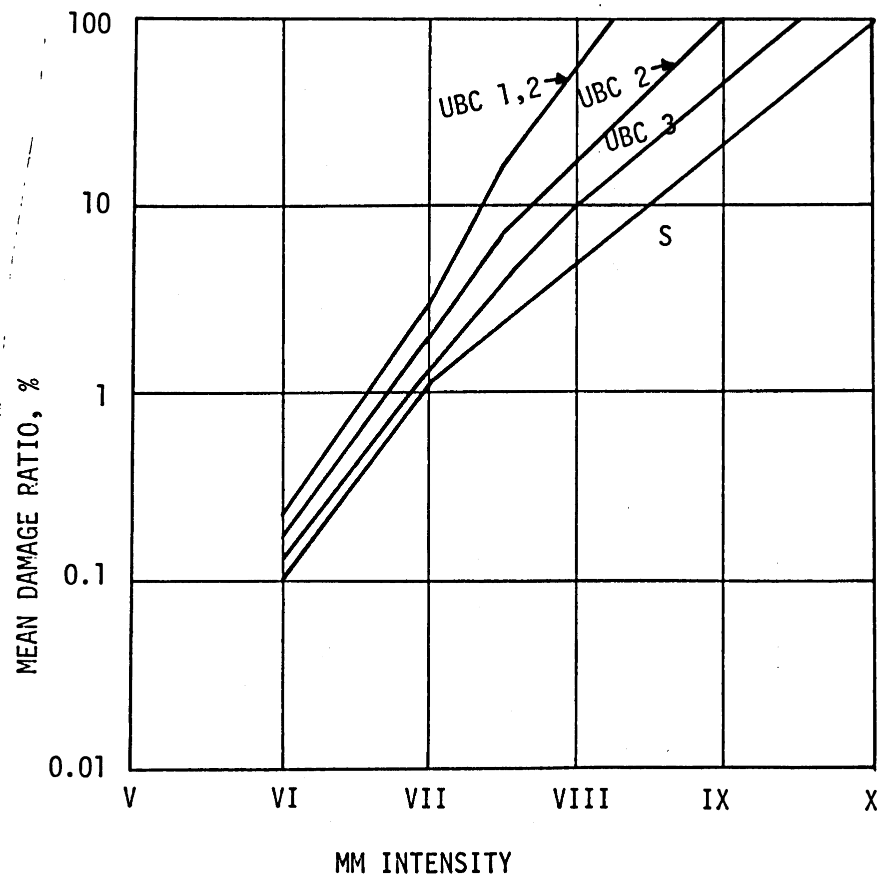


FIGURE 3.4 Mean damage ratio for high-rise buildings designed under different UBC zone requirements. S denotes "superzone" with twice the lateral force for Zone 3 (from Whitman, 1973).

In later analyses, it is not the value of $\sigma_{\ell nDR|I}$ that counts, but rather the ratio

$$\beta_D = \frac{b_D}{\sigma_{\ell n\ DR|I}} \tag{3.7}$$

Table 3.2 collects some statistics and subjective evaluations of the parameters b_D and β_D. The values from Newmark (1974) and Vanmarcke (1971) are not strictly compatible with those from Benjamin (1974) and Whitman (1973) because the former refer to given peak ground acceleration, a, instead of given MM intensity. For example, Newmark suggests values of $\sigma_{\ell n(\text{response})}/a$ for ordinary buildings and nuclear reactor structures and equipment (parameter BETA in his Table 3). If the level of response is proportional to a and a varies by a factor of 2 per unit of I, then the response varies also by a factor of 2 per unit of I and estimates of β_D in Eq. 3.7 can be found from

$$\beta_D = \frac{\ell n\ 2}{\sigma_{\ell n(\text{response})/a}} \tag{3.8}$$

This last relationship has been used to calculate values of β_D in Table 3.2 from estimates of $\sigma_{\ell n(\text{response})|a}$.

In making a model of probabilistic resistance, the following assumptions are introduced:

1. "Failure" occurs when the damage ratio reaches a critical value D*;
2. The conditional expected ℓn damage ratio is a linear function of I, of the type in Eq. 3.6; and
3. The conditional distribution of ℓn given I is normal, with constant variance σ_D^2.

		b_D	β_D
Benjamin (1974)	upper bound lower bound	0.484 0.347	1.64 1.54
Whitman (1973)	Post-1947 Buildings San Fernando, I=6 San Fernando, I=7 San Fernando, I=7.5	 ≃1.13 ≃1.13 ≃0.91	 ≃1.88 ≃2.13 ≃1.90
Newmark (1974)	Nuclear Reactor Structure Nuclear Reactor Equipment		1.33 1.16
Vanmarcke (1971)	I≃6.7 I≃7.8	≃0.68 ≃0.59	1.24 1.53

TABLE 3.2 Values of b_D and β_D in Eqs. 3.6 and 3.7.

It follows from 1, 2, and 3 that the probability of failure from an earthquake of intensity I is

$$P_f(I) = \Phi\,[(D^* - a_D - b_D I)/\sigma_D] \tag{3.8}$$

in which Φ is the standard normal CDF. From the fact that $P_f(I)$ is also the probability that $R < I$, one concludes that, when expressed in units of MM intensity, R has normal distribution $N(m_R, \sigma_R^2)$, with parameters

$$\begin{cases} m_R = (D^* - a_D)/b_D \\ \sigma_R^2 = \sigma_D^2/b_D^2 = \beta_D^{-2} \end{cases} \tag{3.9}$$

Also, in terms of normalized intensity $Y' = I' = (I - m_R)/\sigma_R$, resistance R' has standard normal distribution.

Results on the distribution type for R are not expected to change if ground motion intensity at the site is measured in terms of any other parameter Y in Eq. 3.1.

3.2 Calculation of Seismic Risk

Given the models of seismic hazard in Table 3.1 and given normality of the distribution of R, the mean failure rate λ_f is conveniently calculated from Eq. 3.5c. Results should first be obtained when all uncertainty is of the frequentist type (when the parameters of Λ' and f_R are known) and then consider a few cases in which parameter uncertainty is included.

3.2.1 Cases with Known Seismic Hazard

Combinations of $R' \sim N(0, 1)$ with any of the functions Λ' in Table 3.1 are considered first. Sensitivity to the distribution of R' is then checked by assuming that resistance has gamma distribution and that Λ' has untruncated linear form. In all these cases, λ_f is found by closed-form integration of Eq. 3.5c.

a. Linear Hazard

$$\lambda_f = \frac{\lambda'}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\beta' y' - y'^2/2} \, dy' = \lambda' e^{\beta'^2/2} \tag{3.10}$$

For hazard expressed in units of MM intensity and using Table 3.2 for β_D, one finds that the parameter $\beta' = \beta/\beta_D$ (β is the slope of the seismic hazard curve) varies between 0.6 and 1.2, with a typical value of 0.9. The constant λ' in Eq. 3.10 is the mean rate at which earthquakes occur with intensity in excess of the mean resistance; hence, the quantity $\gamma = e^{\beta'^2/2}$ can be viewed as a penalty factor for uncertain resistance; it increases with σ_R ($= 1/\beta_D$) and with β and is 1 if either parameter is zero. In the present model, typical values of γ are in the range 1.2 to 2.1, but much higher values may result from using nonlinear Λ' models or from including statistical uncertainty on σ_R into the analysis.

b. Truncated Linear Hazard

In this case λ_f is given by

$$\lambda_f = \frac{\lambda'}{\sqrt{2\pi}} \int_{-\infty}^{y_1'} e^{-\beta' y' - y'^2/2} \, dy'$$

$$= \lambda' \, e^{\beta'^2/2} \, \Phi(y_1' + \beta') \tag{3.11}$$

in which $y_1' = (y_1 - m_R)/\sigma_R$ is normalized upper bound intensity. The factor $e^{\beta'^2/2}\Phi(y_1' + \beta')$, which includes the effect of truncation, is plotted in

Fig. 3.5 as a function of y_1' for selected values of β'. For $y_1' > 1$, the mean failure rate in Eq. 3.11 is essentially the same as for the untruncated case. In order to reduce λ_f to 1/2 the value for untruncated hazard, it must obviously be that $y_1 = m_R - \sigma_R\beta$, i.e., the upper bound intensity at the site must be about one standard deviation of resistance below the mean resistance.

An example of application to Boston is given in Fig. 3.6 and Table 3.3. The figure shows five truncated or untruncated linear approximations to the hazard curves of Fig. 3.1. For the case of resistance R (in units of MM intensity) with normal distribution N(8, $(0.8)^2$), the parameters $\beta' = \beta\sigma_R$, λ', and $y_1' = (y_1 - m_R)/\sigma_R$ for the five approximations are given in Table 3.3, columns 2, 3, and 4, respectively. Column 5 gives the associated failure rate. Consideration of the modeling assumptions behind each case indicates that truncation of the site intensity-frequency law has relatively small effect on λ_f unless truncation is significantly below the mean resistance (compare cases 2 and 3). However, truncation of the epicentral intensity-frequency relationship is more important because it also reduces the value of λ'. Compare case 1 (truncation at I = XII for all the sources) with cases 3 and 5 in which upper bound intensities between 6.3 and 7.3 were assumed for seismic sources near Boston. The increase of β' due to truncation of epicentral intensity has only a minor balancing effect.

Integration of Eq. 3.5c for cases c and d in Table 3.1 gives the following failure rates.

c. Truncated Linear First Derivative of Λ

$$\lambda_f \doteq \lambda' \left[e^{\beta'^2/2}\,\Phi(y_1'+\beta') - e^{-\beta' y_1'}\,\Phi(y_1')\right] \tag{3.12}$$

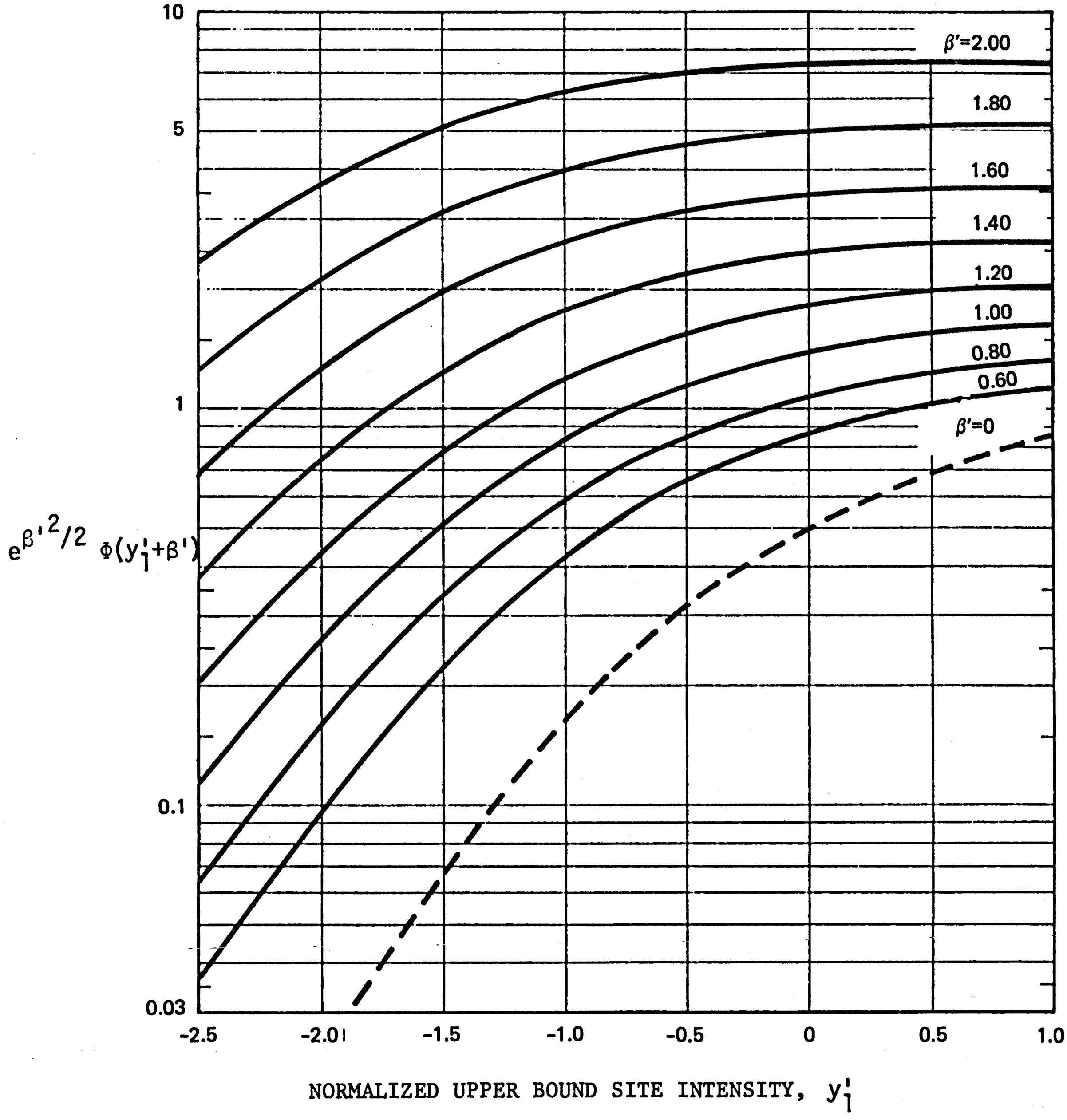


FIGURE 3.5 Truncated linear hazard model. Factor $e^{\beta'^2/2}\ \Phi(y_1'+\beta')$ for a few values of β'.

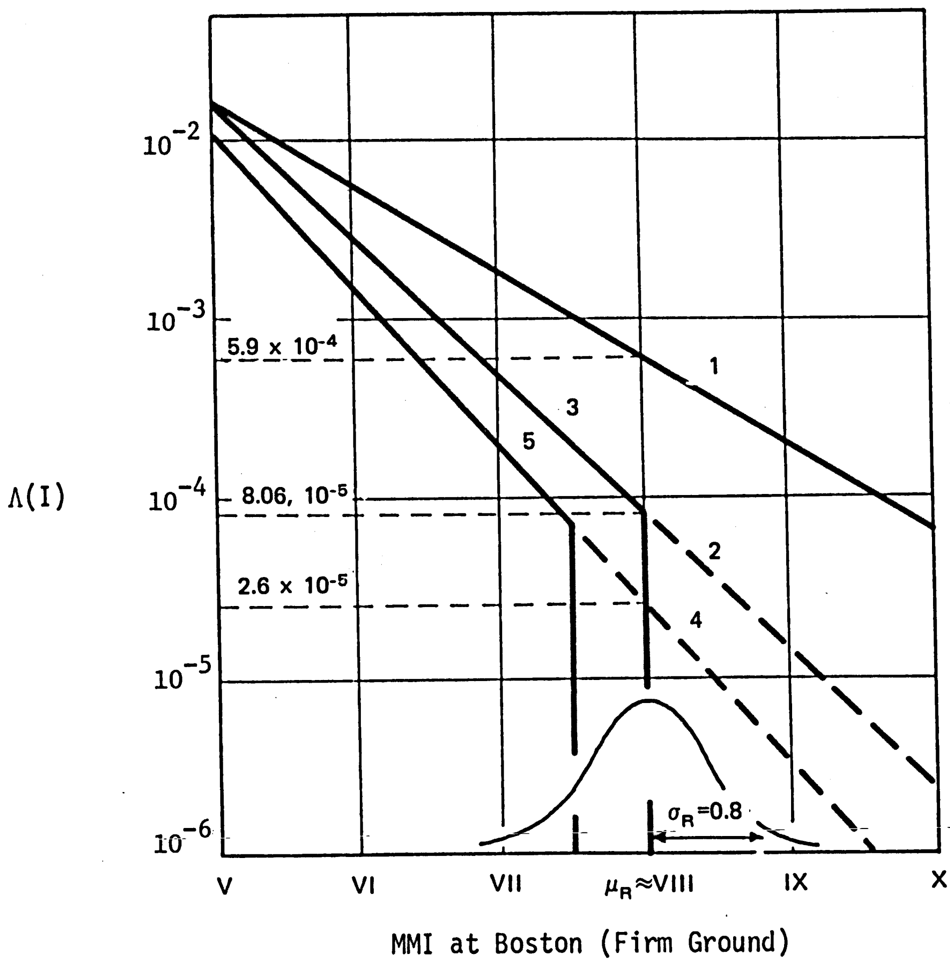


FIGURE 3.6 Linear and truncated linear approximations to the hazard functions in Fig. 3.1.

HAZARD CURVE (Fig. 3.6)	$\beta' = \beta\sigma_R$	λ'	y_1'	λ_f
1	0.88	5.9×10^{-4}	∞	8.7×10^{-4}
2	1.55	8.6×10^{-5}	∞	2.9×10^{-4}
3	1.55	8.6×10^{-5}	0	2.7×10^{-4}
4	1.58	2.6×10^{-5}	∞	9.1×10^{-5}
5	1.58	2.6×10^{-5}	-0.625	5.3×10^{-5}

TABLE 3.3 Parameters and failure rate for the five hazard curves in Fig. 3.6.

d. Truncated Linear Second Derivative of Λ

$$\lambda_f = \lambda' \left\{ e^{\beta'^2/2}\, \Phi(y_1'+\beta') - e^{-\beta' y_1'} [(1+\beta' y_1')\Phi(y_1') + \beta'\phi(y_1')] \right\} \tag{3.13}$$

in which ϕ is the standard normal probability density function. The parameter λ' in the last two equations is the mean rate of failure for $R = m_R$, in the case of no truncation, i.e., $y_1 \to \infty$.

One should be cautioned against directly comparing the mean crossing rates of the last three models because when using them to approximate actual hazard functions, one would conceivably select different values of y_1' and possibly also of λ' and β'.

e. Quadratic Hazard

$$\lambda_f = \frac{\lambda'}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\alpha' y'^2 - \beta' y'}\, e^{-y'^2/2}\, dy$$

$$= \frac{\lambda'}{\sqrt{2\alpha'+1}}\, e^{\beta'^2/(4\alpha'+2)} \tag{3.14}$$

Drawbacks of the quadratic hazard model are that it has no upper bound and that it violates the condition that Λ should be a nonincreasing function of y. Results using this model are typically very close to those from model c.

f. Linear Hazard and Gamma Distribution of Resistance

In order to test sensitivity of risk to the shape of the distribution of R, suppose that R' has shifted gamma distribution with zero mean and unit variance, hence with probability density function

$$f_{R'}(y') = \begin{cases} 0 , & y' < -D \\ \dfrac{D[D(y'+D)]^{D^2-1}}{\Gamma(D^2)} \; e^{-D(y'+D)} , & y' \geq -D \end{cases} \tag{3.15}$$

in which D is a positive constant. Plots of this function are shown in Fig. 3.7 for selected values of D. Notice that the above gamma distribution reduces to a shifted exponential for D = 1 and to the standard normal for $D \to \infty$. For Λ' that corresponds to the untruncated linear model, the mean failure rate is

$$\begin{aligned} \lambda_f &= \frac{\lambda' D}{\Gamma(D^2)} \int_{-D}^{\infty} e^{-\beta' y'} \; [D(y'+D)]^{D^2-1} \; e^{-D(y'+D)} \; dy' \\ &= \lambda' \; e^{\beta' D} \left(\frac{D}{D+\beta'}\right)^{D^2} \end{aligned} \tag{3.16}$$

The ratio between this rate and that for normal resistance distribution with the same mean value and variance (see Eq. 3.10) is given by

$$\frac{\lambda_f(\text{gamma})}{\lambda_f(\text{normal})} = \left(\frac{D}{D+\beta'}\right)^{D^2} \; e^{\beta' D - \beta'^2/2} \tag{3.17}$$

This ratio is shown in Fig. 3.8 as a function of β' for D = 1, 2, 3, 5, 8. The failure rate ratio is generally smaller than 1, primarily because the gamma density vanishes for $y' < -D$. This fact shows the importance of left truncation of the resistance distribution. However, if one excludes the artificious cases when $D \leq 1$ and for typical values of β' (e.g., $\beta' < 1.5$), the mean failure rate is insensitive to the present change of resistance

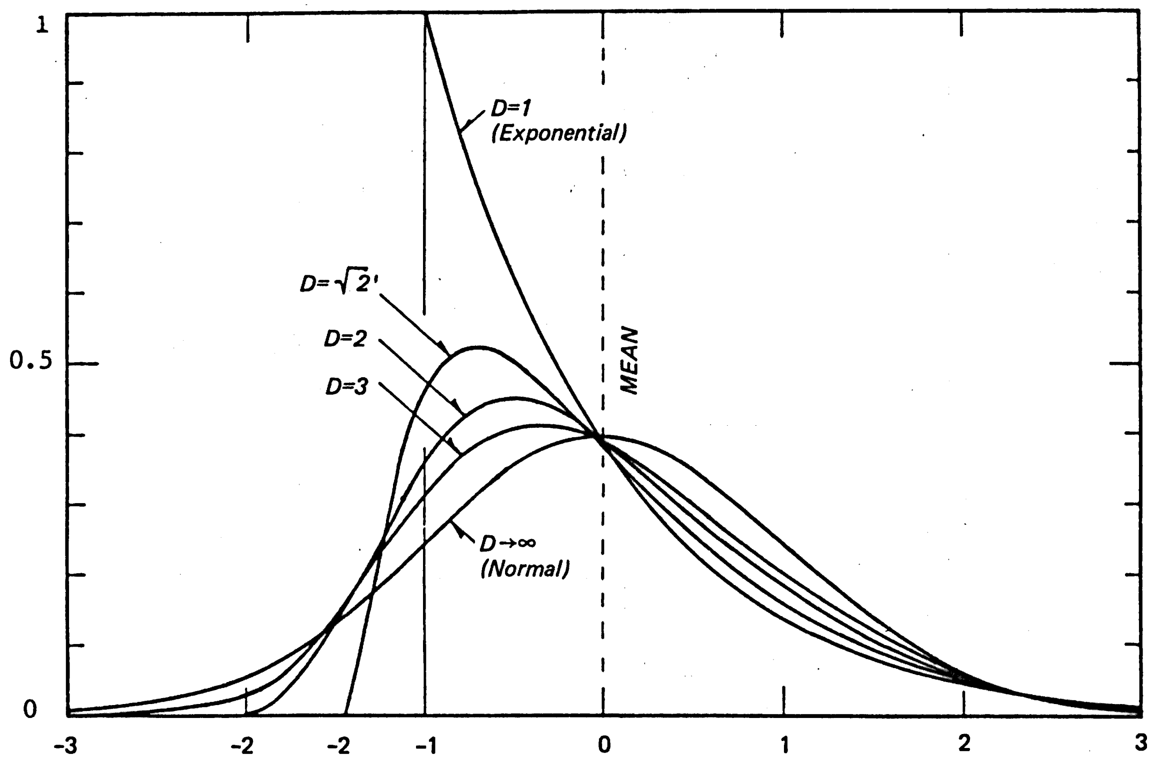


FIGURE 3.7 Shifted gamma densities with zero mean and unit variance, Eq. 3.15.

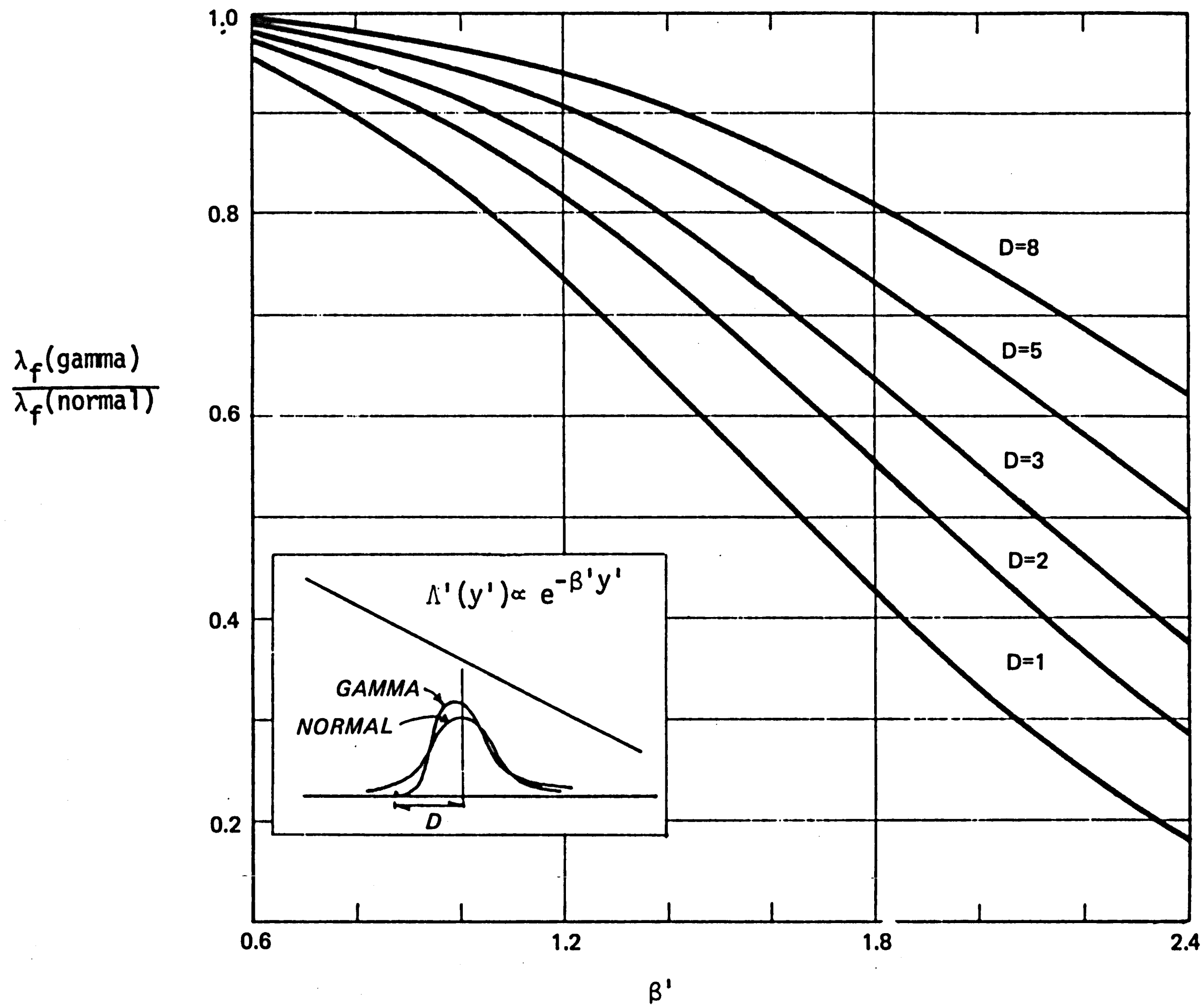


FIGURE 3.8 Plots of the mean failure rate ratio $\frac{\lambda_f(\text{gamma})}{\lambda_f(\text{normal})}$ in Eq. 3.17.

distribution. Higher sensitivity is found with respect to distributions that allocate more probability to values of R below the mean.

3.2.2 Cases with Uncertain Seismic Hazard

Non-frequentist uncertainty on the mean failure rate λ_f is contributed by:

1. Possible errors in the seismic hazard procedure (M_T might not belong to $\mathfrak{M}$; see Sec. 1.2). One might call these "model errors";
2. Uncertainty on which model in $\mathfrak{M}$ corresponds to the true model M_T, given that $M_T \in \mathfrak{M}$. If $\mathfrak{M}$ is a parametric family of models, then this source of error might be called "parameter uncertainty";
3. Errors in the form of the distribution of R and on the resistance parameters, m_R and σ_R^2.

Except for very few cases, the distribution of λ_f induced by these possible errors and uncertainties must be found through straightforward but tedious numerical integration, which requires that very many risk calculations be made, each conditional on different modeling assumptions or parameter values. The fact that realistic approximations to the hazard function lead to explicit expressions for λ_f suggests an alternative two-step procedure which, although approximate, is much more expeditious. First, one quantifies uncertainty on the hazard function $\Lambda(y)$ due to possible errors in the assumptions and to uncertainty on the parameters of the seismic hazard procedure. (Results on uncertainty on Λ will be given later in Secs. 4, 5, and 6.) Then one relates uncertainty on $\Lambda(y)$ to the probability distribution or at least to the expected value of λ_R. The second step becomes particularly simple if $\Lambda(y)$ has the form of model a in Table 3.1, with uncertain parameters λ' and β'. Uncertainty on the parameters of the resistance distribution can also be considered. The results that follow on the mean of λ_f are limited to this case. In Bayesian analysis, one is interested only in the probability $E[P_f(t)]$ and for small risk this is closely approximated by $E[\lambda_f] \cdot t$. In classical analysis, one would typically be interested in the full distribution of λ_f (or of $P_f(t)$), but one can still use the mean value $E[\lambda_f]$ to calculate the bias of estimates that do not account for uncertainty on seismic hazard or on resistance distribution parameters. In some cases the full distribution of λ_f is obvious from its dependence on the uncertain parameters; in other cases it can be found in approximation or through numerical convolution.

Under the assumption that Λ' has untruncated exponential form (model a in Table 3.1) and that R has normal distribution $N(m_R, \sigma_R^2)$, non-frequentist uncertainty is limited to the four parameters λ', β', m_R, and σ_R^2. Results on $E[\lambda_f]$ for all combinations of known/unknown parameters with appropriate distributions are given in Veneziano (1975); those for known resistance distribution are reproduced here in terms of the ratio γ between $E[\lambda_f]$ and the mean failure rate when the uncertain parameters are fixed to mean values, Eq. 3.10. Hence, γ can be interpreted as a bias factor due to parameter uncertainty. Let m_R and σ_R^2 be known. Then for uncertain λ' or β' one finds:

1. If β' is known and λ' is uncertain with lognormal distribution such that $\ell n\ \lambda' \sim N\ (m_{\ell n \lambda'}, \sigma^2_{\ell n \lambda'})$, then

$$\gamma_{\lambda'} = \frac{E_{\lambda'}[\lambda_f]}{\lambda_f|\lambda'=\exp(m_{\ell n \lambda'})} = \exp(\sigma^2_{\ell n \lambda'}/2) \tag{3.18}$$

   In the range of $\exp(\sigma^2_{\ell n \lambda'})$ from 0.4 to 1.1, which corresponds to a 1-sigma uncertainty factor on λ' between 1.5 and 3, the ratio $\gamma_{\lambda'}$ varies from 1.1 to 1.8. In this case, λ_f has itself lognormal distribution.

2. If λ' is known and β' has normal distribution $N(m_{\beta'}, \sigma^2_{\beta'})$, then the mean failure rate ratio $\gamma_{\beta'}$ is

$$\gamma_{\beta'} = \frac{E_{\beta'}[\lambda_f]}{\lambda_f|_{\beta'=m_{\beta'}}} = (1-\sigma^2_{\beta'})^{-1/2} \exp[m^2_{\beta'} \sigma^2_{\beta'}/2(1-\sigma^2_{\beta'})] \tag{3.19}$$

For realistic distribution parameters ($m_{\beta'}$ between 0.8 and 1.6 and $\sigma_{\beta'}$ between 0.1 and 0.3), $\gamma_{\beta'}$ varies in the range 1.01 to 1.20, with the upper limit very sensitive to the assumed maximum value of $\sigma_{\beta'}$.

3. If λ' and β' are independent with the marginal distributions given above, then the factor $\gamma_{\lambda',\beta'}$ is simply the product of the marginal factors, $\gamma_{\lambda',\beta'} = \gamma_{\lambda'}\gamma_{\beta'}$. Most often, however, the assumption of independence is inappropriate. It is found that if $\ell n\ \lambda'$ and β' have correlation coefficient ρ, then the factor $\gamma_{\lambda',\beta'}$ can be decomposed into three partial factors:

$$\gamma_{\lambda',\beta'} = \gamma_{\lambda'}(\rho) \cdot \gamma_{\beta'} \cdot \gamma_\rho \tag{3.20}$$

where $\gamma_{\beta'}$ is given by Eq. 3.19,

$$\gamma_{\lambda'}(\rho) = \exp[(1-\rho^2)\ \sigma^2_{\ell n\lambda'}/2]$$

and

$$\gamma_\rho = \exp[\rho\sigma_{\ell n\lambda'}(\rho\sigma_{\ell n\lambda'}+2m_{\beta'}\sigma_{\beta'})/2(1-\sigma^2_{\beta'})]$$

For $\rho = 0$, Eq. 3.20 reproduces the result for independent parameters. If $\rho = 1$, then uncertainty on λ' is totally explained by β' and

$\gamma_{\lambda'}(1) = 1$; if in addition $\sigma_{\ell n\lambda'} = 0$, then λ' is known and $\gamma_{\lambda',\beta'} = \gamma_{\beta'}$. Explicit results when m_R and σ_R^2 are uncertain (with conjugate distributions in random sampling) are not quoted here. Whereas uncertainty on m_R has moderate consequences on $E[\lambda_f]$, uncertainty on σ_R^2 may be very important because it considerably increases the probability of small resistance values to the point that for exponential hazard, many failures may result from ground motions of small intensity. For detailed results, see Veneziano (1975).

Although limited by the simple forms of the hazard and resistance distribution functions, results of this section should be kept in mind when evaluating the consequences of errors in hazard analysis; in particular, errors that affect the upper bound intensity and errors in the seismicity rates and in the slope of intensity-frequency relationships.

4. ERRORS FROM UNCERTAINTY ON MODEL PARAMETERS

Error analysis in engineering seismic risk encompasses two problems: (1) evaluation of non-frequentist uncertainty on seismic hazard, and (2) calculation of the consequences of this uncertainty on risk. Problems of the second type have been examined in the last section, whereas the remainder of the report focuses on the first task. This is a complicated one because of the multiplicity of error sources and seismic hazard procedures. It is subdivided into three parts and each part treated in a separate section.

The present section looks at the variation of hazard due to modeled uncertainty on or to reasonable variations of the parameters, while Sec. 5 quantifies the consequences of unmodeled errors. Hence, it is assumed here that the true model M_T is of the type hypothesized in hazard analysis (that $M_T \in \mathfrak{M}$) whereas the opposite assumption is made in Sec. 5. In application, there is a tendency to rely on expert opinion in order to include nonstatistical information and reduce estimation errors. Expert-based estimates are themselves not error-free. Conceptual and mathematical models and some results on the variability of expert opinion will be presented in Sec. 6.

Returning to the objective of the present section, the aim is to quantify the uncertainty on hazard (e.g., on the mean exceedance rate $\Lambda(a)$) due to the uncertainty of the form and parameters of the seismicity model, e.g., to uncertainty on shape, size, and location of earthquake sources, on the activity of each source (magnitude distribution and upper bound, earthquake occurrences in time and space), on the form and parameters of the attenuation function, and on site amplification effects.

Use of probabilistic analysis to propagate uncertainty through hazard calculations is made difficult by two factors: (1) the relationship between the uncertain quantities and hazard may not be explicit but rather given in algorithmic form and (2) it is difficult to quantify uncertainty about some aspects of the model, e.g., the magnitude distribution or the spatial dependence of earthquake epicenters. As a consequence, one must resort to different procedures for different parameters, ranging from exact analytical treatment to numerical sensitivity. Powerful, although approximate, analytical techniques have been recently developed by McGuire and Shedlock (1980) and will be reviewed first in Sec. 4.1. Results from parameter sensitivity analyses given the form of the model will follow in Sec. 4.2. Finally, Sec. 4.3 will comment on the effect of changing basic assumptions in the hazard model.

4.1 Analytical Results

Cases of parameter uncertainty that lend themselves to analytical treatment become fewer as complexity of the model increases. The following model, simple but realistic, has been studied by McGuire and Shedlock (1980).

Earthquake sources are independent. They generate events at Poisson times, with truncated exponential distribution of magnitude. Earthquakes originate from fault ruptures and, given magnitude, rupture length has truncated lognormal distribution (truncation is due to the finite length of the faults). Peak ground motion characteristics at the site are related by an attenuation function to magnitude and shortest distance to the rupture segment. Conditional on magnitude and distance, they too have lognormal distribution. Uncertainty on hazard at the site is studied by treating model parameters either as known constants, or as mutually independent random variables.

The main approximation of the method consists of treating earthquake sources as points in space, thus replacing the random distance to the site with a given constant for each source. Similarly, single values are used to replace magnitude distributions.

Seismic hazard is given in terms of the peak ground acceleration a_{P_0} that corresponds to a specified annual probability of exceedance, P_0.

It was found by McGuire and Shedlock that the distribution of a_{P_0} due to parameter uncertainty has a mean value (= Bayesian estimate) practically identical to the value calculated by fixing all parameters to their mean value. This unbiased property of conventional hazard procedures was attributed to almost linear dependence of a_P on the uncertain parameters. (But notice that bias may be large if there are errors in model assumptions, as shown in Sec. 5, or if hazard is nonlinear in the uncertain parameters; see e.g. Fig. 4.4).

Sources and magnitude of non-frequentist uncertainty vary from case to case. In an application to the San Francisco Bay area, McGuire and Shedlock considered the following parameters to be uncertain: the depth of energy release, the activity rate and Richter b value for each fault, and the mean attenuation relationship from magnitude to acceleration. An additional source of uncertainty for sites in the central Mississippi Valley was taken to be the exact geographical location of faults. For each active fault, the method that these authors propose to calculate the distribution of a_{P_0} due to parameter uncertainty is as follows.

First, one uses standard methods of hazard analysis to calculate a_P as a function of P, with all uncertain parameters and relationships fixed to mean values. In particular, the value of a_P for $P = P_0$ is considered to be the mean value of a_{P_0} when parameter uncertainty is included.

One then calculates an "effective magnitude" M_e and an "effective distance" R_e, defined as the mean magnitude and mean distance of events that cause peak site acceleration larger than a_{P_0}, i.e.,

$$\begin{cases} M_e = \sum\limits_{M} M \cdot P[A > a_{P_0} | M] \cdot P[M]/P_0 \\ \\ R_e = \sum\limits_{R} R \cdot P[A > a_{P_0} | R] \cdot P[R]/P_0 \end{cases} \tag{4.1}$$

The method proceeds by calculating sensitivity factors on the peak site acceleration a_{P_0} due to variations of the uncertain parameters. Parameters are varied one at a time, and in each case, one utilizes the quantities in Eq. 4.1 to replace lengthy seismic hazard convolutions with closed-form calculations using point sources and single magnitude values.

For each parameter combination (parameters are assumed to have discrete distribution) one finds a_{P_0} by using the best estimate of a_{P_0} and previously calculated sensitivity factors. One also finds the associated probability as the product of individual parameter probabilities.

The final result is a discrete distribution of a_P. In the case of a single active fault, this approximate distribution was compared with the exact distribution from repeated seismic hazard analysis for each parameter combination and found to be very accurate.

Numerical results using realistic representations of uncertainty are as follows. For the San Francisco Bay area, the coefficient of variation of $a_{0.002}$, $V_{a_{0.002}}$, ranges from less than 0.2 at sites located about 50 km from the main faults to more than 0.4 near these faults. Intermediate coefficients of variation are obtained at sites far away from the earthquake sources. The reason for this peculiar dependence of $V_{a_{0.002}}$ on distance R from the high intensity region is that a large contribution to uncertainty on a_{P_0} comes from uncertainty on the mean attenuation function. Data to which attenuation relationships are fitted are abundant at distances of about 50 km; and for these distances, alternative attenuation functions are nearly the same. For sites with $R \approx 50$ km, more important sources of hazard uncertainty are the activity rate and the Richter b value, whereas close to the faults, another important source of variation of a_{P_0} is the depth of energy release.

In the calculation of $V_{a_{0.002}}$ for the central Mississippi Valley, uncertainty on the actual location of the faults was also modeled. In this case, location of the earthquake sources and depth of energy release were found to be the main sources of hazard uncertainty, especially at sites close to the faults. Away from the epicentral region, uncertainty on attenuation becomes relatively more important. The largest values of $V_{a_{0.002}}$ in excess of 0.4 were found in the regions of higher activity.

4.2 Sensitivity Analyses

Quite likely, methods of error analysis of the semianalytical type described in the previous section or of the type by enumeration will become of common use in the near future. They allow one to represent hazard not in terms of a single frequency-intensity curve, but as a continuous or discrete family of curves, each associated with an exceedance probability. This representation of seismic hazard is being advocated by several authors, e.g. Kaplan (1980) and Kaplan and Garrick (1980), as a way to express confidence in the results of standard procedures of hazard analysis. The same representation corresponds to the no-mixing option of Sec. 2.2; it has the conceptual appeal of avoiding the combination of probabilities with different interpretations and of separating objective hazard from subjective or non-frequentist uncertainty. Its main drawbacks are (1) the nonavailability of practical instruments of computation (the algorithm by McGuire and Shedlock is the first of this kind to be proposed) and (2) the added complication and expense of analysis. Also, as observed by McGuire and Shedlock, the representation of seismic hazard is useful when making decisions about research and data gathering efforts, but it is not clear how one should utilize it when dealing with the safety of given facilities. In the latter context, it seems that only the average hazard counts. If this is so and if, as found by McGuire and Shedlock, the average hazard is nearly identical to the hazard when uncertain parameters are fixed to their mean values, then the seismic safety of individual facilities can be calculated through conventional, one-curve analysis algorithms, ignoring parameter uncertainty.

One may conclude that both types of hazard analysis (with and without mixing frequentist and non-frequentist uncertainty) are useful, but in different circumstances. When faced with a specific problem, one should first ask whether quantified non-frequentist uncertainty about frequentist hazard would be of use. If the answer is no, as it seems to be the case for the safety evaluation of given facilities, then standard procedures with best estimates of the parameters are usually adequate. If, on the other hand, decisions depend on non-frequentist uncertainty about hazard, as it is the case in research resource allocation, then quantification of this uncertainty is also needed.

There are cases of the first type (only the average hazard is relevant) that cannot be solved by simply performing one hazard analysis with parameters fixed to mean values, because the results would be biased or because it is difficult to define or work with mean parameter values. For example, it may be impractical to define a mean random arrival process, given several alternative representations of earthquake times, and it may be impossible to combine the opinions of different experts to obtain a "mean opinion." In this case, expectation of hazard should be taken after each model has been analyzed (possibly in combination with input from each of several experts). Associated with each model-expert combination there is a probability and a conditional hazard curve and the average hazard is obtained by weighing the conditional hazard functions by the associated probabilities.

Sensitivity of the average hazard to modeling alternatives depends on the site under consideration and on the expert whose opinion is being considered. This dependence has been amply documented in a recent study by TERA Corp. (1980) of seismic hazard at sites of the central and eastern United States. The same study shows that sensitivity to reasonable parameter variations depends also on the level of hazard (the return period) being considered and on the measure of site intensity; e.g., it is different for peak ground acceleration (PGA), peak ground velocity (PGV), and for peak spectral response depending on the natural frequency. It is therefore clear that only general qualitative conclusions can be drawn about the influence of parameter variations and errors on seismic hazard. When dealing with specific situations, these conclusions should be taken as guidelines, not as absolute facts.

Some results from the TERA Corp. (1980) study of seismic hazard in the central and eastern United States will be discussed. Nine nuclear central power plant sites were analyzed using input information from ten different experts. Each expert responded to a comprehensive questionnaire on seismic source zonation, seismicity of each identified source, and spatial attenuation of motion intensity, providing not only best estimates of the parameters but also quantified uncertainty on ranges of values. In the analysis, zonation and seismicity parameters (mean rate of occurrence, decay of the magnitude-recurrence law, and upper bound magnitude) were allowed to vary from expert to expert. Three different attenuation models were used and kept the same for all experts. Two of them have the form

$$\ell n(GM) = C_1 + C_2 I_0 + C_3 r + C_4 \ell n\, r \qquad (4.2)$$

in which GM is ground motion intensity at the site (PGA, PGV, or spectral response), the coefficients C_i have the values in Table 4.1 for the attenuation of peak acceleration and velocity, and r is epicentral distance in kilometers. The two sets of coefficients in the table correspond to using different MM intensity attenuation relationships: one,

$$I_S = I_0 + 3.2 - 0.0011r - 2.7 \log_{10} r \,, \quad \begin{matrix} I_S \geq IV \\ r \geq 20 \text{ km} \end{matrix} \qquad (4.3)$$

Intensity Attenuation	GM	Units	C_1	C_2	C_3	C_4
Ossippee	PGA	cm/s^2	1.18	0.56	-0.007	-0.189
	PGV	cm/s	-1.84	0.64	-0.006	-0.167
Gupta-Nuttli	PGA	cm/s^2	2.70	0.56	-0.0007	-0.733
	PGV	cm/s	-0.50	0.64	-0.0006	-0.645

TABLE 4.1 Constants in the attenuation function of Eq. 4.2 for GM = PGA, PGV (after TERA, 1980).

was adapted from Gupta and Nuttli (1976); the other,

$$I_S + I_0 + 0.774 - 0.00117r - 0.302 \ln r \tag{4.4}$$

was obtained from analysis of the Ossippee 1940 earthquake data. A comparison of median attenuation models is shown in Figs. 4.1 and 4.2. The figures include three other attenuation relationships, one of which (denoted August 1979) was used by TERA in a previous evaluation of seismic hazard. The comparison is in terms of sustained acceleration (= 0.7 PGA) for magnitudes 5.5 and 6.5 and uses the relationship $I_o = 2m_b - 3.5$, where m_b = body-wave magnitude. The two more recent models (denoted by "JAN 80" in the figures) intersect around 20 and 200 km; the Ossippee attenuation is more conservative in this intermediate range of epicentral distances and less conservative for $r > 200$ km.

For each (median) attenuation function, attenuation uncertainty was expressed by modeling $\ln(GM)$ as a truncated normal random variable with standard deviation σ. The value of σ was considered to be either 0.7 or 0.9 (values around 0.6 result from direct regression of acceleration data in the west, whereas values about 0.9 are often obtained from the mathematical combination of various sources of uncertainty). Use of the untruncated normal distribution for $\ln(GM)$ is not satisfactory, especially for large median values of GM and large σ; for example, in the case when $\sigma = 0.9$, a $3-\sigma$ departure from the mean of $\ln(GM)$ corresponds to an unrealistically high factor of about 15 on GM. Truncation of the distribution of $\ln(GM)$ was therefore introduced, at 2 or 3 standard deviations from the mean. A model with variable σ was also considered, with a value of 0.6

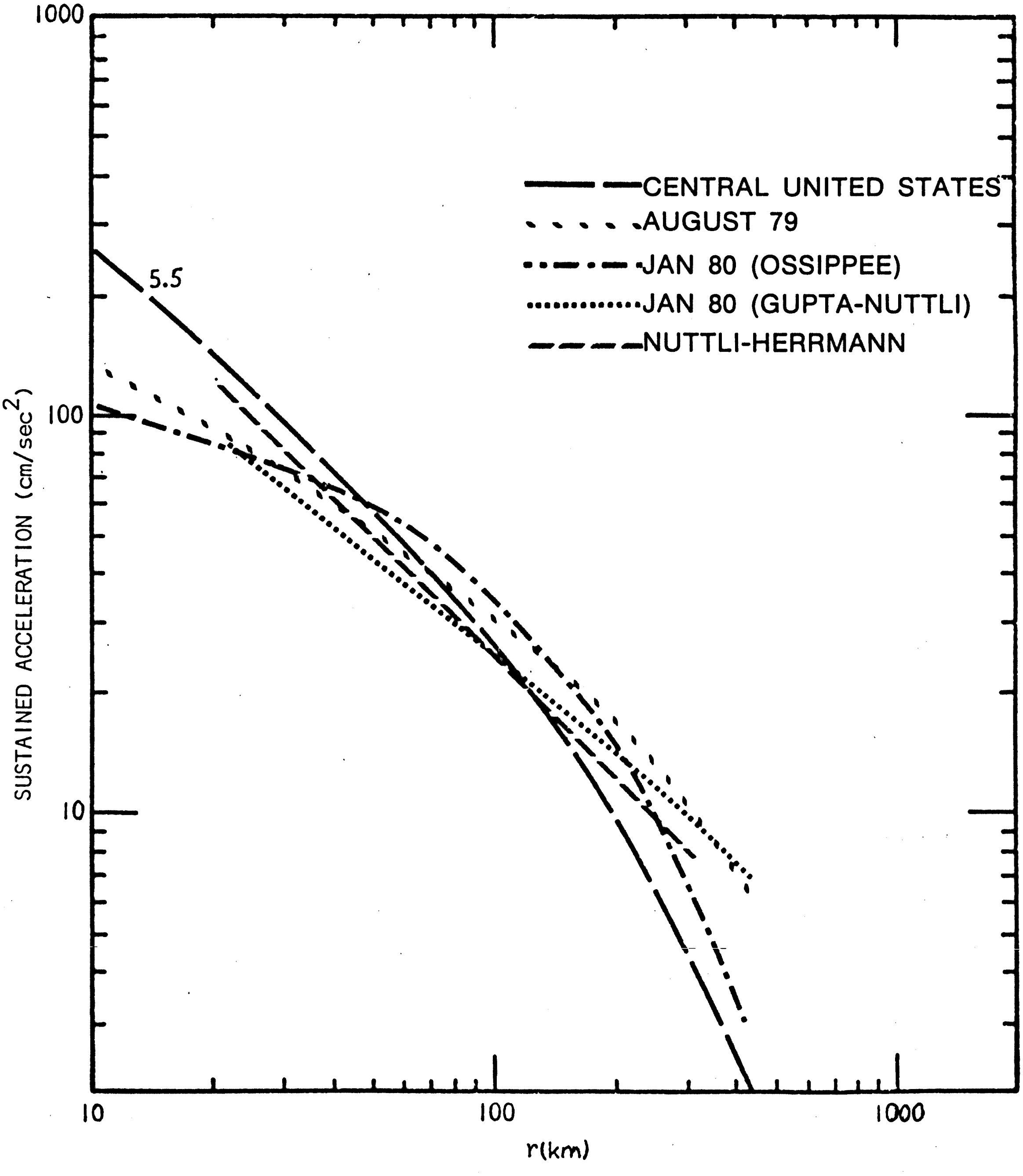


FIGURE 4.1 Comparison of attenuation models for magnitude 5.5 (TERA, 1980)

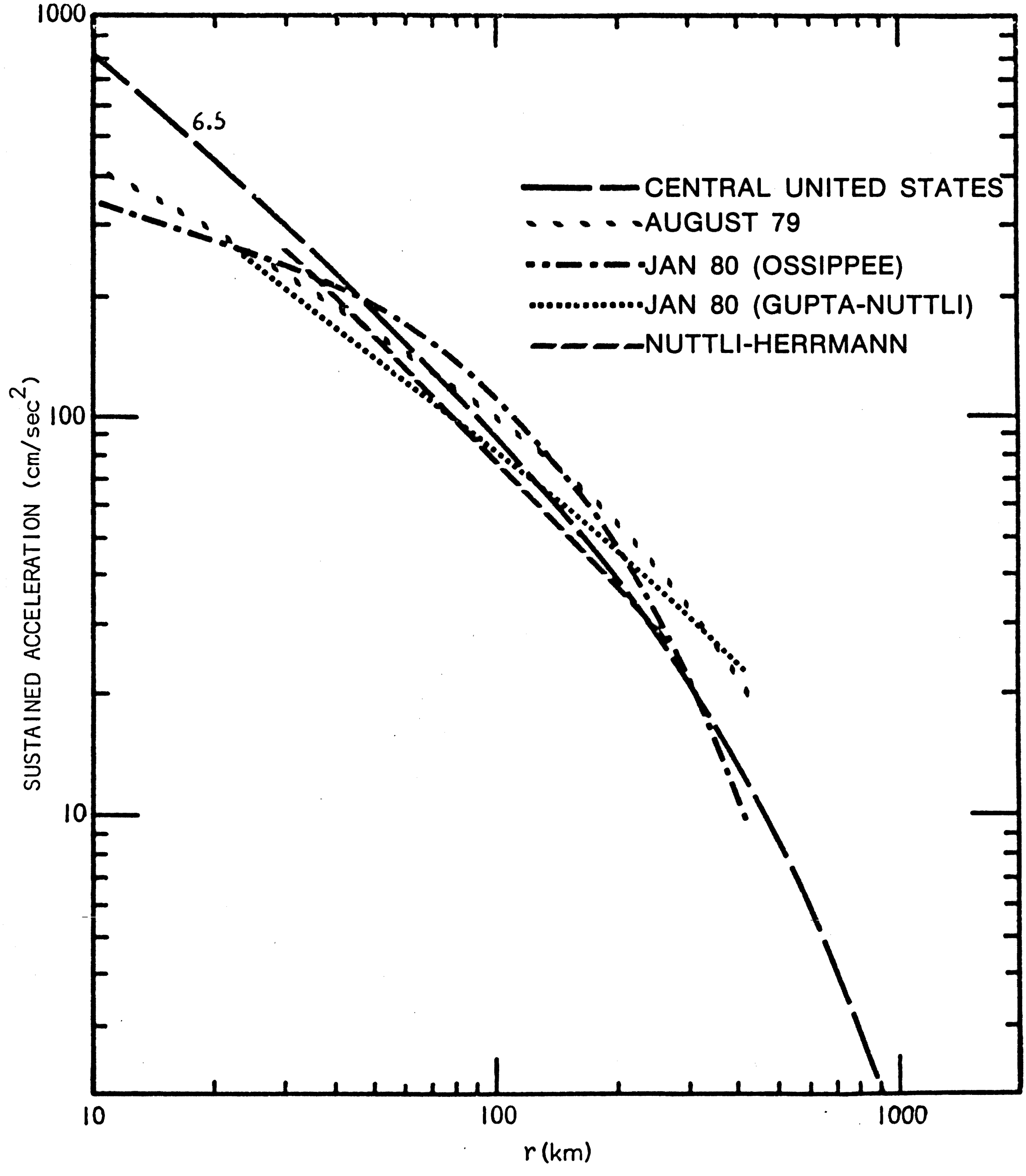


FIGURE 4.2 Comparison of attenuation models for magnitude 6.5 (TERA, 1980)

for $r < 40$ km and 0.9 for larger distances. Unfortunately, there is little physical or statistical basis for selecting the truncation point, which is often a critical parameter.

Sensitivity analyses were conducted with respect to possible interpretations of each expert input and, in all cases, with respect to the median attenuation model and the attenuation error distribution mentioned above. The main results are as follows.

In nearly all cases, sensitivity to upper bound magnitude and zonation alternatives as specified by each expert was found to be small. Sensitivity to the upper bound magnitude M_U is a function of the slope of the magnitude-frequency relationship and of the return period. If the return period is short (e.g., 1000 years or less) and the slope is steep, then M_U is usually unimportant. For all sites located in the Central Stable Region, sensitivity to zonation alternatives was found to be negligible because the main contributing sources are the host region with rather well-defined boundaries, and the New Madrid region, which is too distant to reflect minor variations in its boundaries. Higher sensitivity was noticed at northeastern sites where zonation is more complex and zonation uncertainty is larger. However, also in this case, the variation of hazard when using zonations from <u>different</u> experts exceeds the variation associated with zonation uncertainty of individual experts.

The median attenuation model and the distribution of attenuation error were found to be the most critical parameters, especially for PGV and long-period spectral ordinates. Sensitivity of PGA and PGV to the median attenuation at four sites in the central United States and five sites in the eastern United States are summarized in Tables 4.2 and 4.3, respectively.

CENTRAL U.S. SITES

SUMMARY OF ATTENUATION RESULTS
(Variation in percent at 1,000-year return period)

	PGA		PGV	
From (1) to (2)	BRP, La Crosse	no change	BRP	30% decrease
	Palisades	10% decrease	La Crosse	35% decrease
	Dresden	15% decrease	Palisades	50% decrease
			Dresden	60% decrease
From (2) to (3)	BRP	5% decrease	BRP	5% increase
	La Crosse	no change	La Crosse	25% increase
	Palisades	10% increase	Palisades	35% increase
	Dresden	15% increase	Dresden	45% increase
From (1) to (3)	BRP	5% decrease	BRP, La Crosse	20% decrease
	La Crosse, Palisades	no change	Palisades	35% decrease
	Dresden	5% increase	Dresden	45% decrease

(1) August attenuation
(2) Ossippee
(3) Gupta-Nuttli

TABLE 4.2 Sensitivity of PAG and PGV to the median attenuation function at four sites in the central United States (TERA, 1980).

EASTERN U.S. SITES

SUMMARY OF ATTENUATION RESULTS

(Variation in percent at 1,000-year return period)

	PGA	PGV	
From (1) to (2)	no change	Oyster Creek	35% decrease
		Millstone	35% decrease
		Connecticut Yankee	35% decrease
		Yankee Rowe	45% decrease
		Ginna	50% decrease
From (2) to (3)*	15% decrease	10 to 15% decrease, although E_8 and E_{10} show increases from 0% to 20%	

(1) August attenuation
(2) Ossippee
(3) Gupta-Nuttli

* These variations are estimates as the Gupta-Nuttli attenuation was run for three experts only: E_3, E_8 and E_{10}.

TABLE 4.3 Sensitivity of PGA and PGV to the median attenuation function at five sites in the eastern United States (TERA, 1980).

Sensitivity varies from site to site depending on its relative location with respect to the main sources; it is highest for PGV and for sites where hazard is largely influenced by distance sources.

At sites in the east, sensitivity of PGA to the median attenuation is smaller than in the central United States because of absence of important distant sources (the St. Lawrence source is too far to give appreciable contributions). However, sensitivity of PGV remains high.

The distribution of the attenuation error is also critical, especially for high intensities (long return periods). For a return period of 1000 years, changing the truncation point from two to three standard deviations increases the PGA from 20 to 35 percent if $\sigma = 0.9$ and from 15 to 25 percent if $\sigma = 0.7$. For PGV, these quantities are respectively 8 to 15 percent and 6 to 12 percent. Higher sensitivity, up to 45 percent, is found for a return period of 4000 years. Reducing the attenuation error variance at short distances (using $\sigma = 0.6$ for $r < 40$ km and $\sigma = 0.9$ for $r > 40$ km, as opposed to $\sigma = 0.9$ for all r) produces very minor changes in the hazard, except for the few sites where hazard depends on near sources.

As one would expect, the importance of far-away sources increases with increasing period (from PGA to PGV) and the importance of sources with higher upper bound magnitude increases with decreasing hazard. Also, the relative importance of different sources depends, in some cases rather importantly, on the form of the attenuation function.

Average hazard and sensitivity results for firm-soil sites in Boston have been obtained by Cornell and Merz (1975). The zonation for the base case is shown in Fig. 4.3 and best estimates of the seismicity parameters for each source are given in Table 4.4. The attenuation law, in terms of

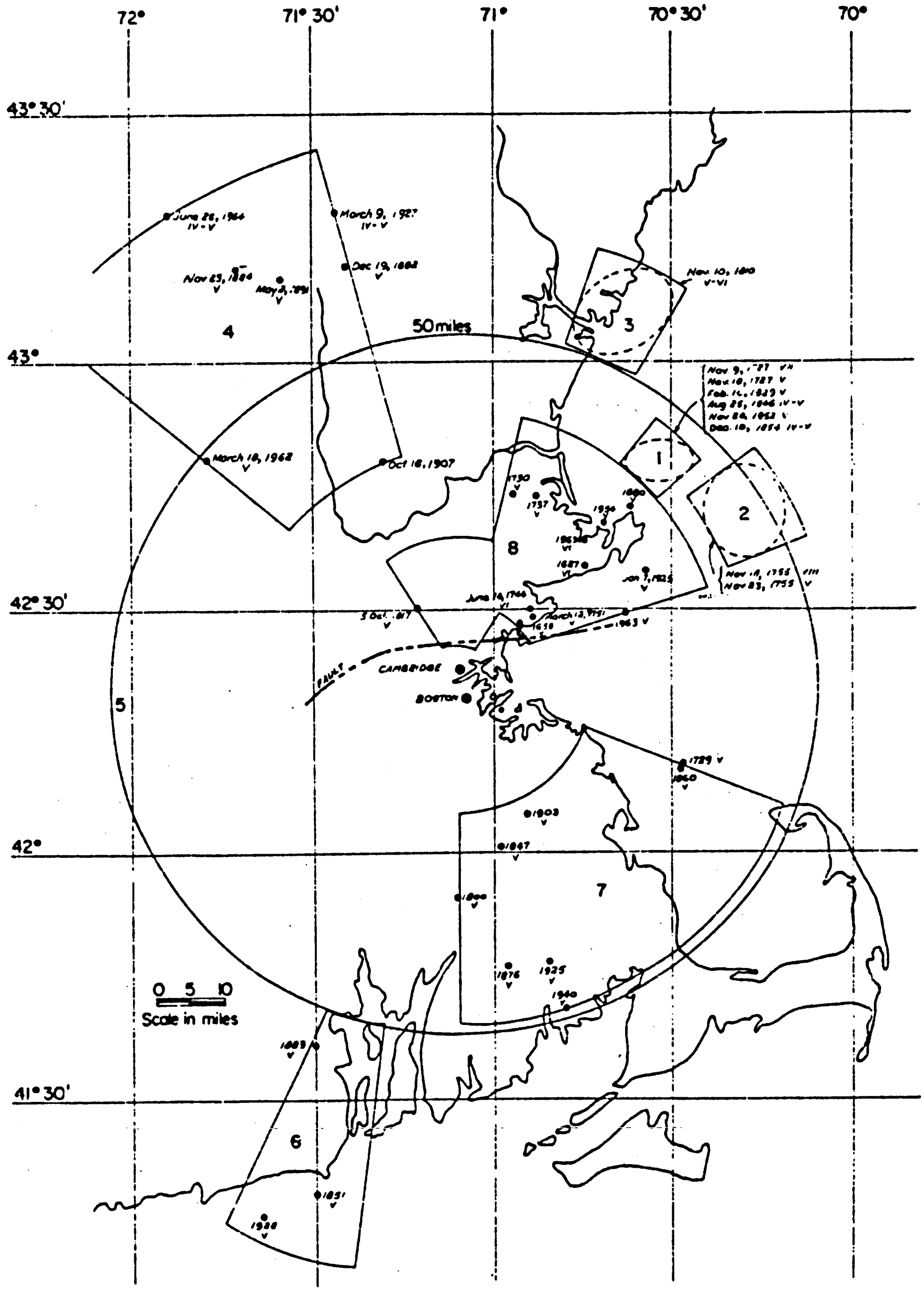


FIGURE 4.3 Zonation of the region around Boston for base-case hazard calculation (Cornell and Merz, 1975).

SOURCE	No. of events with MMI $\geq$ 4.25/year/10,000 km^2	Slope	Upper bound intensity
1	3.16	0.48	7.7
2	0.42	0.48	8.7
3	0.180	0.48	6.7
4	0.150	0.48	7.3
5	0.010	0.48	6.3
6	0.215	0.48	6.7
7	0.200	0.48	6.7
8	0.380	0.48	7.3

TABLE 4.4 Best estimates of source parameters for hazard analysis of Boston (Cornell and Merz, 1975).

Modified Mercalli intensity, was estimated from the isoseismals of large historical earthquakes in the region; it has the form

$$I_S = \begin{cases} I_0 , & r < 10 \text{ miles} \\ 2.6 + I_0 - 1.3 \ln r, & r \geq 10 \text{ miles} \end{cases} \tag{4.5}$$

Attenuation intensity is expressed by an additive normal random variable with mean 0 and standard deviation 0.2. The main sensitivity cases considered by Cornell and Merz correspond to the following variants of the base case:

1. I_S increased or decreased by 1/2 unit;
2. Source 8 replaced with a more conservative or a less conservative source.
3. The upper bound epicentral intensity of all sources increased or decreased by 1/2 unit.
4. The upper bound epicentral intensity for sources 1 and 2 (off Cape Ann) increased to XII.

Hazard curves corresponding to these and other parameter perturbations are shown in Fig. 3.1. Decreasing site intensity by 1/2 unit in the attenuation equation (curve AL) produces appreciable changes only for short return periods. For long return periods the hazard remains the same because the contributing events are primarily those very close to the site, within a 10-mile radius. Conservative modification of Source 8 (curve GP) produces similar qualitative results, with noticeable effects only for short return periods. More important sensitivity is found with respect to the upper bound intensities. If these bounds are increased by 1/2 unit (curve UBP), then the hazard curve moves to the right by an amount of up to 0.4 for long return periods. Larger increases of intensity over long return periods are observed if the Cape Ann sources are assigned upper bound intensity XII (curve UB12).

For calculation of the composite (average) hazard curve, probabilities were subjectively assigned to discrete alternatives about the parameters of the attenuation function, the geometry of Source 8, and the upper bound intensities. These probabilities are given in Table 4.5. They correspond to 36 different cases. The resulting composite hazard curve is compared in Fig. 4.4 with the curve for parameters fixed to most likely values. The very significant divergence for return periods larger than 10,000 years is due to the possibility that the maximum intensity is XII for the sources near Cape Ann.

Additional sensitivity analyses for Boston have been conducted by TERA (1980) with respect to expert input, median attenuation model (Eq. 4.4 vs. Eq. 4.5), and dispersion of the attenuation error (σ = 1.0 in place of σ = 0.2). The results are summarized in Table 4.6 for return periods of 1,000 and 10,000 years. As shown in Fig. 4.5, the Gupta-Nuttli attenuation model produces site-intensity values about 1/2 unit higher than the Cornell-Merz attenuation. The effect on hazard is a variation of about 0.3 units of intensity. More sensitive parameters are the expert opinion and the dispersion of the attenuation error. Although the expert in the TERA analysis basically agreed with the zonation by Cornell and Merz, he assigned different values to upper bound intensities and occurrence rates, with a net effect of decreasing the hazard by approximately 0.9 units of intensity. An even more critical choice is that of σ which, when varied from 0.2 to 1.0,

	LOW	BEST ESTIMATE	HIGH	
ATTENUATION LAW	0.25	0.50	0.25	
ZONATION	0.25	0.50	0.25	MMI XII for Cape Ann
MMI UPPER BOUNDS	0.20	0.30	0.20	0.30

TABLE 4.5 Subjective probabilities for the calculation of the composite hazard curve (Cornell and Merz, 1975).

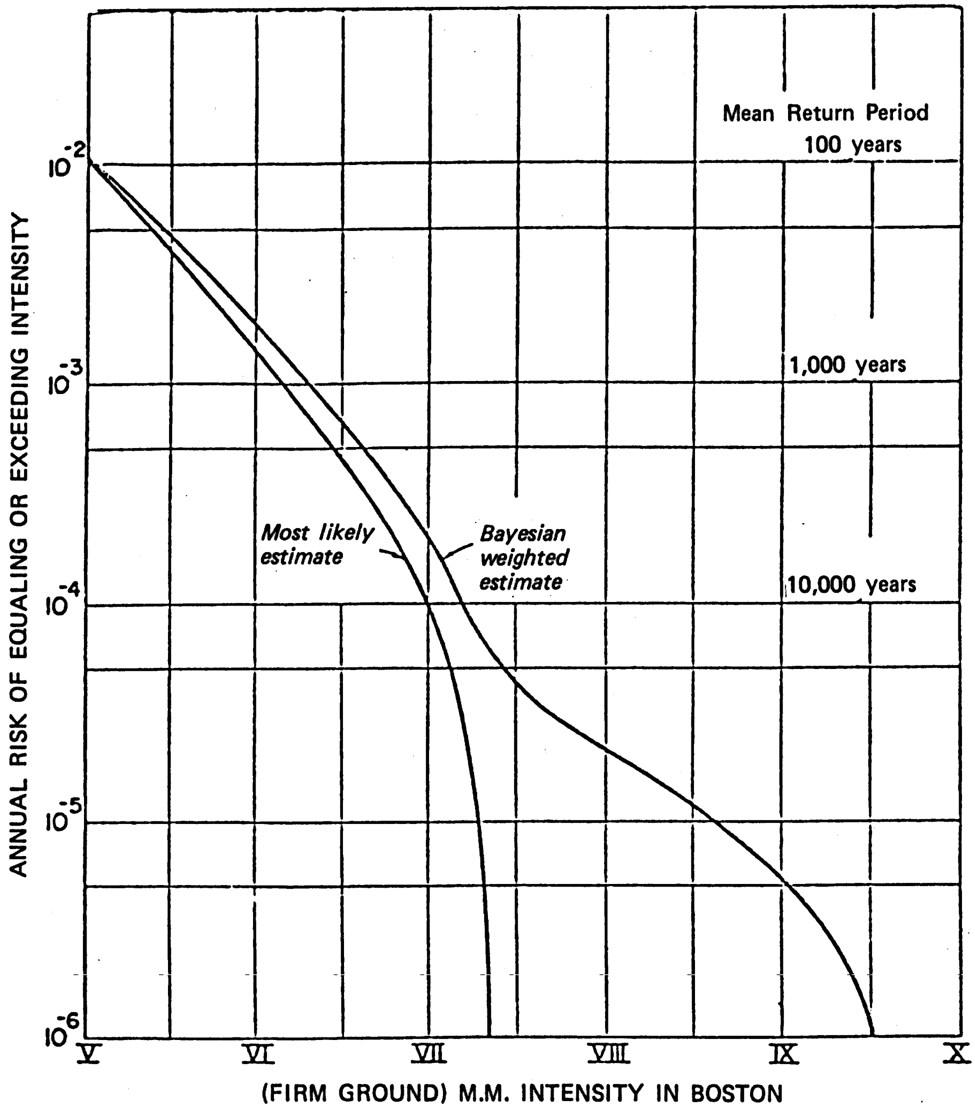


FIGURE 4.4 Most likely and composite hazard curves for Boston (Cornell and Merz, 1975).

INTENSITY COMPARISONS
(INTENSITY IN MMI UNITS)

	C-M Analysis	TERA ANALYSIS			
	C-M BEST ESTIMATES $\sigma_I = 0.20$	C-M DATA C-M ATTENUATION $\sigma_I = 0.20$	EXPERT 8 DATA C-M ATTENUATION $\sigma_I = 0.20$	EXPERT 8 DATA C-M ATTENUATION $\sigma_I = 1.0$	EXPERT 8 DATA GUPTA ATTENUATION $\sigma_I = 1.0$
Return Period 1,000 Years	6.20	6.30	5.45	6.55	6.90
Return Period 10,000 Years	7.00	6.80	5.90	7.35	7.65

TABLE 4.6 Sensitivity analyses in the TERA (1980) report for hazard in Boston "C-M" refers to Cornell and Merz (1975).

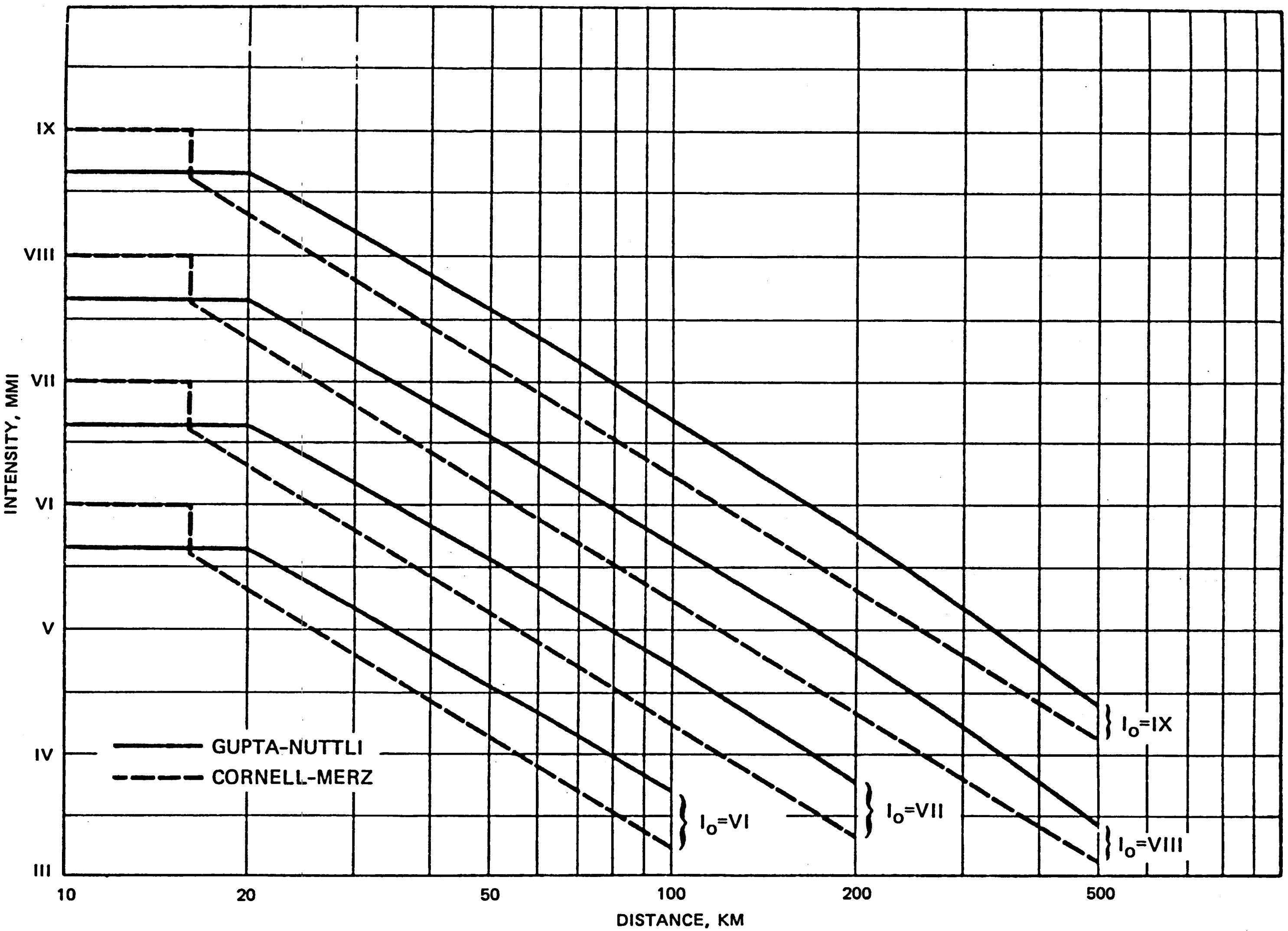


FIGURE 4.5 Comparison of median attenuation relationships for Boston (TERA, 1980).

produces an increase of site intensity of 1.1 and 1.45 units, for return periods of 1,000 and 10,000 years, respectively.

Sensitivity results for the La Crosse site (Fig. 4.6) from Dames and Moore (1980) and TERA (1980) are given next. Sensitivity involves zonation, recurrence parameters, and median attenuation. In all cases, the attenuation error on ℓn(PGA) is modeled as an untruncated normal variable with standard deviation σ = 0.6. Two alternative zonations are considered: the Central Stable Region only, or the Central Stable Region and the Wisconsin Arch source shown in Fig. 4.6. In each case, sensitivity is evaluated also with respect to inclusion or exclusion of the New Madrid source. The median attenuation functions, based on the Ossippee and Gupta-Nuttli models, are the same as described earlier in this section. Table 4.7 gives the seismicity parameters, as assigned by Dames & Moore (D&M) and by TERA (E_0). Especially noticeable in the table is the difference in mean activity rate (which is much higher in the TERA report) and in the upper bound magnitude (the three values in the last column correspond to the best estimate and to lower and upper bounds). The difference in the best estimate is about 1/2 magnitude unit for all sources.

Hazard results in Table 4.8 show that the effects on PGA of changing the mean rate a and the upper bound magnitude M_U are about 15 and 10 percent, respectively. Sensitivity to the inclusion or exclusion of the New Madrid source is entirely dependent on the attenuation model: it is negligible for Ossippee and about 25 percent for Gupta-Nuttli. Finally, the effect of including or excluding the Wisconsin Arch source is not large but about the same for both attenuation functions.

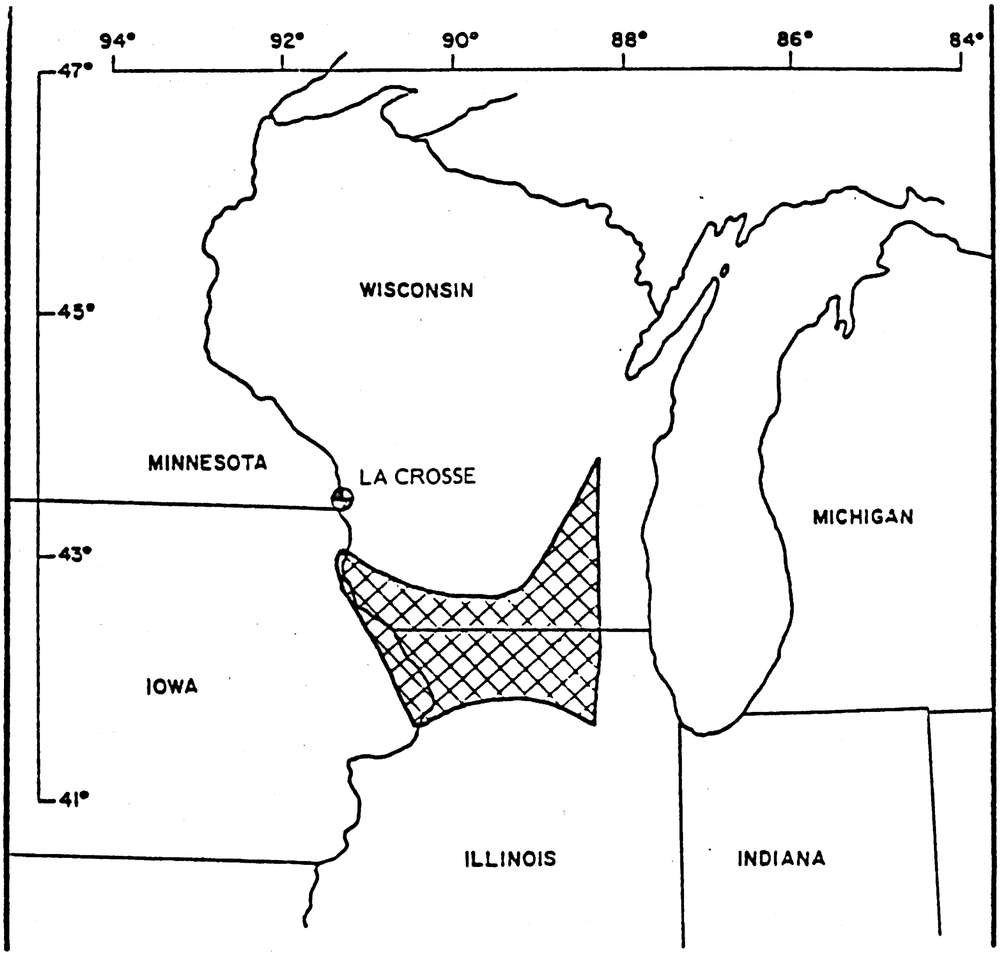


FIGURE 4.6 La Crosse site and the Wisconsin Arch source (Dames and Moore, 1980).

SOURCE PARAMETERS

ZONATION	SEISMIC SOURCE(S)	ACTIVITY RATE: D & M EVENTS IN 175 YEARS WITH $m_b \geq 4.6$	ACTIVITY RATE: E_9 EVENTS IN 175 YEARS WITH $m_b \geq 4.6$	ACTIVITY RATE: E_9 EVENTS IN 175 YEARS WITH $m_b \geq 4.0$	b-VALUE	UPPER MAGNITUDE: D & M	UPPER MAGNITUDE: E_9 *
1	Central Stable Region	10.6	20.9	102.6	0.92	5.8	5.5-6.1-6.2
2	Central Stable Region	8.1	16.6	80.7	0.92	5.0	5.0-5.3-5.5
	Nothern Illinois (Wisconsin Arch)	3.2	4.3	20.9	0.92	5.8	5.5-6.1-6.2

* Upper magnitude bounds and best estimate.

TABLE 4.7 Seismicity parameters for La Crosse according to Dames and Moore (1980) and TERA (1980).

PGA (g) AT LA CROSSE FOR
1000 AND 10,000 YEAR RETURN PERIOD

Ground Motion Model	Central Stable Region without New Madrid		Central Stable Region and New Madrid							
	D&M a value D&M M_u		D&M a value D&M M_u		E_9 a value D&M M_u		E_9 a value E_9 M_u		E_9 a value* E_9 M_u	
	RP 1	RP 2	RP 1	RP 2	RP 1	RP 2	RP 1	RP 2	RP 1	RP 2
Ossippee	0.061	0.15	0.061	0.15	0.081	0.18	0.085	0.20	0.085	0.20
Gupta-Nuttli	0.052	0.12	0.065	0.14	0.075	0.15	0.076	0.17	0.079	0.17
	New Madrid Excluded		Central Stable Region and Northern Illinois (E_9) and New Madrid							
	D&M a value D&M M_u		D&M a value D&M M_u		E_9 a value D&M M_u		E_9 a value E_9 M_u		E_9 a value* E_9 M_u	
	RP 1	RP 2	RP 1	RP 2	RP 1	RP 2	RP 1	RP 2	RP 1	RP 2
Ossippee	0.058	0.12	0.058	0.12	0.068	0.14	0.084	0.17	0.085	0.17
Gupta-Nuttli	0.049	0.096	0.062	0.12	0.066	0.13	0.074	0.15	0.075	0.15

* Considering m_b=4.0 and greater

RP 1 = 1,000 year return period
RP 2 = 10,000 year return period

TABLE 4.8 Sensitivity results for La Crosse site (TERA, 1980).

All parameter sensitivity results up to here refer to essentially identical seismic hazard models. If the model is changed, then not only the numerical value of estimated hazard will change (see Sec. 4.3) but also sensitivity to parameter variations may be different. One example is the line source model of Der Kiureghian and Ang (1975). The main departure from the standard model is that earthquakes are now associated with fault ruptures of finite dimension and site effects are related to the (minimum) distance to the rupture zone as opposed to the epicentral or focal distance. As a consequence of this modification, calculated hazard increases and parameter sensitivity changes. With regard to the latter effect, Der Kiureghian and Ang challenge the conclusions by Cornell and Vanmarcke (1969) based on point-source models, that

1. The major contribution to hazard comes from frequent, small-size earthquakes from sources close to the site.
2. Hazard is seldom sensitive to the upper tail of the magnitude distribution or to the upper bound magnitude.

From the line-source model, Der Kiureghian and Ang find that

1. Distant sources may have significant effects on seismic hazard when large-magnitude earthquakes are possible.
2. Seismic hazard is sensitive to the upper bound magnitude M_U, particularly for $M_U \leq 8.5$.

These conclusions are especially significant for long return periods. The main reason why point and line source models have different sensitivity to high magnitude earthquakes is that these events are usually associated with long rupture zones, which may propagate close to the site and produce effects much larger than expected from epicentral distance.

For practical applications, Der Kiureghian and Ang conclude that the choice of M_U is immaterial at low intensity levels, that in all cases events of extremely large size (magnitude larger than 9) give negligible contribution to hazard, and that the choice of M_U is critical only for long return periods and when contributing sources exist far from the site with M_U below 8.5. They also conclude, somewhat contrary to findings in the TERA study, that sensitivity to M_U is not much affected by the value of the slope parameter b.

Fig. 4.7 shows a comparison of the point and line source models and some sensitivity results for the latter model. All curves in the figure estimate hazard in San Francisco and neglect attenuation uncertainty. With reference to the line-source model, curve 1 represents the base case evaluation; curve 2 results from neglecting events with magnitude below 4.5; and curve 3 corresponds to increasing the slope of the magnitude recurrence law from 1.22 to 1.30. The remaining two curves account for incompleteness of the data when fitting the magnitude recurrence relationship (curve 4) and reinterpret the rate of activity along the San Andreas fault (curve 5). The last curve may be considered to quantify sensitivity to zonation. In all cases, the upper bound magnitude was fixed to 8.5. Sensitivity is clearly small, especially when compared with the effect of including or excluding attenuation uncertainty. The latter effect and the importance of the choice of the median attenuation curve is demonstrated in Fig. 4.8 for which the three attenuation models of Table 4.9 have been used, with and without uncertainty. It is interesting to note that models 2 and 3, which are substantially different in their median predictions, produce nearly identical results when attenuation uncertainty is included, whereas the opposite is true for models 1 and 2.

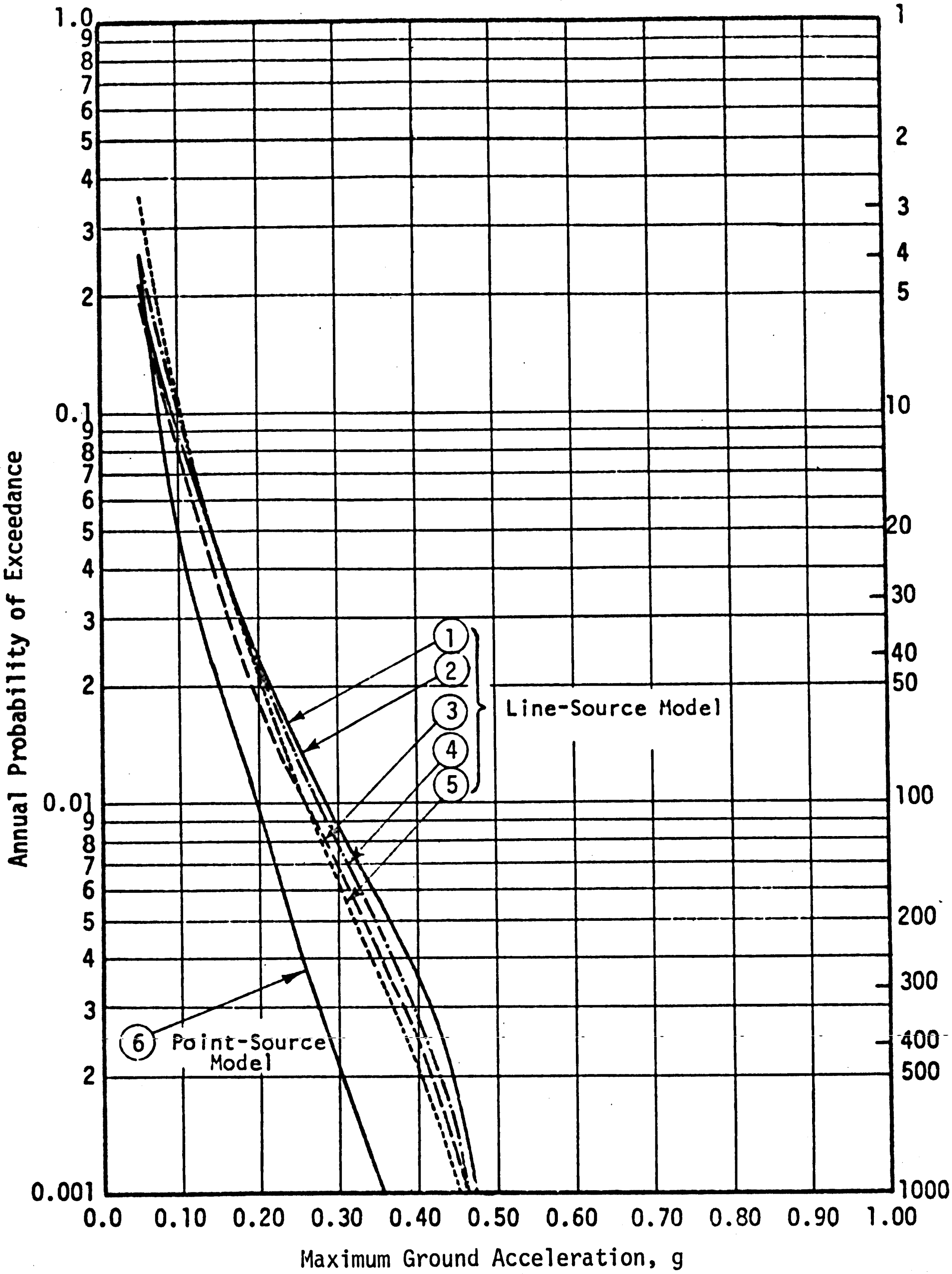


FIGURE 4.7 Comparison of point and line source models and sensitivity analysis for San Francisco (Der Kiureghian and Ang, 1975).

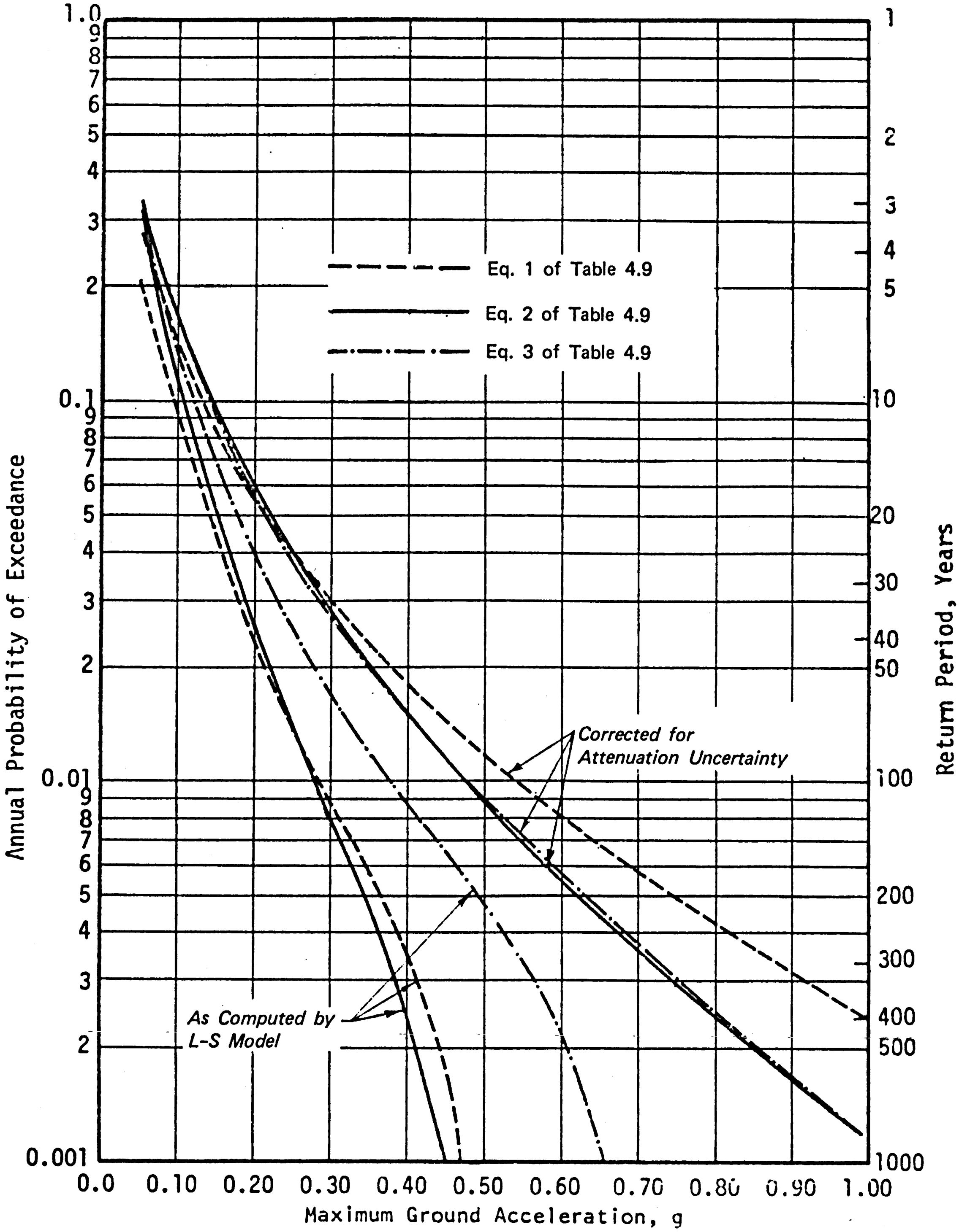


FIGURE 4.8 Sensitivity of the line-source model to median attenuation and to attenuation uncertainty. Hazard estimates are for San Francisco (Der Kiureghian and Ang, 1975).

Eq.	MEDIAN ATTENUATION	σ
1	$a = 1320e^{0.58m}\ (r+25)^{-1.52}$	0.894
2	$a = 1080e^{0.50m}\ (r+25)^{-1.32}$	0.755
3	$a = 472.3e^{0.64m}\ (r+25)^{-1.301}$	0.573

TABLE 4.9 Attenuation models used in the sensitivity analysis of Fig. 4.8.

The exact determination of source boundaries is usually not of critical importance for "area" sources, but there are notable exceptions. One exception is when different interpretations are possible for the mechanism of earthquake generation. For example, it has been suggested by Long (1976) and McKeown (1978) that high intensity seismicity in some regions of the eastern United States might be due to (or might be statistically associated with) the presence of plutons. More frequent situations in which source geometry plays an important role are when it is uncertain whether a major fault is active or not. A study summarized in Donovan and Bornstein (1978) shows that the change of hypothesis about activity or inactivity of a fault in the Los Angeles areas leads to error factors of up to 50 percent on site acceleration for a given exceedance probability. This should be regarded as a rather extreme case. Sensitivity of hazard to upper bound magnitude depends on the value of the maximum magnitude and on the location of the site relative to the source, but it is now recognized that this parameter is not as critical as previously believed (see, for example, Cornell, 1980; McGuire and Shedlock , 1980; Donovan and Bornstein, 1978).

The diversity of conclusions from sensitivity studies in this section demonstrates the impossibility of giving absolute rules about the influence of parameter uncertainty on hazard. At most, one can identify sources of uncertainty that tend to dominate the results; among these sources are the form of the attenuation function and the distribution of the attenuation error. Expert opinion should also be included as a major source of hazard variation, as will be said in Sec. 6.

4.3 Alternative Models

Alongside the development of what we now regard as standard seismic hazard procedures, many alternative modeling hypotheses have been explored.

Much work has been done on the representation of earthquake times (sometimes also locations) through random point processes other than Poisson. The motivating idea is that earthquakes produce changes in the state of the physical system and must therefore affect the occurrence of other seismic events at close times and locations. Perhaps the simplest way to introduce dependence of earthquakes in time is through so-called renewal and clustering models.

In a renewal point process, interarrival times T_1, T_2, ... are independent and identical variables as in the Poisson model, but not necessarily with exponential distribution. Therefore, memory extends only between consecutive events. Kameda and Ozaki (1979) fitted a simple renewal model to historical data from Kyoto. They found that these data are compatible with a piecewise constant hazard function $\nu(t) = f_T(t)/[1-F_T(t)]$, with only two values: ν_1 for small t and $\nu_2 > \nu_1$ for large t, and compared hazard predictions from this model with those when $\nu(t) \equiv \nu$ (Poisson events). After observing that in the Kyoto catalogue, longer interarrival times usually terminate in larger earthquakes, Kameda and Ozaki generalized the model to account for this dependence and again compared results with the standard assumptions of Poisson arrivals and independent sizes. The difference in hazard is sometimes large, depending on the interval of prediction and on the time of occurrence of the last earthquake. One set of results is shown in Fig. 4.9. Each curve in the figure gives the

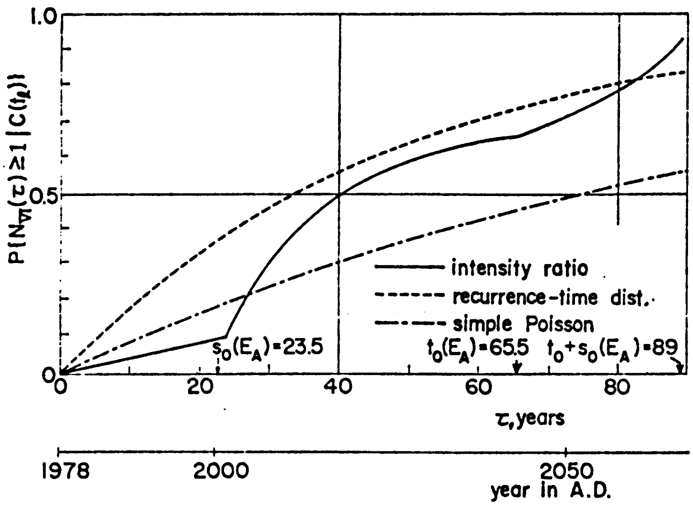


FIGURE 4.9 Probability of future earthquakes of JMA intensity VI or greater in Kyoto, starting from year 1978. Poisson and renewal models. (Kameda and Ozaki, 1979).

probability of at least one earthquake in Kyoto with JMA (Japan Meterological Agency) intensity VI or greater (approximately MM intensity IX or greater) as a function of time, starting from the year 1978. The dashed-dotted line corresponds to the hypothesis of Poisson events with independent sizes. The other two lines are for arrivals according to the renewal process: the dashed line continues to assume independent sizes, whereas the solid line incorporates dependence of size on interarrival time. In the latter case, the distribution of intensity has either of two forms, depending on whether the earthquake occurs when $\nu = \nu_1$ or $\nu = \nu_2$. It is interesting to observe that in the present case and for short to moderate time intervals, dependence of size on interarrival time counteracts the effect on hazard of non-Poisson arrivals. A different renewal process of earthquake times was proposed earlier by Esteva (1976).

Clustering models have been developed more extensively (see for example, Vere-Jones and Davies, 1966; Walley, 1976; Schlien and Toksoz, 1970; Knopoff and Kagan, 1977). Each cluster comprises one "main shock" and the associated family of "aftershocks." Clusters are typically assumed to occur at Poisson times (and Poisson locations in space), in spite of some evidence of secondary "clustering of clusters" over long times and distances (Walley, 1976). Clustering has of course the effect of increasing hazard at sites and times close to sites and times of past earthquake occurrences. However, it is found that if the main event is also the most intense event of the cluster and aftershocks are confined to a small region near the epicenter of the parent earthquake, then the hazard consequence of excluding aftershocks is small (Merz and Cornell, 1973).

An interesting extension of the concept of uncertain seismicity rate λ has been explored by Walley (1976), who considers as random the intensity function $\lambda(X_1, X_2, t)$ of the Poisson process of epicentral coordinates X_1 and X_2 and times t. This assumption transforms the Poisson process into a so-called doubly-stochastic Poisson process. Realizations of doubly-stochastic Poisson processes are very similar to realizations of clustering models. The effect of seismic hazard is also similar. In an application to New Zealand, the same author finds that hazard is sensitive to the depth of energy release (a similar conclusion for sites near major faults is reached by McGuire and Shedlock, 1980) and cautions that significant errors may come from the assumption of stationary seismicity.

On the basis that major earthquakes release large amounts of strain energy and that such energy gradually accumulates in time, one would expect that the interarrival time be related to the magnitude of the previous and the next earthquake. Convenient mathematical representations of this dependence have been given in terms of Markov and semi-Markov processes (Patwardhan *et al*., 1978, Kiremidjian *et al*., 1980). Hazard predictions are sometimes very different from those using the Poisson model, depending on the time and magnitude of the last major earthquake.

As remarked previously in Sec. 4.2, hazard is usually sensitive to the attenuation function, and even more to the variance and upper tail distribution of the attenuation error. The variance depends not only on the data set and on which quantities are used to characterize source and site intensity, but also on the method by which the data are analyzed. Multistep procedures which utilize prediction of intermediate parameters lead to particularly large estimates of dispersion (Cornell *et al*., 1979).

Errors on hazard at New Zealand sites from neglecting azimuth-dependence of the attenuation are reported in Walley (1976). He finds that when an isotropic model is replaced with a set of highly eccentric elliptical iso-seismals, hazard at some sites varies by 0.6 to 0.8 units of MM intensity.

Of course, the effect of anisotropy is less pronounced if seismic activity is uniform over large regions.

Sensitivity of hazard to including or neglecting the <u>spatial dimension of fault rupture</u> (point versus line source model) has already been reported in Sec. 4.2; see for example Fig. 4.7.

5. ERRORS FROM MODELING ASSUMPTIONS

As said earlier in Sec. 1.2, the correct but unknown model M_T does not necessarily belong to the family of models $\mathcal{M}$. The latter comprises only simple models to which, for convenience of analysis, M_T is assumed to belong. For example, all models in $\mathcal{M}$ might assume isotropic attenuation on the geographical plane, whereas actual attenuation may be azimuth-dependent due to oriented geological structures or to the directional propagation of fracture. One should expect that if $M_T \notin \mathcal{M}$, bias will result from assuming the contrary. If bias is small, then the seismic hazard procedure is said to be "robust." Robustness is a desirable property because M_T is unknown and certainly not of the elementary type hypothesized for analysis. In what follows, a true model M_T is assumed and a few common estimators of seismic hazard, $\hat{H}$, are evaluated against the associated true hazard, H_T. Special attention is given to the conditional bias $(H_T - \underset{Z|M_T}{E}[\hat{H}])$ in which Z is statistical sampling information. In reality, the analyst is ignorant of M_T and H_T, his only statistical information being a historical sample Z, e.g., of ground accelerations at the site. The advantage of knowing H_T is that it will be possible here to evaluate bias and error variance of various seismic hazard procedures which the analyst might choose to use. Contrary to findings by McGuire and Shedlock (1980) for the case when M_T belongs to $\mathcal{M}$ (see Sec. 4.1), error bias, not variance, is found to be the primary source of mean square error of the hazard estimators, especially at high levels of intensity. Material in this section is for the most part from Schumacher (1977) and from unpublished work by Veneziano et al. (1977).

5.1 True Model and Hazard Estimators

Suppose that the true hazard at the site, H_T, corresponds to the following probabilistic representation of regional seismicity, M_T:

a. There is only one source, in the form of a disc with the center at the site and radius r. Earthquakes occur as Poisson points in time, with a mean annual rate of $\nu_t = \nu\pi r^2$ while in space; epicenters are uniformly and independently distributed inside the source.

b. Earthquake magnitudes M_1, M_2, ... are independent random variables with identical, doubly-truncated exponential distribution

$$F_M(m) = K_1[1 - e^{-b(m-m_0)}] , \quad m_0 \leq m \leq m_1 \tag{5.1}$$

in which m_0 and m_1 are lower and upper bounds to M, b is a positive constant, and $K_1 = \{1 - \exp[-b(m_1-m_0]]\}^{-1}$ is a normalization constant.

c. Peak acceleration at the site, a, is the only intensity parameter of interest. It is related to magnitude M and epicentral distance R as (Esteva, 1970; Cornell, 1971; Der Kiureghian and Ang, 1977)

$$a = b_1 \, e^{b_2 M} \, (R+b_4)^{-b_3} \, \varepsilon \tag{5.2}$$

in which the b_i are constant parameters and ε is a lognormal random variable with independent realizations for different earthquakes.

d. ~~Hazard is measured in terms of the probability G(a, t) that ground~~ acceleration at the site will exceed a, at least once in t years.

Under these assumptions, the correct mathematical representation of hazard, $H_T = G_T(a, t)$, is given by

$$H_T = G_T(a, t) = e^{-\nu_t t G_a(a)} \tag{5.3}$$

in which $G_a(a)$ is the correct complementary CDF of peak ground acceleration at the site due to an event of random magnitude and epicentral location. The function $G_a(a)$ has not simple analytical form, but it can be calculated numerically, e.g., by the method of Cornell and Merz (1975).

Given a finite record of peak ground accelerations at the site, the quantities ν_t and $G_a(a)$ in Eq. 5.3 cannot be calculated exactly. For values of a of practical interest, the error in estimating $G_T(a, t)$ is contributed primarily by errors in the estimation of $G_a(a)$. Therefore, it will be assumed here that ν_t is known and inference will be limited to $G_a(a)$. Moreover, one may limit consideration to acceleration values above a threshold a_{min} of engineering interest and estimate the conditional complementary CDF,

$$G_{a|a>a_{min}}(a) = G_a(a)\ G_a(a_{min}) \tag{5.4}$$

for $a > a_{min}$. This last function is plotted in Fig. 5.1 (dashed lines) for the following parameter values:

$$\begin{cases} a_{min} = 10 \text{ (cm/sec}^2\text{)} \\ r = 100 \text{ km} \\ m_0 = 4,\ m_1 = 8,\ b = 2 \\ b_1 = 463.2,\ b_2 = 0.64,\ b_3 = 1.301,\ b_4 = 25 \\ \ell n\ \varepsilon \sim N(0, \sigma^2_{\ell n\varepsilon})\ ,\ \sigma_{\ell n\varepsilon} = 0,\ 0.6 \end{cases} \tag{5.5}$$

The values of b_1, b_2, b_3, and b_4 are from Esteva (1976). It is further noticed that inference of the complementary CDF of a, given that $a > a_{min}$, is equivalent to inference of the complementary CDF of any of the following random variables:

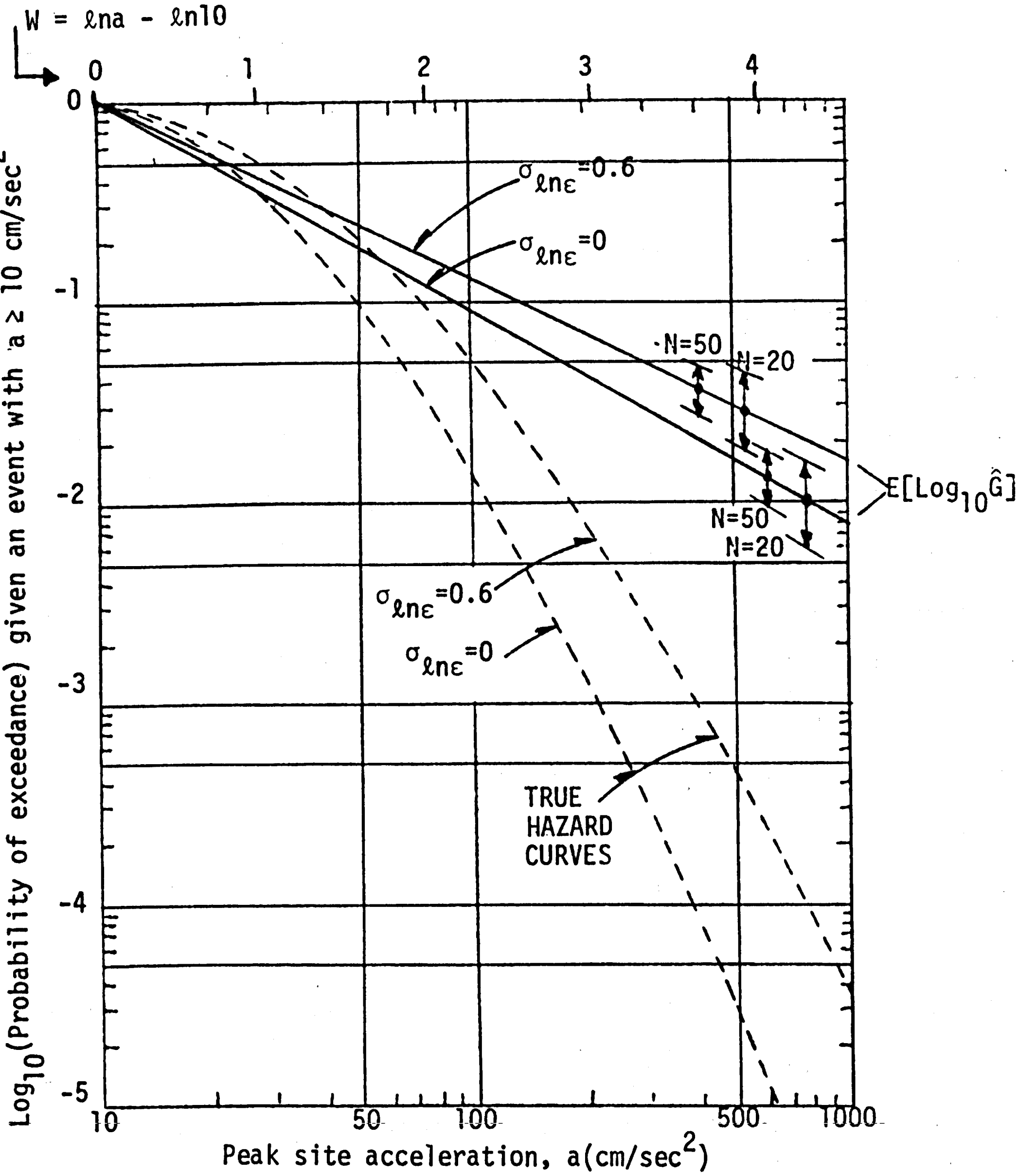


FIGURE 5.1 True hazard curves and estimated hazard curves for H = untruncated exponential family and $\hat{H}$ = maximum-likelihood estimator of G_a.

$$a' = a - a_{min}$$
$$Y = \ell n\ a \tag{5.6}$$
$$W = \ell n\ a - \ell n\ a_{min}\ ,\ W > 0$$

It is convenient here to characterize estimators of hazard as inference rules on $G_Y(y)$ or on $G_W(w)$.

Estimators of $H_T = G_W(w)$ may differ because of differences in the class $\mathcal{H}$ of hazard functions to which G_W is assumed to belong or because of different inference rules, e.g, maximum likelihood, least squares, Bayes. Both aspects have important effects on $\hat{H} = G_W(w)$, especially for large w. In selecting a specific estimator one looks for simplicity, richness ($\mathcal{H}$ should contain a variety of hazard functions so that at least one is close to the unknown true function G_W), and accuracy. The difference $e(w) = G_W(w) - \hat{G}_W(w)$ is a random process with the conditional second moment function

$$B_e(w) = E[e^2(w)|G_W] \tag{5.7}$$

This function should be as small as possible, at least for a certain class of true complementary CDF's, G_W.

Richness and accuracy are related. If $\mathcal{H}$ is a narrow family, then G_W might be very different from all the functions in $\mathcal{H}$. As a consequence, $\hat{G}_W$ is likely to be heavily biased and $B_e(w)$ to be large. At the other extreme, if $\mathcal{H}$ is very rich, then bias may be small but the variance of e(w) is large and $B_e(w)$ is again large. A good choice of $\mathcal{H}$ is one that achieves a balance between bias and error variance for a given class of actual hazard functions and one that, at the same time, is easy to analyze. Bias and error variance have been calculated for the following estimation rules and parametric families of distributions:

Maximum-likelihood Estimation with

$\mathcal{H}$ = one-parameter gamma family.

$\mathcal{H}$ = two-parameter truncated-exponential family.

$\mathcal{H}$ = two-parameter Weibull family.

$\mathcal{H}$ = two-parameter lognormal family.

Bayesian Estimation with

$\mathcal{H}$ = one-parameter exponential or two-parameter truncated-exponential family.

$\mathcal{H}$ = two-parameter Weibull family.

$\mathcal{H}$ = two-parameter lognormal family.

In some other cases, hazard at the site was estimated indirectly by first fitting a regional (as opposed to site) model (source configuration, seismicity parameters, attenuation function) to statistical data. A brief account of these more complicated analyses is given in Sec. 5.4.

For all site models, sample information was assumed in the form of an independent sample of size N, $Z = \{W_1, \ldots, W_N\}$ from the true distribution of W, and estimation error statistics were obtained through repeated Monte Carlo simulation of Z. Specifically, three sets of independent samples were used: Set 1 = 20 samples each of length N = 20 from the distribution of W that corresponds to $\sigma_{\ell n\varepsilon} = 0$; Set 2 = same as set 1 but with $\sigma_{\ell n\varepsilon} = 0.6$; Set 3 = 17 samples each of length N = 50 from the distribution of W that corresponds to $\sigma_{\ell n\varepsilon} = 0.6$. Some approximate analytical results are also given.

5.2 Performance of Simple Maximum-Likelihood Estimators

Given $\mathcal{H}$ as a set of complementary CDF's and given a sample Z from G_W, the maximum-likelihood estimator G_W is the function in $\mathcal{H}$ for which the probability (density) of Z is maximum. Detailed analysis follows for the parametric families $\mathcal{H}$ given previously in Sec. 5.1.

a. $\mathcal{H}$ = one-parameter family of gamma distributions

Let $\mathcal{H} = \{GA(K, \lambda),\ K = \text{shape parameter, known}\}$ be a parametric family of gamma complementary CDF's with λ as the only parameter. Consider first the case when K = 1, i.e., when GA(K, λ) is the exponential function

$$G_W(w) = e^{-\lambda w} \quad , \quad w \geq 0 \tag{5.8}$$

Then, as a function of $Z = \{W_1, \ldots, W_N\}$, the maximum-likelihood estimator of λ is

$$\hat{\lambda} = 1/\overline{W} \text{ , in which } \overline{W} = \frac{1}{N} \sum_{i=1}^{N} W_i \tag{5.9}$$

An accurate and convenient approximation to the distribution of $\overline{W}$ is the gamma distribution with (exact) mean $m_{\overline{W}} = m_W$ and (exact) variance $\sigma_{\overline{W}}^2 = \sigma_W^2/N$. In fact, this would be the correct distribution of $\overline{W}$ if W itself had gamma distribution, as opposed to one of the dashed distributions in Fig. 5.2. Under this approximation, $-\ell n G_W(w) = \hat{\lambda} w$ has so-called inverted gamma-1 distribution (Raiffa and Schlaifer, 1961) and after a few substitutions one finds

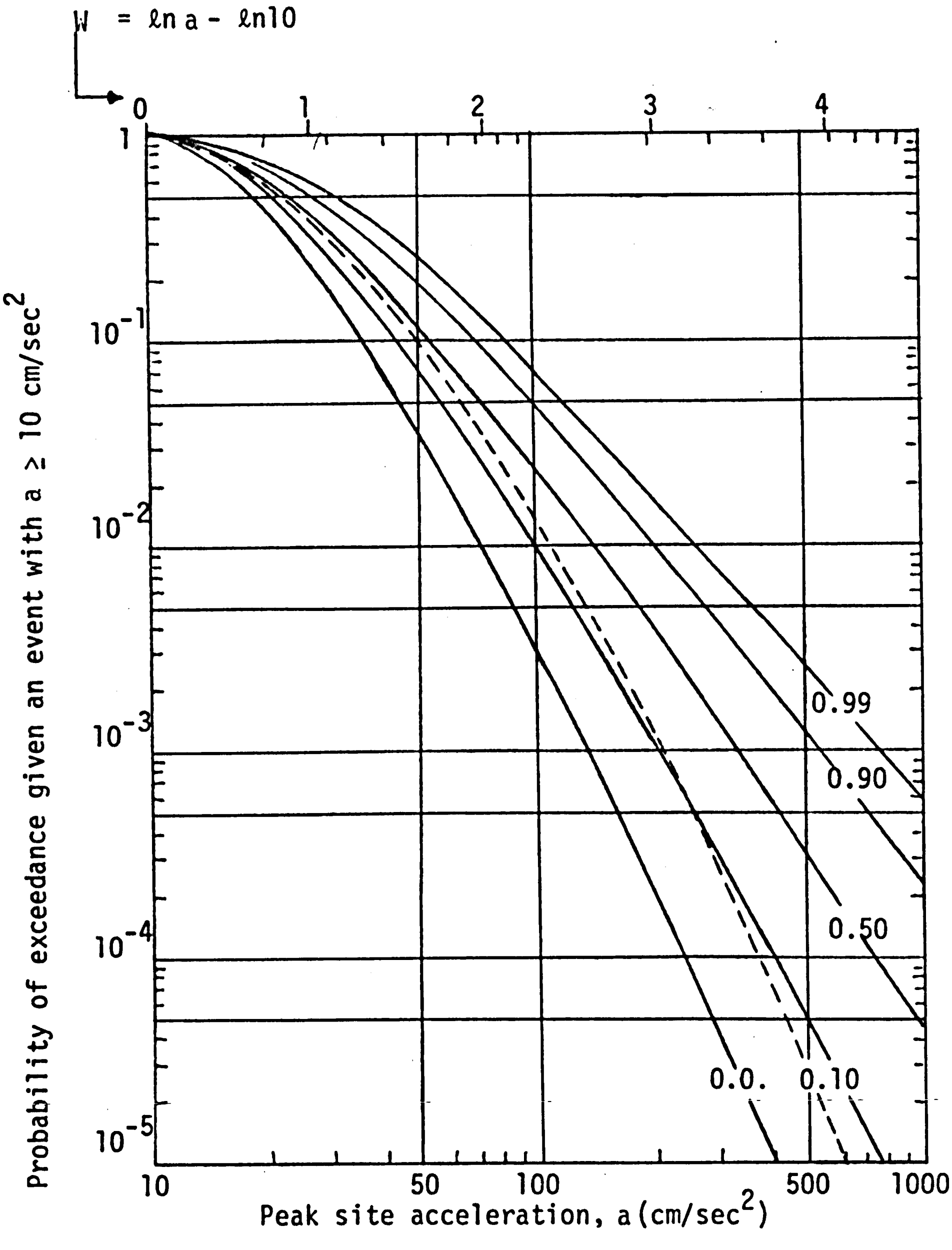


FIGURE 5.2 True hazard (dashed line) and fractiles of the maximum-likelihood estimator (solid lines) when W is assumed to have gamma distribution with shape parameter K = 3. $\sigma_{\ell n \varepsilon} = 0$.

$$
\begin{cases}
E[-\ell n\hat{G}_W(w)] = \dfrac{w}{m_W}\,\dfrac{N}{N-V_W^2} \quad , \quad N > V_W^2 \\[2ex]
\mathrm{Var}[-\ell n\hat{G}_W(w)] = \dfrac{w^2}{m_W^2}\,\dfrac{N^2V_W^2}{(N-V_W^2)^2(N-V_W^2)} \quad , \quad N > 2V_W^2
\end{cases}
\tag{5.10}
$$

in which V_W is the coefficient of variation of W. The mean square error of the maximum-likelihood estimator of $\ell nG_W(w)$ is

$$
\begin{aligned}
B_e(w) &= b_e^2(w) + \mathrm{Var}[-\ell n\ \hat{G}_W(w)] \\
&= [\ell n\ G_W(w) - \frac{w}{m_W}\frac{N}{N-V_W^2}]^2 + \frac{w^2}{m_W^2}\frac{N^2V_W^2}{(N-V_W^2)^2(N-2V_W^2)} \quad , \quad N > 2V_W^2
\end{aligned}
\tag{5.11}
$$

in which b_e denotes bias. Plots of the expected value and of the ±1 standard deviation interval of $\ell n\ \hat{G}_W(w)$ are shown in Fig. 5.1 for N = 20,50 and $\sigma_{\ell n\varepsilon}$ = 0, 0.6. (These analytical approximations are in excellent agreement with simulation results.) For large w, divergence of the expected estimator from the actual hazard indicates that the mean square error in Eq. 5.11 is due mostly to bias; hence, poor performance of the present estimator must be ascribed to the inappropriate choice of $\mathcal{H}$, which contains ~~no model close to the actual distribution, especially in the upper tail.~~ For the same reason, increasing the sample size above N = 20 produces no appreciable improvement.

The previous analysis is readily extended to the case of nonexponential gamma families (K ≠ 1). The maximum-likelihood estimator of λ is now

$$\hat{\lambda} = K/\overline{W} \tag{5.12}$$

with a distribution that is again closely approximated by an inverted gamma-1 function. Given the first two moments of W or the parameters K_W and λ_W that produce the best gamma approximation to G_W (for the case with $\sigma_{\ell n\varepsilon} = 0$ these parameters are $K_W = 4.0$ and $\lambda_W = 4.28$), given the shape parameter K of the $\mathcal{H}$ family and the sample size N, one can use tables of the gamma distribution to calculate fractiles of $\hat{\lambda}$ and $\hat{G}_W(w)$. For example, the fractiles of $\hat{G}_W(w)$ in Fig. 5.2 correspond to $\sigma_{\ell n\varepsilon} = 0$, K = 3, and N = 20. Comparison with Fig. 5.1 shows considerable reduction of bias. Increasing the sample size is more effective than for the exponential family because a larger fraction of the mean square error is contributed by the variance term.

b. $\mathcal{H}$ = <u>two-parameter family of truncated-exponential distributions</u>

The likelihood is now maximized among complementary CDF models of the type

$$G_W(w) = \begin{cases} 1 - K_1(1-e^{-\beta w}), & w \leq w_1 \\ 0\ , & w > w_1 \end{cases} \tag{5.13}$$

in which β and w_1 are unknown positive parameters and $K_1 = (1-e^{-\beta w_1})^{-1}$. The maximum-likelihood estimator of w_1 is $\hat{w}_1 = \max W_i$, but no simple analytical expression is available for $\hat{\beta}$. This last parameter can be found by numerically solving the extreme problem

$$\max_{\hat{\beta} > 0} \left\{ e^{-\beta N\overline{W}} \left(\frac{\hat{\beta}}{1-e^{-\hat{\beta}\hat{w}_1}}\right)^N \right\} \tag{5.14}$$

Due to the form of $\hat{w}_1$, the maximum-likelihood estimator of $\ell n G_W(w)$ is biased, with $E[-\ell n\ \hat{G}_W(w)] = \infty$ for all w and all finite N. It is therefore more meaningful to consider sample estimates of $E[\hat{G}_W(w)]$ since this quantity is finite nonzero for all w. Results are shown as curves G_{TE} in Figs. 5.3, 5.4, and 5.5. Sample estimates of the standard deviation of $\hat{G}_W(w)$ and sample ranges are also indicated.

In maximum-likelihood estimation, the one-parameter exponential and the two-parameter truncated exponential families have opposite behavior with regard to tail probabilities. The exponential family (as all unbounded, one-parameter families) extrapolates the upper tail from the body of the distribution where the data lie, whereas the truncated exponential family (as all families with upper-truncation parameter) makes no extrapolation beyond the largest sample value. In both cases, small-risk estimates are strongly biased.

c. $\mathcal{H}$ = <u>two-parameter Weibull family</u>

The Weibull distribution with parameters α and c has a complementary cumulative function

$$G_W(w) = e^{-\alpha w^c} \quad , \; w > 0, \; \alpha > 0, \; c > 0 \tag{5.15}$$

Given the independent sample $W_1, \ldots, W_N$, the maximum-likelihood estimator of c can be calculated numerically from the condition

$$\frac{N}{\hat{c}} + \sum_{i=1}^{N} \ell n\, W_i - \frac{N \sum_{i=1}^{N} (W_i)^{\hat{c}}\, \ell n\, W_i}{\sum_{i=1}^{N} (W_i)^{\hat{c}}} = 0 \tag{5.16}$$

and then $\hat{\alpha}$ can be found from

$$\hat{\alpha} = \frac{N}{\sum_{i=1}^{N} (W_i)^{\hat{c}}} \tag{5.17}$$

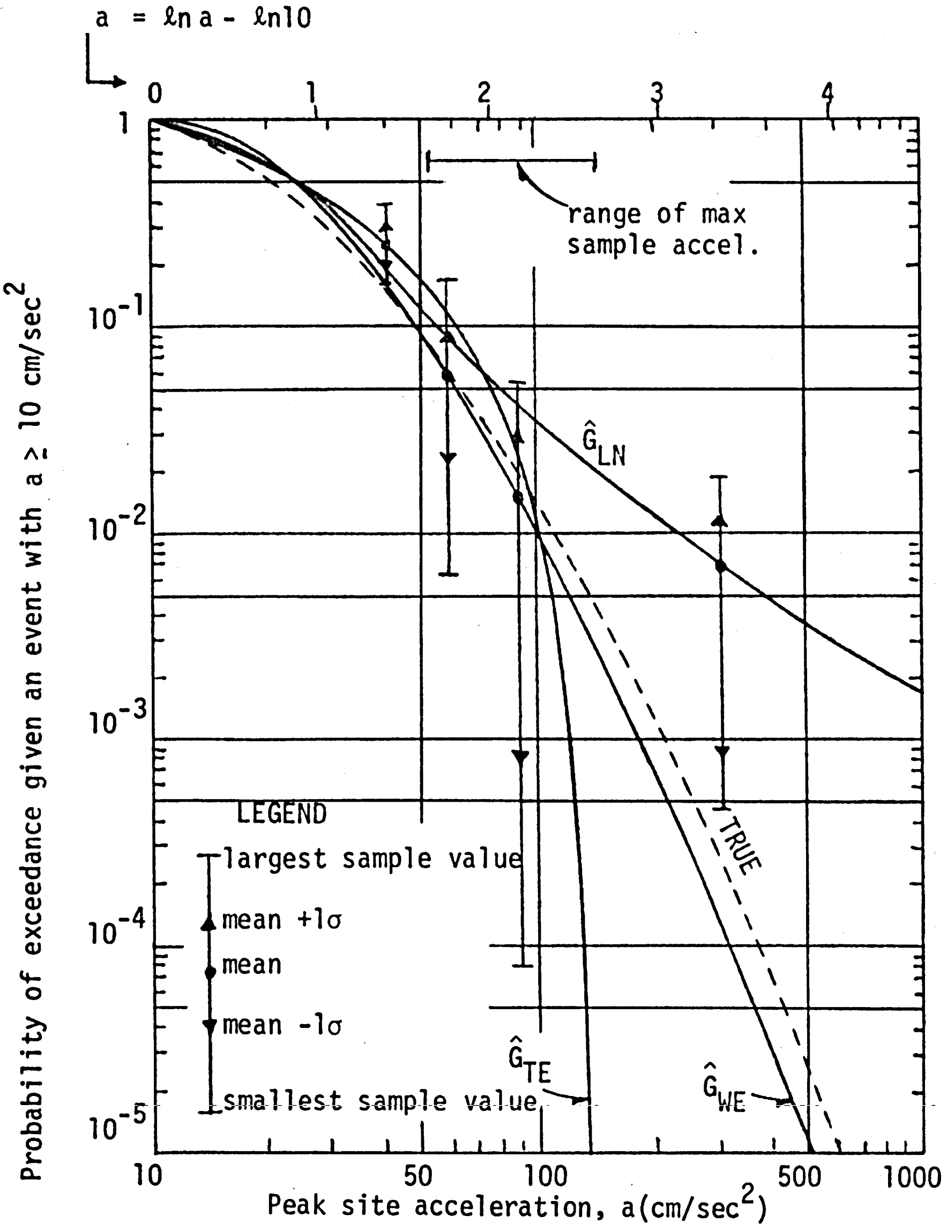


FIGURE 5.3 True hazard and sample statistics of maximum-likelihood hazard estimators when W is assumed to have truncated exponential (TE), Weibull (WE), or lognormal (LN) distribution. Twenty samples of size 20 from the true distribution of W. $\sigma_{\ell n\varepsilon} = 0$.

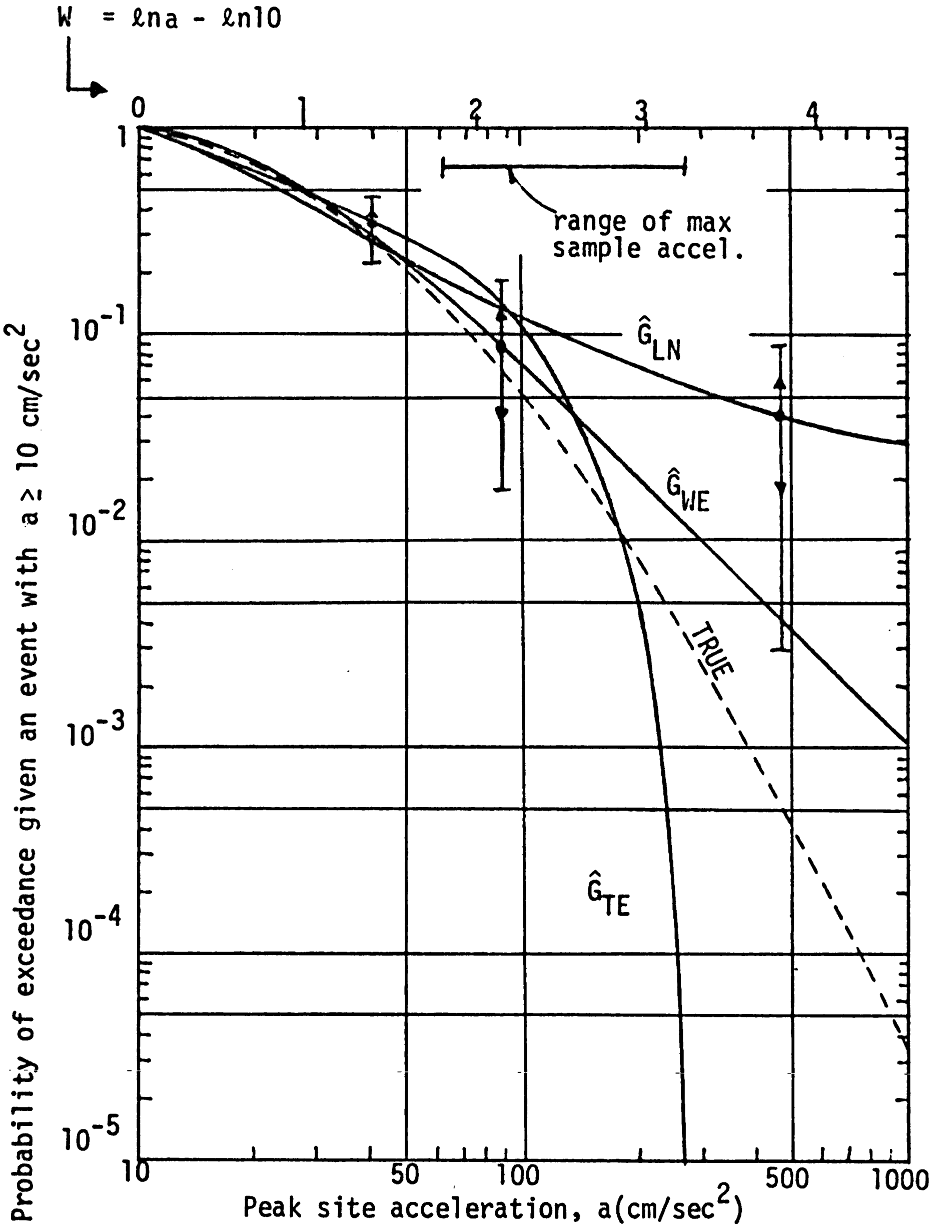


FIGURE 5.4 True hazard and sample statistics of maximum-likelihood hazard estimators when W is assumed to have truncated exponential (TE), Weibull (WE), or lognormal (LN) distribution. Twenty samples of size 20 from the distribution of W. $\sigma_{\ell n \varepsilon} = 0.6$.

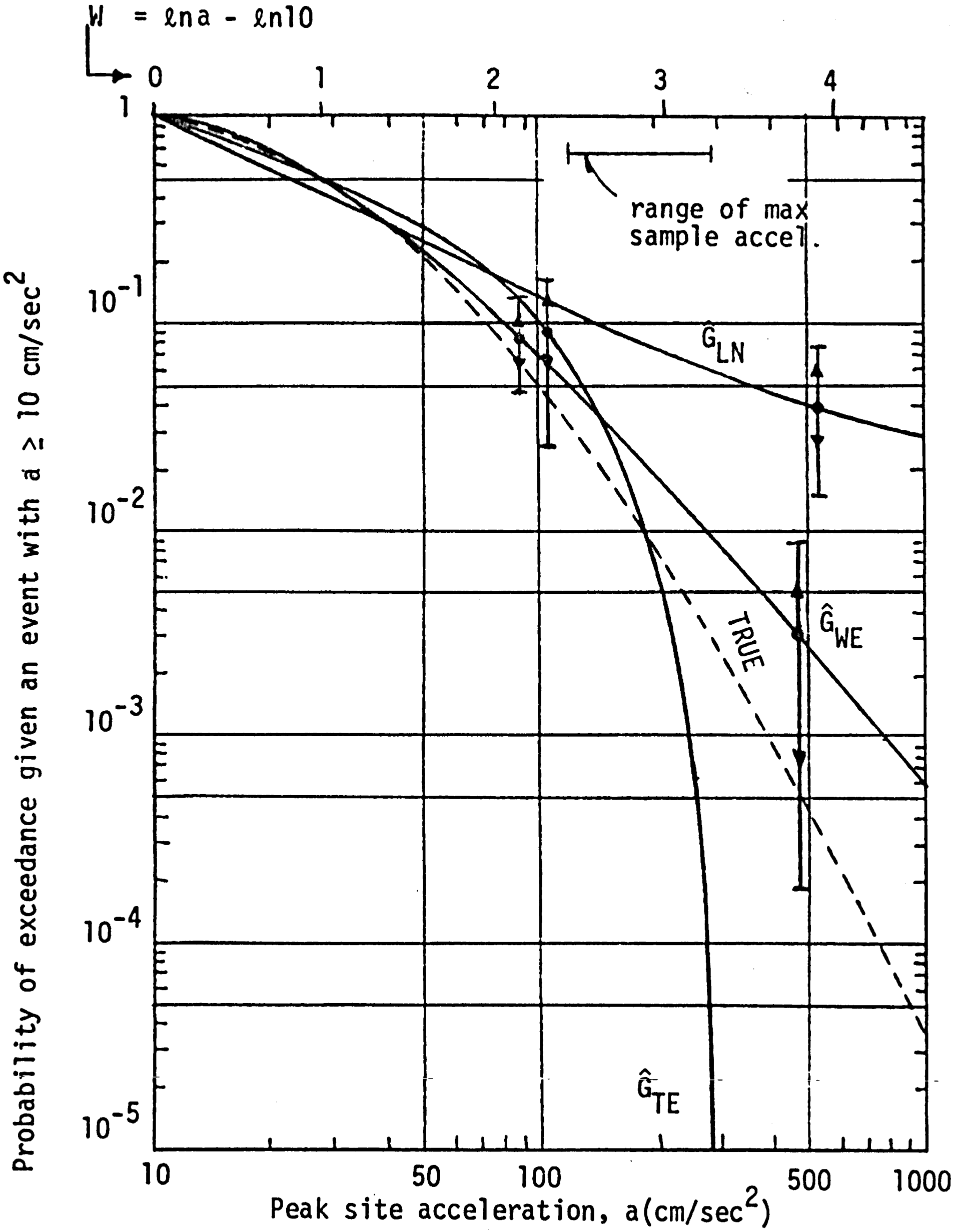


FIGURE 5.5 True hazard and sample statistics of maximum-likelihood hazard estimators when W is assumed to have truncated exponential (TE), Weibull (WE), or lognormal (LN) distribution. Seventeen samples of size 50 from the true distribution of W. $\sigma_{\ell n\varepsilon} = 0.6$.

Results in the form of sample mean values (curves $\hat{G}_{WE}$), sample standard deviations, and ranges of $\hat{G}_W(w)$ are shown in Figs. 5.3, 5.4, and 5.5. The Weibull family includes distributions that are close to the time distribution of W. Hence, the bias is smaller and the percentage reduction of mean square error from increasing the sample size is larger than for distribution families considered before.

d. $\mathcal{H}$ = <u>two-parameter lognormal family</u>

If $\ell n\, W$ is assumed to have normal distribution with unknown mean $m_{\ell nW}$ and variance $\sigma^2_{\ell nW}$, then the maximum-likelihood estimators of these parameters are

$$\hat{m}_{\ell nW} = \frac{1}{N}\sum_{i=1}^{N} \ell n\, W_i$$
$$\hat{\sigma}^2_{\ell nW} = \frac{1}{N}\sum_{i=1}^{N} (\ell n\, W_i - \hat{m}_{\ell nW})^2 \tag{5.18}$$

and the maximum-likelihood complementary CDF of W is

$$\hat{G}_W(w) = 1 - \Phi\left(\frac{\ell nw - \hat{m}_{\ell nW}}{\hat{\sigma}_{\ell n\, W}}\right) \tag{5.19}$$

in which Φ = standard normal CDF. Sample results are shown in Figs. 5.3, 5.4, and 5.5 as $\hat{G}_{LN}$ curves. The lognormal estimator performs better than the exponential but not as well as the Weibull estimator, mainly because the tail of the fitted lognormal distributions falls off too slowly.

5.3 Performance of Simple Bayesian Estimators

Under the assumption that G_W belongs to a parametric family $\mathcal{H}$ of complementary CDF's, uncertainty on the parameters can be included in seismic hazard calculations through Bayesian analysis. If $\underline{\theta}$ is the vector of uncertain parameters, then knowing $\mathcal{H}$ means knowing the conditional complementary CDF, $G_{W|\underline{\theta}}(w, \underline{\theta})$. The prior distribution of $\underline{\theta}$, $F'_{\underline{\theta}}(\underline{\theta})$, is assumed here to be noninformative of either of the following types: for components θ_i with unrestricted values,

$$dF'_{\theta_i}(\theta) \propto d\theta \tag{5.20}$$

whereas for components that are bounded on one side, as $\theta_i \gtrsim \theta_0$,

$$dF'_{\theta_i}(\theta) \propto \frac{d\theta}{\theta-\theta_0}, \quad \theta > \theta_0 \tag{5.21}$$

A priori independence is assumed, so that $dF_{\underline{\theta}}(\underline{\theta}) = \prod_i dF_{\theta_i}(\theta_i)$. By using the distributions in Eqs. 5.20 and 5.21, one eliminates dependence of estimated hazard on nonstatistical information.

Given the independent sample $Z = \{W_1, \ldots, W_N\}$, the posterior CDF of $\underline{\theta}$ is, from Bayes' theorem,

$$dF''_{\underline{\theta}}(\underline{\theta}) \propto dF'_{\underline{\theta}}(\underline{\theta})\, L_{\underline{\theta}|Z}(\underline{\theta}) \tag{5.22}$$

in which the likelihood function $L_{\underline{\theta}|Z}$ satisfies

$$L_{\underline{\theta}|Z}(\underline{\theta}) \propto \prod_{i=1}^{N} \left[-dG_{W|\underline{\theta}}(W, \underline{\theta})/dW\Big|_{W=W_i}\right] \tag{5.23}$$

Finally, the posterior complementary CDF of W is calculated by using the Total Probability Theorem, which gives

$$G''_W(w) = \int_{\text{all } \underline{\theta}} G_{W|\underline{\theta}}(w, \underline{\theta}) dF''_{\underline{\theta}}(\underline{\theta}) \tag{5.24}$$

This last function is the Bayesian estimator of $G_W(w)$. More explicit results are given next for a few parametric families $\mathcal{H}$.

a. $\mathcal{H}$ = <u>one-parameter exponential or two-parameter truncated exponential family</u>

Let $\theta = \lambda$ and denote by $G_{W|\theta}$ the function in Eq. 5.6. Then $L_{\lambda|Z}(\lambda) \propto \lambda^N \exp(-\lambda N\overline{W})$ with $\overline{W}$ = sample mean, and for a prior distribution of λ in the form of Eq. 5.21, the posterior distribution of W' has complementary CDF (use Eqs. 5.22 and 5.24),

$$G''_W(w) = \left(1 + \frac{w}{N\overline{W}}\right)^{-N} \quad , w \geq 0 \tag{5.25}$$

It can be shown (Schumacher, 1977) that this is also the posterior complementary CDF of W if $\mathcal{H}$ is the family of truncated-exponential distributions in Eq. 5.13 and λ and w_1 have jointly noninformative prior density, $f'_{\lambda,w_1}(\lambda, w_1) \propto (\lambda w_1)^{-1}$, for $\lambda \geq 0$, $w_1 \geq 0$.

b. $\mathcal{H}$ = <u>two-parameter Weibull family</u>

For $\underline{\theta} = [\alpha, c]$ and $G_{W|\underline{\theta}}$ in Eq. 5.15, the likelihood function of $\underline{\theta}$ given Z is

$$L_{\alpha, c|Z}(\alpha, c) \propto \alpha^N c^N \exp\left\{-\alpha \sum_{i=1}^{N} (W_i)^c\right\} \prod_{i=1}^{N} (W_i)^{c-1} \tag{5.26}$$

If both α and c have prior distribution in the form of Eq. 5.21, then Eq. 5.24 yields

$$G''_W(w) = \frac{1}{K} \int_0^\infty c^{N-1} \prod_{i=1}^{N} (W_i)^{c-1} \frac{\Gamma(N)dc}{[\sum_{i=1}^{N} (W_i)^c + w^c]^N} \tag{5.27}$$

in which $K = G''_W(0)$ is a normalization constant. The integral of Eq. 5.27 must be evaluated numerically.

c. $\mathcal{H}$ = <u>two-parameter lognormal family</u>

Let $m_{\ell nW}$ and $\sigma^2_{\ell nW}$ have prior joint density function

$$f'_{m_{\ell nW}, \sigma^2_{\ell nW}}(m, \sigma^2) \propto \frac{1}{\sigma^2} \quad , \quad \sigma^2 \geq 0 \tag{5.28}$$

Then Bayesian analysis (Raiffa and Schlaifer, 1961) shows that the random variable

$$S(W) = (\ell n\ W - \hat{m}_{\ell nW}) \left[\frac{N(N-1)}{N+1}\right]^{1/2} \left[\sum_{i=1}^{N} (\ell n\ W_i - \hat{m}_{\ell nW})^2\right]^{1/2} \tag{5.29}$$

has posterior Student-t cumulative distribution $F_{t_{N-1}}$, with N-1 degrees of freedom. From the fact that S is a monotonically decreasing function of W one concludes,

$$G''_W(w) = P[S(W) > S(w)] = 1 - F_{t_{N-1}}(S(w)) \tag{5.30}$$

Simulation results for the Bayesian estimators in Eqs. 5.25, 5.27, and 5.30 are shown in Figs. 5.6, 5.7, and 5.8. Previous considerations about

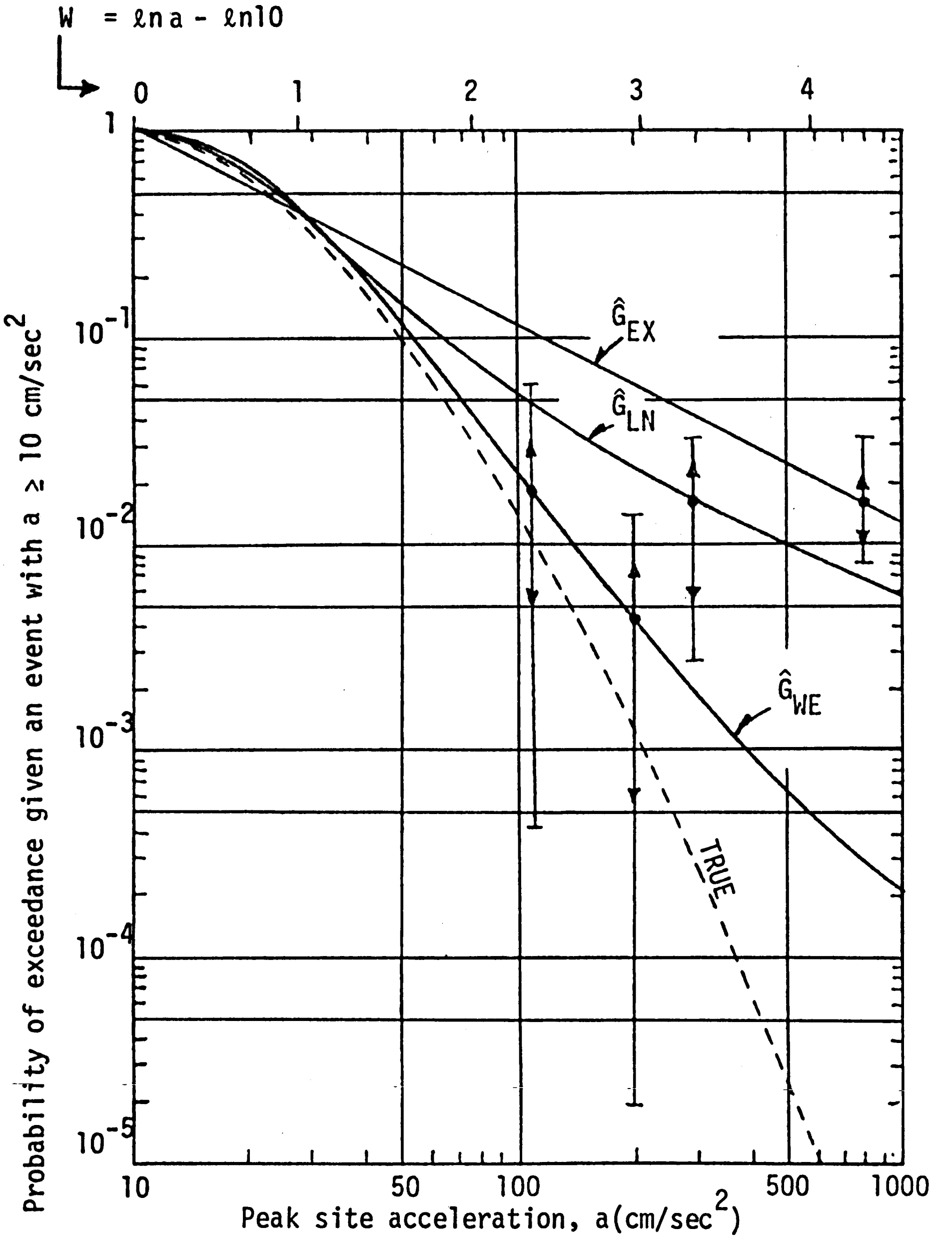


FIGURE 5.6 True hazard and sample statistics of Bayesian estimators when W is assumed to have exponential or truncated exponential (EX), Weibull (WE), or lognormal (LN) distribution. Twenty samples of size 20 from the true distribution of W. $\sigma_{\ell n \varepsilon} = 0$.

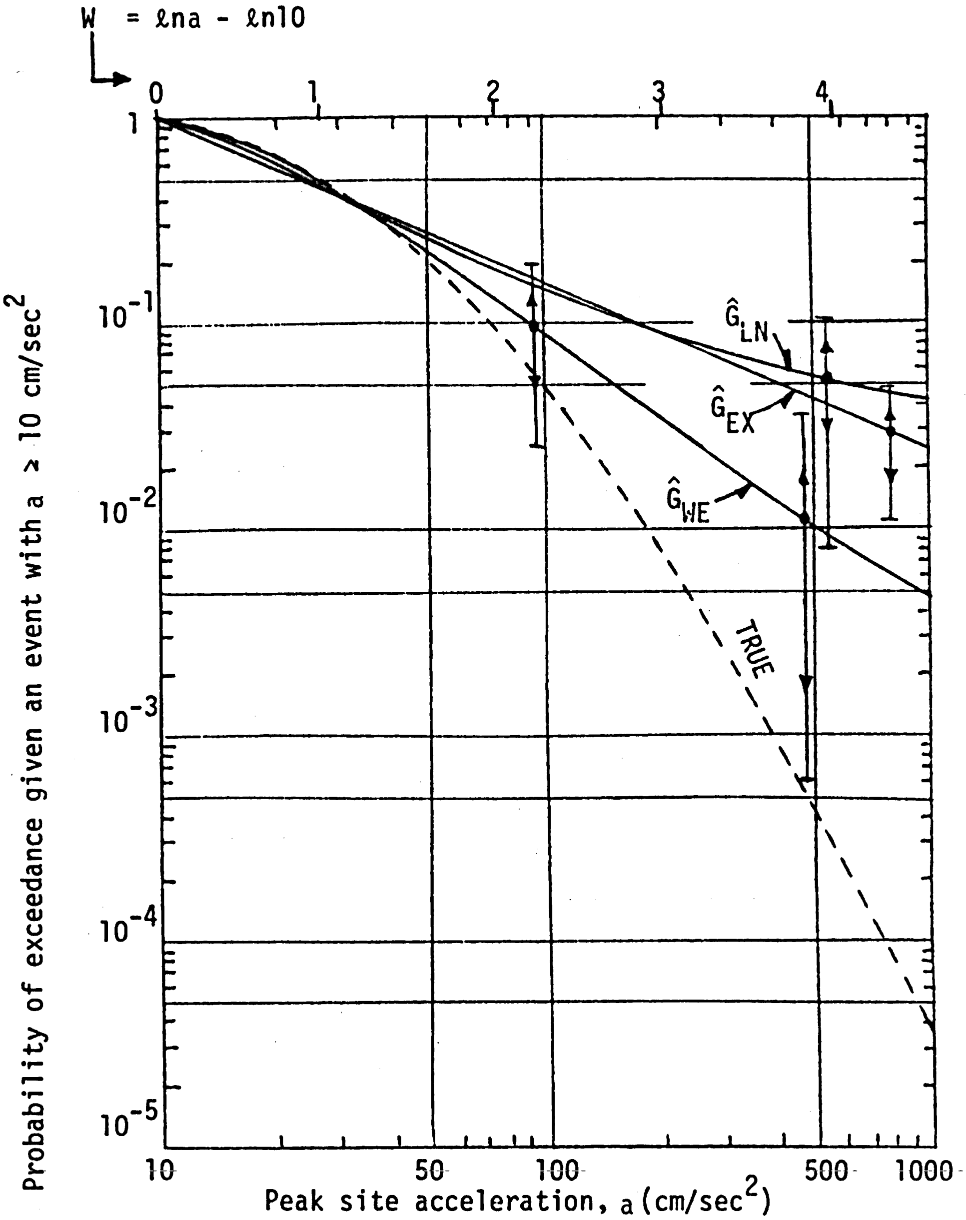


FIGURE 5.7 True hazard and sample statistics of Bayesian estimators when W is assumed to have exponential or truncated exponential (EX), Weibull (WE), or lognormal (LN) distribution. Twenty samples of size 20 from the true distribution of W. $\sigma_{\ell n\varepsilon} = 0.6$.

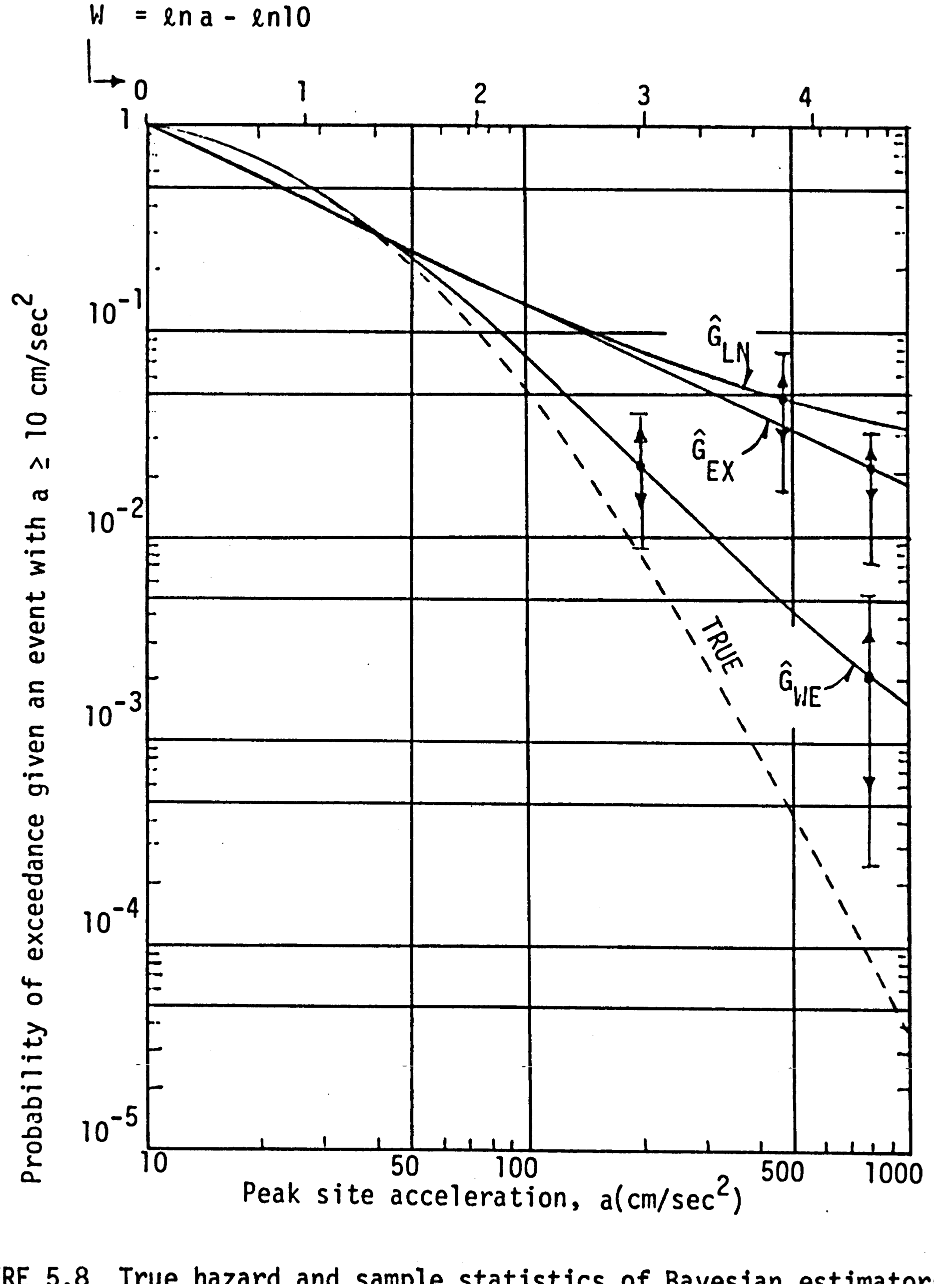


FIGURE 5.8 True hazard and sample statistics of Bayesian estimators when W is assumed to have exponential or truncated exponential (EX), Weibull (WE), or lognormal (LN) distribution. Seventeen samples of size 50 from the true distribution of W. $\sigma_{\ell n\varepsilon} = 0.6$.

maximum-likelihood estimators still apply. However, Bayesian estimators with noninformative prior distribution of the parameters tend to be more conservative in the range of high ground motion intensities. The most evident case is that of the truncated-exponential family, but differences in mean estimated hazard are noticeable also in the Weibull case (compare Figs. 5.5 and 5.8). One reason for conservatism is that Bayesian estimation accounts for uncertainty on parameters in which the complementary CDF is very nonlinear. Bayesian and maximum-likelihood estimators coincide in the limit, as $N \to \infty$.

5.4 More Complicated Estimators

The previous results indicate that hazard estimators are very sensitive to the choice of the distribution family, $\mathcal{H}$. One may attempt to reduce sensitivity by using richer families of distributions, e.g., families with three or four parameters, or composite distribution sets $\mathcal{H} = \{\mathcal{H}_1, \ldots, \mathcal{H}_q\}$ in which $\mathcal{H}_i$ is itself a distribution family, e.g., one of those considered in Secs. 4.2 and 4.3. The inference rule may still be of the maximum-likelihood or Bayesian type, the latter in the form of a weighted average of all the distributions in $\mathcal{H}$. A Bayesian approach with composite $\mathcal{H}$ has been developed by Wood et al. (1974). An approximate, simpler "pseudo-Bayesian" rule has been suggested by Grigoriu (1976), of the type

$$G_W''(w) \propto \sum_{i=1}^{q} P_i' L_i \, G_{W_i}''(w) \tag{5.31}$$

in which P_i' = prior probability that $G_W \in \mathcal{H}_i$.

L_i = likelihood of the maximum-likelihood model in $\mathcal{H}_i$.

G_{W_i}'' = posterior Bayesian complementary CDF of W, given that $G_W \in \mathcal{H}_i$.

When using maximum-likelihood or pseudo-Bayesian inference, the estimated hazard function critically depends on the likelihood ratios L_2/L_1, L_3/L_1, ..., L_q/L_1. For $\mathcal{H} = \{\mathcal{H}_1, \mathcal{H}_2\}$ with $\mathcal{H}_1$ the two-parameter lognormal family and $\mathcal{H}_2$ the two-parameter Weibull family, and for 10 independent samples of size 20 and 50, ratios L_2/L_1 are given in Table 5.1. All samples have been pseudo-randomly simulated from the time distribution of W that corresponds to $\sigma_{\ell n\varepsilon} = 0.6$. Although in the case of these samples a Weibull distribution always exists with likelihood higher than that of any lognormal distribution, the large sample variability of L_2/L_1 implies, at least for N = 20, large statistical variability of the pseudo-Bayesian estimator. This fact in turn indicates that recourse to a richer family of distributions produces better results in the range of high intensities, only if the sample size is sufficiently large (e.g., N > 50).

All model estimators $\hat{M}$ (= $\hat{G}_W$) and model families $\mathcal{M}$ (= $\mathcal{H}$) considered up to here in Sec. 5 are of the <u>site</u> type, i.e., they directly describe seismic hazard at the site of interest. If historical data are also available on earthquake source parameters (e.g., epicentral coordinates and magnitude), then $\mathcal{H}$ can be replaced with a family $\mathcal{M}$ of <u>regional models</u>, e.g., of the same type as the true model M_T in Sec. 5.1. As an example, let $\mathcal{M}$ collect all regional models of the same type as M_T with given parameters $m_0 = 4$, $b_4 = 25$, $\sigma_{\ell n\varepsilon} = 0$ and uncertain parameters m_1, b, r, b_1, b_2, and b_3. The analyst must estimate these parameters from data on magnitude, distance, and peak site acceleration. Statistical sampling is from the true mechanism with parameters in Eq. 5.5 and $\sigma_{\ell n\varepsilon} = 0.6$. The difference between assumed and true variance of the attenuation residual term ε is therefore the only source of modeling error. Samples have the form, $Z = \{(M, R, a)_1, \ldots, (M, R, a)_N\}$ in which $(M, R, a)_i$ stands for magnitude, epicentral distance,

N = 20				N = 50			
Sample #	L_2/L_1	#	L_2/L_1	Sample #	L_2/L_1	#	L_2/L_1
1	11.4	6	7.6	1	446	6	3354
2	4.1	7	13.7	2	4836	7	25
3	14.7	8	20.3	3	522	8	1615
4	1.3	9	403.2	4	2489	9	44
5	354.1	10	9.8	5	22420	10	2161

TABLE 5.1 Likelihood ratio L_2/L_1 for maximum-likelihood Weibull and lognormal distributions. Independent samples of size N are simulated from the true distribution of W for $\sigma_{\ell n\varepsilon} = 0.6$.

and peak site acceleration for the ith earthquake. For samples of this type and noninformative prior distribution

$$f'_{m_1,b,r,\ell nb_1,b_2,b_3}(m_1,b,r,\beta,b_2,b_3) \propto \frac{1}{m_1 br} \tag{5.32}$$

Schumacher (1977) found an expression for the posterior complementary CDF, $G''_W(w)$. The average of this function over 20 independent samples, each of size N = 20, is shown in Fig. 5.9 as curve $\hat{G}$. One-standard deviation intervals are also indicated. The large bias is due primarily to uncertainty on the upper-bound parameter m_1 of the magnitude distribution.

The most important consideration from the error analyses of this section is that, when hazard estimates are based solely on statistical information, large biases may result. Biases are sometimes due to the estimation procedure (recall earlier comparison between maximum-likelihood and Bayesian inference of distributions with upper truncation) but they are more likely due to possible errors in modeling assumptions (see the variation of mean hazard with $\mathcal{H}$).

One way to reduce the variability of hazard estimates is to utilize nonstatistical information, e.g., by replacing the posterior distribution of the parameters in Bayesian analysis with subjectively assessed distributions. The latter distributions are usually narrower because they are based not only on historical data, but also on nonstatistical considerations about local geology, earthquake mechanisms, and similarity with other seismic regions. The technique relies on "expert opinion" and is becoming increasingly popular, especially for the assessment of hazard in regions with poor historical records. It is itself not free of difficulties, however, as will be said in the next section.

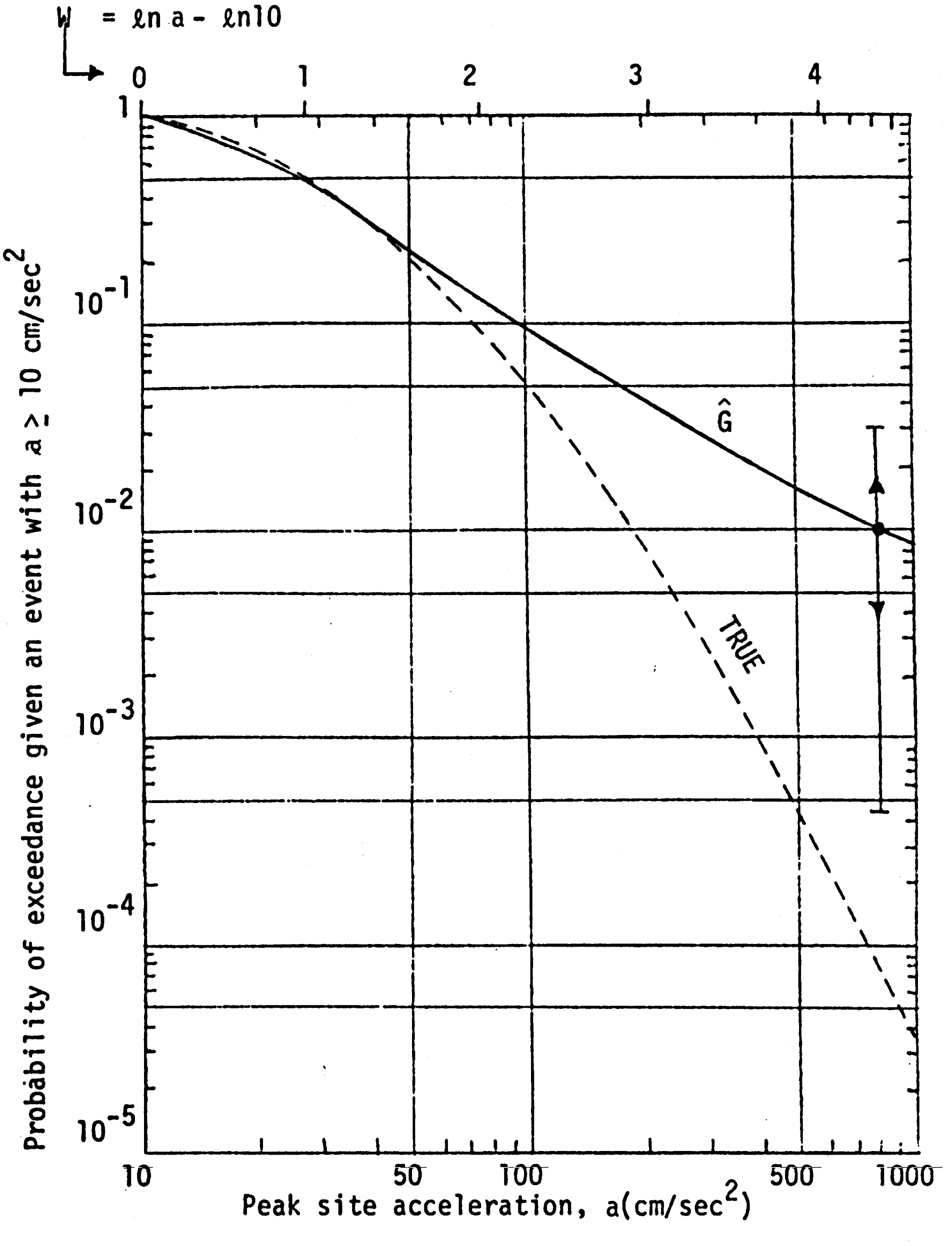


FIGURE 5.9 True hazard and sample statistics (from 20 samples of size 20) of a Bayesian estimator. The estimator corresponds to assuming that the regional earthquake mechanism is of the same type as the actual mechanism, with $\sigma_{\ell n\varepsilon} = 0$ and a few uncertain parameters. For the true mechanism, $\sigma_{\ell n\varepsilon} = 0.6$.

6. THE USE OF EXPERT OPINION

The basic problem when turning to experts for the quantification of parameter uncertainty is that subjective estimates are not totally reliable. For this reason one typically asks for the opinion of several experts and from each of them obtains a separate hazard estimate, $\hat{H}_i$. How should one interpret the difference between $\hat{H}_i$ and $\hat{H}_j$? What is the procedure to combine the $\hat{H}_i$ and produce a single (sometimes called consensus) hazard estimate $\hat{H}$? How does accuracy of $\hat{H}$ depend on the number n of experts? On their degree of expertise?

The first step toward answering these questions is to make a conceptual model of professional judgement. Conceptual models should then be given mathematical form and analyzed. The present section follows this ideal path by first introducing two alternative interpretations of the cause for diversity in expert opinion. Each interpretation contains elements of truth. The mathematical model that follows in Sec. 6.2 incorporates features of both interpretations. It is flexible and simple. There is obviously no contention of completeness in dealing with a topic which has been under study for decades, primarily by nonengineers; see for example Winkler (1968), de Groot (1974), Morris (1974), Dalkey (1975).

6.1 Conceptual Models

There are two seemingly conflicting interpretations of the diversity of expert opinion. One interpretation views each estimate $\hat{H}_i$ as a noisy observation of the true hazard H_T; the other considers each estimate as a sample from a random process that expresses objective professional uncertainty about H_T.

Observation model. Through mathematical and cognitive procedures, each expert operates on statistical and physical information to produce an estimate $\hat{H}_i$ of H_T. The difference between $\hat{H}_i$ and H_T is attributed to statistical estimation or observation errors, as if experts were "observing" the true hazard by means of imperfect instruments. Different experts may use different sets of data and different theories and procedures, thus producing different estimates.

Sampling model. It is recognized at the outset that statistical information and physical knowledge are limited so that no matter how many experts are questioned, there will always be uncertainty about the actual hazard. Estimates $\hat{H}_i$ are considered as random samples that reflect this irreducible, professional uncertainty. Therefore, one should not reduce the $\hat{H}_i$ to a single "consensus" estimate, $\hat{H}$. Rather, the full diversity of opinions should be retained and used to quantify professional uncertainty on H_T.

Both models are valuable. For example, it is logical to think of an objective irreducible uncertainty on H_T as predicated by the sampling model. However, viewing expert estimates as I.I.D. variables from a hypothetical "professional distribution" is not correct. Not only does each expert operate on a different fraction of the total information available, but he also makes errors depending on professional experience, degree of familiarity with hazard models, etc. One might think, therefore, of some additional type of error in the $\hat{H}_i$, of the type emphasized by the observational model. This will be called personal error. Whereas professional uncertainty is the same for all experts, personal uncertainty varies. This variation is one of the reasons why one may want to give different weights to different estimates. In the next section, mathematical models will be developed to represent all these components of uncertainty. First a quantitative model of the observational point of view will be developed and then it will be generalized to incorporate the idea of professional uncertainty. It is always assumed that the objective is to estimate (uncertainty on) the mean rate Λ at which a given ground acceleration level a is exceeded at the site. The estimate of Λ based on information from the ith expert is denoted by $\hat{\Lambda}_i$.

6.2 Mathematical Models

According to the observation model, the ith log hazard estimate, $\ell n\hat{\Lambda}_i$, is considered to be a noisy observation of $\ell n\ \Lambda$ with random error ε_i = $\ell n\ \hat{\Lambda}_i - \ell n\ \Lambda$, so that

$$\ell n\ \hat{\Lambda}_i = \ell n\ \Lambda + \varepsilon_i \tag{6.1}$$

The fact that experts use basically the same historical data, geological information, seismicity theories, and hazard procedures makes the error terms in Eq. 6.1 probabilistically dependent. Suppose for convenience that the vector $\underline{\varepsilon} = [\varepsilon_1, \ldots, \varepsilon_n]^T$ has multivariate normal distribution with zero mean and covariance matrix $\underline{\Sigma}_\varepsilon$. If, before obtaining any expert estimate, the hazard $\ell n\ \Lambda$ is independent of $\underline{\varepsilon}$ and has noninformative flat prior distribution, then a posteriori $\ell n\Lambda$ has normal distribution with mean

$$m''_{\ell n\Lambda} = (\underline{1}^T\ \underline{\Sigma}_\varepsilon^{-1}\ \ell n\ \underline{\hat{\Lambda}})/(\underline{1}^T\ \underline{\Sigma}_\varepsilon^{-1}\ \underline{1}) \tag{6.2}$$

and variance

$$\sigma^{2''}_{\ell n\Lambda} = 1/(\underline{1}^T\ \underline{\Sigma}_\varepsilon^{-1}\ \underline{1}) \tag{6.3}$$

where $\underline{1}$ denotes the column vector with n unit components and $\ell n\ \hat{\underline{\Lambda}}$ is the vector of estimates, $\ell n\ \hat{\underline{\Lambda}} = [\ell n\ \hat{\Lambda}_1, \ldots, \ell n\ \hat{\Lambda}_n]^T$. In the case of only one expert, this posterior distribution becomes simply $N(\ell n\ \hat{\Lambda}, \sigma_\varepsilon^2)$.

Due to linearity of the observation model in Eq. 6.1 and normality of $\underline{\varepsilon}$, the posterior mean value in Eq. 6.2 is a linear combination of the expert estimates, with coefficients that depend on the covariance matrix of the errors. The posterior variance, Eq. 6.3, is a nonincreasing function of the number of experts and for given n, increases with the variances and the correlation coefficients of the errors.

If the errors were independent and identically distributed so that $\underline{\Sigma}_\varepsilon = \sigma_\varepsilon^2\ \underline{I}$, then one would find that a posteriori

$$\ell n\ \Lambda \sim N\left(\frac{1}{n} \sum_{i=1}^{n} \ell n\ \hat{\Lambda}_i\ ,\ \frac{1}{n}\ \sigma_\varepsilon^2\right) \tag{6.4}$$

The posterior variance of $\ell n\ \Lambda$ is larger than σ_ε^2/n if errors are positively correlated. A case with correlation which leads to explicit results is that in which experts are treated as equally credible and correlation between estimation errors is the same, irrespective of the identity of the experts. Covariance matrices $\underline{\Sigma}_\varepsilon$ that express these conditions have the equicorrelated form,

$$\underline{\Sigma}_\varepsilon = \sigma_\varepsilon^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{bmatrix} \tag{6.5}$$

This last model may not be accurate in practice if the identity of the experts is known, but is realistic in the case of randomization, i.e., when pairing between experts and estimates is not known or is ignored. The technique of randomization produces suboptimal estimates because it neglects available information, but has the benefit of reducing the complexity of analysis and the number of parameters (to only n, σ_ε^2, and ρ). For $\underline{\Sigma}_\varepsilon$ in Eq. 6.5, the posterior distribution of $\ell n\ \Lambda$ is

$$\ell n\ \Lambda \sim N\left(\frac{1}{n}\sum_{i=1}^{n} \ell n\ \hat{\Lambda}_i \quad , \quad \sigma_\varepsilon^2 \frac{1 + (n-1)\rho}{n}\right) \tag{6.6}$$

The quantity $\frac{n}{1 + (n-1)\rho}$, which can be interpreted as the equivalent number of independent experts, is plotted in Fig. 6.1 as a function of n for selected values of the correlation coefficient ρ. Notice that, as $n \to \infty$, the posterior variance tends asymptotically to

$$\sigma_\infty^2 = \lim_{n\to\infty} \sigma_{\ell n\Lambda}^{2''} = \rho\ \sigma_\varepsilon^2 \tag{6.7}$$

which is a quantity larger than zero, unless $\rho = 0$. This means that an infinite number of expert estimates with common correlation coefficient ρ has the same information content as a single estimate with error variance $\rho\sigma_\varepsilon^2$.

A possible interpretation of the asymptotic posterior distribution $N(m_\infty, \sigma_\infty^2)$ is that this distribution represents <u>irreducible professional uncertainty</u> about the true seismic hazard. This is the uncertainty after considering the opinion of all living experts; it cannot be reduced because it is due to objective limitation of information and knowledge.

Let ε_0 be a random variable with the asymptotic distribution of $\ell n\ \Lambda$ and define variables $\eta_1, \ldots, \eta_n$ with distribution $\eta_i \sim N(0, \sigma_{\eta_i}^2)$. The

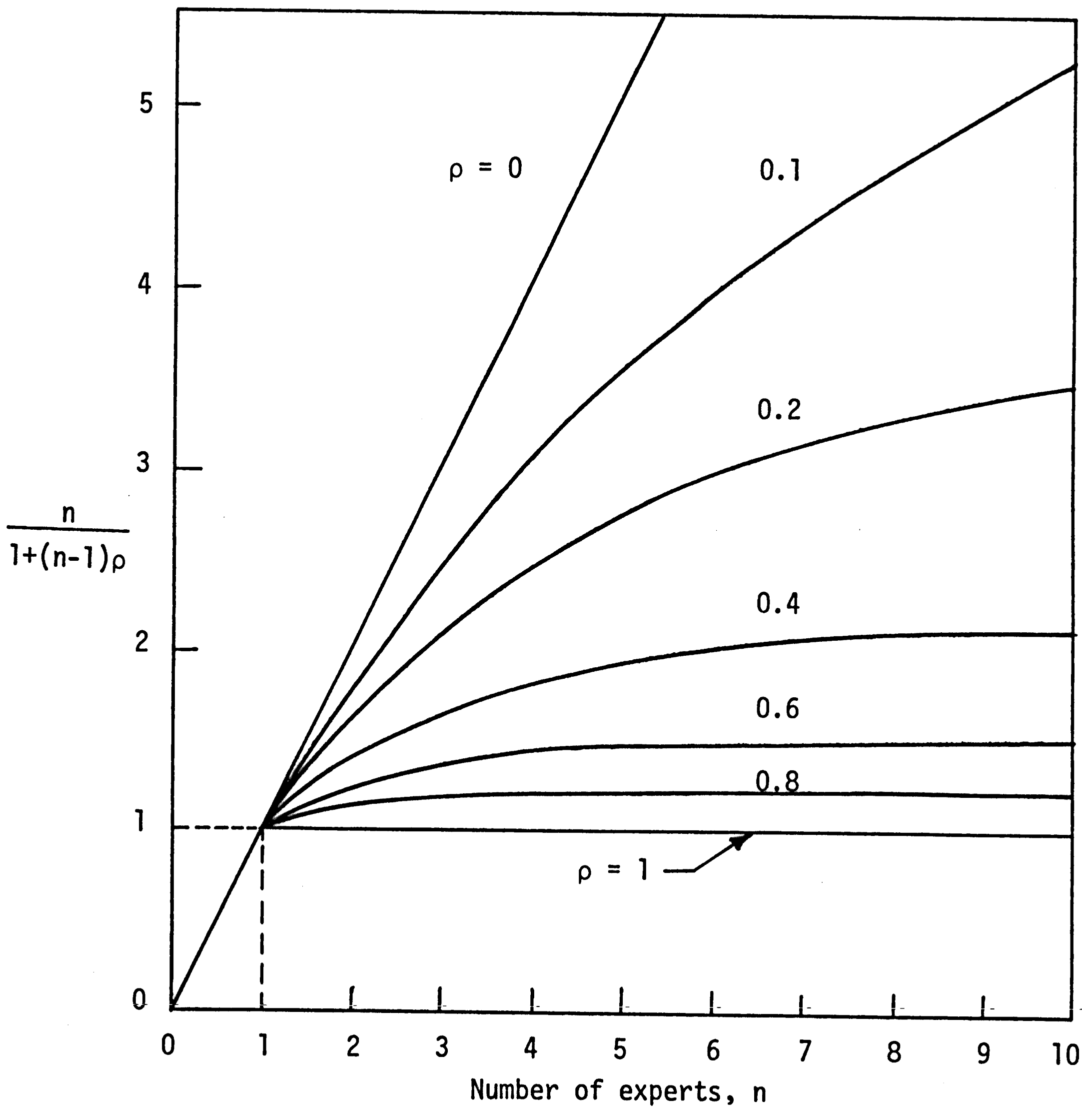


FIGURE 6.1 Equivalent number of independent experts as a function of n and ρ (equicorrelated model).

quantity η_i can be regarded as the component of personal estimation error for the ith expert. With ε_0 and the η_i mutually independent variables, the model of Eq. 6.1 can be written

$$\ell n\ \hat{\Lambda}_i = \ell n\ \Lambda + \varepsilon_0 + \eta_i \tag{6.8}$$

in which $\varepsilon_0 + \eta_i$ has been substituted for ε_i. For example, the covariance matrix in Eq. 6.5 results from setting $\sigma^2_{\eta_i} = (1-\rho)\sigma^2_\varepsilon$ for all experts.

It is clear from Eq. 6.8 that the present mathematical model considers professional uncertainty as systematic error (bias) and associates it with the random variable ε_0. On the contrary, the sampling viewpoint predicates that professional uncertainty is responsible for variation among expert estimates, not for bias. It is believed that both concepts are true and that professional uncertainty has in fact both a bias and a random component--bias due to limitation of information, and random component caused by alternative possible interpretations of seismicity and alternative models for hazard calculation. The model of Eq. 6.8 should then be generalized to

$$\ell n\ \hat{\Lambda}_i = \ell n\ \Lambda + \varepsilon_0 + \xi_i + \eta_i \tag{6.9}$$

in which

$\varepsilon_0 \sim N(0, \sigma_0^2)$ is random bias from limited information.

$\xi_i \sim N(0, \sigma_\xi^2)$ is a random component associated with uncertainty on the interpretation of seismicity and on the model for hazard calculation.

$\eta_i \sim N(0, \sigma_\eta^2)$ is personal error.

Correlation may exist among some of these variables. Professional uncertainty corresponds to the random variable, $\gamma = \varepsilon_0 + \varepsilon_i \sim N(0, \sigma_0^2 + \sigma_\xi^2)$. The objective of analysis is to estimate the distribution of γ, *i.e.*, the variance terms σ_0^2 and σ_ξ^2.

Unfortunately, it is clear that the model postulated by Eq. 6.9 is not inferable from the assumed data structure, $\{\ell n\, \hat{\Lambda}_1, \ldots, \ell n\, \hat{\Lambda}_n\}$. Specifically, the variance of ε_0 cannot be estimated and it is impossible to separate η_i from ξ_i unless correlation is introduced. One can think of other, realistic data structures that would make estimation feasible, e.g., structures that are typical of analysis-of-variance problems.

Let us start with separation between ξ_i and η_i. A way to estimate the variances σ_ξ^2 and σ_η^2 is to first specify a realistic set of alternative seismicity interpretations and hazard models and form a matrix of estimates

$\ell n\, \hat{\Lambda}_{ij}$ = estimate of $\ell n\, \Lambda$ using information from expert i and the jth combination of seismicity interpretation and hazard model

(6.10)

Standard methods of analysis of variance (e.g., Draper and Smith, 1966) will then provide estimates of σ_ξ^2 and σ_η^2.

The estimation of σ_0^2 is somewhat more complicated. One way to obtain information on σ_0^2 is through analyses of the type in Sec. 5, i.e., by pseudo-randomly generating alternative data sets (possibly also alternative seismic environments) and by calculating the associated variation of estimated hazard.

6.3 Actual Variability of Expert Opinion

In a recent study of Seismic Hazard in the Eastern United States (TERA, 1979), 10 seismologists responded to a comprehensive questionnaire which included information on source geometry, seismicity rate and magnitude distribution for each source, and form and parameters of the attenuation function. Before answering the questionnaire, experts were provided with available maps of seismicity and with exhaustive catalogues of historical events. In spite of this attempt to equalize the information available to all experts, opinions ranged widely on a few critical parameters, notably, on the appropriate configuration of earthquake sources, on the slope of the recurrence relationship, and on the upper bound size. Based on response from each expert, 10 separate estimates of hazard at various locations of the eastern and central United States were obtained. Results at one site, in terms of pseudo-velocity spectra for a 1000-year return period, are shown in Fig. 6.2. The range of estimates corresponds to factors on the spectrum of typically 2 or 3, with peak values of about 7 for long-period oscillators. With reference to the model of Eq. 6.9, this variability should be attributed mainly to the term η_i and less to the term ξ_i because of the similarity of hazard analysis procedures (but the interpretation of seismicity was allowed to vary from expert to expert).

A self-ranking method (experts were asked to quantify their degree of confidence in answering various parts of the questionnaire) was devised to finally produce a single hazard estimate $\hat{H}$ as a weighted average of the individual curves. However, precision of $\hat{H}$ was not quantified.

Another example is shown in Fig. 6.3 (from Cornell, 1980). In this case, three independent consulting teams obtained estimates of seismic hazard at the site of the Diablo Canyon nuclear power plant, in California.

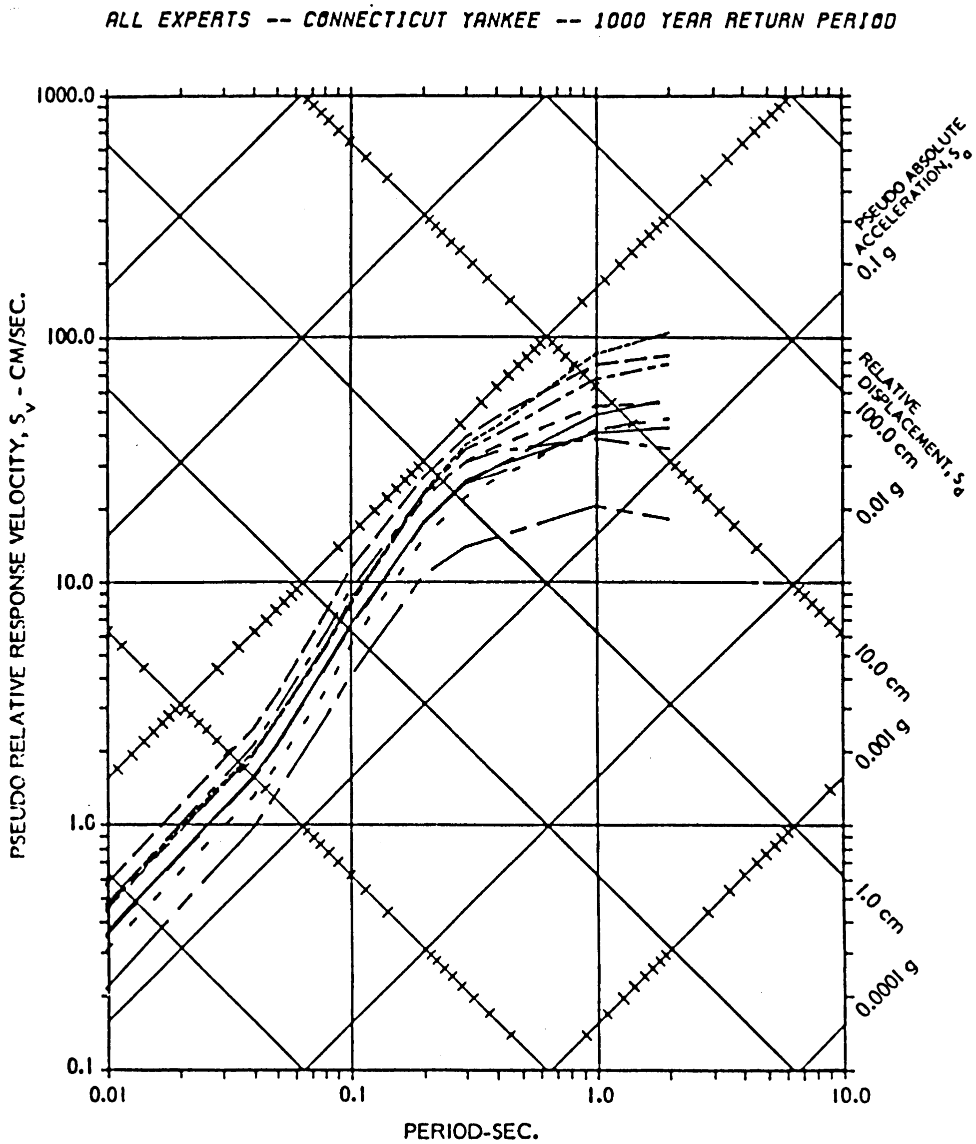


FIGURE 6.2 Diversity of expert opinion in the assessment of seismic hazard at Connecticut Yankee power station site (from TERA, 1979)

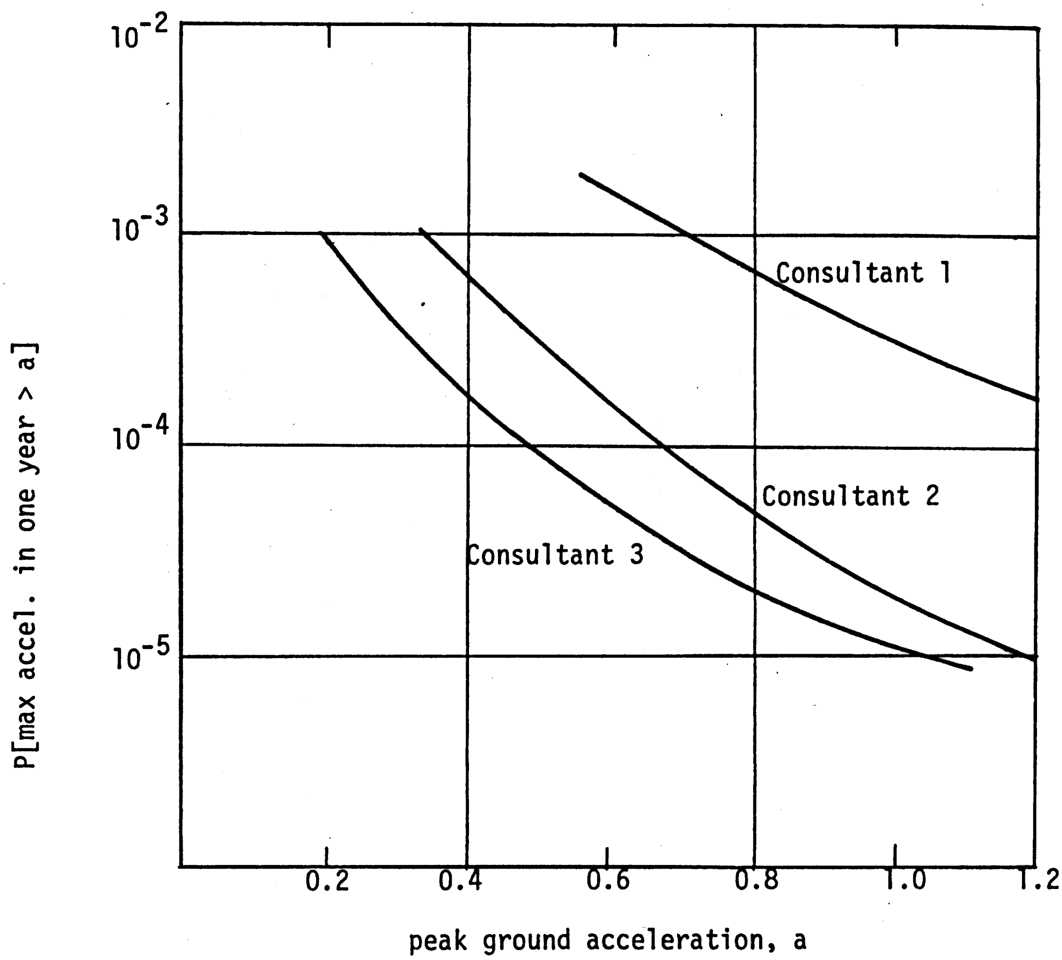


FIGURE 6.3 Seismic hazard curves for Diablo Canyon (Cornell, 1980).

Although the sample is very small, the range of estimates corresponds to a factor of about 10 on probability given a and of about 3 on acceleration given the probability level.

Another useful source of information on the bias and error of subjective quantification and updating of uncertainty is the data collected by experimental psychologists (Phillips, Mays, and Edwards, 1966; Winkler, 1967; Vlek, 1973; Staël Von Holstein, 1971; Rapoport and Wallsten, 1972; to name but a few). Distortions due to conservatism and hedging are rather well documented; so are techniques for controlling them, such as feedback and cross-checking.

7. CONCLUSIONS

It is just natural that first theories and models are developed and then their limitations are understood. In the case of seismic hazard analysis, the first step is essentially completed but not the second, so that quoting from a contemporary survey of the field (Cornell, 1980), "the probabilistic procedures of applied seismic hazard analysis are at the moment very stable."

By contrast, understanding of the accuracy of even the standard methods is at best incomplete. The very definition of accuracy (with respect to what?) has never been given in clear, unequivocal terms. Studies of the subject are fragmentary and uncoordinated; they look at different sources of error and use different methods and language in a rather unscientific fashion. Besides the conceptual need to give order to the subject, there is the practical necessity of deciding what to do about uncertainty on the true hazard. While waiting for an answer, there is a consensus, de facto, that one should be conservative, although nobody has come up with a measure of conservatism, even less with a rationale.

The purpose of this report is to gather available information on errors in seismic hazard analysis and, more important, to provide a conceptual framework to the subject. Toward the latter end, it was found necessary to first define a "true probabilistic hazard" against which accuracy could be measured. An objective definition of hazard has been given in Sec. 2, as the relative frequency in time with which earthquake events of engineering interest occur at the site. The body of the report--Secs. 4, 5, and 6--deals with errors in the estimation of true hazard--statistical errors in Secs. 4 and 5, judgmental errors in Sec. 6. Specifically, Sec. 4 looks at errors from modeled uncertainty (on parameters or on alternative hazard procedures) whereas Sec. 5 deals with unmodeled errors from making wrong assumptions about the relevant laws of nature (form of the true hazard function or characteristics of seismicity). Errors of the latter type are more elusive and often more severe than those that are explicitly modeled. Hence, judging accuracy only from modeled uncertainty (e.g., about parameters such as the Richter b slope or the upper bound magnitude) may give a false sense of confidence in a model.

With the increasing use of nonstatistical information in the form of professional judgment, the study of errors in the subjective estimation and updating of uncertainty becomes part of the accuracy problem. A related issue is that of how to combine hazard estimates from different seismologist experts. Reflections on this problem are offered in Sec. 6.

It is realized that the engineering accuracy and value of a model cannot be quantified only in terms of errors. Consequences of errors rather than errors are important, uncertainty on seismic risk rather than on seismic hazard. Although most of the literature to date deals with the latter topic, a section (Sec. 3) has been included here on seismic risk and its dependence on parameter uncertainty.

It is apparent that more research is needed on virtually every aspect of this complex problem and that it is of utmost importance that it be carried out with much more coordination than we have seen in the past.

REFERENCES

Aitchison, J. (1964), "Bayesian Tolerance Regions," J. Royal Statist. Soc., B, Vol. 26, pp. 161-210.

Aitchison, J. and Dunsmore, I. R. (1975), Statistical Prediction Analysis, Cambridge University Press, London.

Aitchison, J. and Schulthorpe, D. (1965), "Some Problems of Statistical Prediction," Biometrika, Vol. 52, Nos. 3 and 4. p. 469.

Banon, H. and Veneziano, D. (1981), "Seismic Safety of Reinforced Concrete Members and Structures," Earthquake Engineering And Structural Dynamics, (to appear).

Benjamin, J.R. (1974), "Probabilistic Decision Analysis Applied to Earthquake Damage Surveys," Research Report, Earthquake Engineering Research Institute.

Chew, V. (1968), "Simultaneous Prediction Intervals," Technometrics, Vol. 10, pp. 323-330.

Cornell, C.A. (1968), "Engineering Seismic Risk Analysis," Bull. Seism. Soc. Am., 58, pp. 1583-1606.

Cornell, C.A. (1971), "Probabilistic Analysis of Damage to Structures under Seismic Loads," Chapter 27 of Dynamic Waves in Civil Engineering, D.A. Howells, I.P. Haigh, and C. Taylor, Eds., Wiley Interscience, N.Y.

Cornell, C.A. (1980), "Probabilistic Seismic Hazard Analysis: A 1980 Assessment," Proc., US-Yugoslavia Earthquake Engineering Research Seminar, Skopje, Yugoslavia, June 30-July 3.

Cornell, C.A. and Vanmarcke, E.H. (1969), "The Major Influences on Seismic Risk," Proc. Fourth World Conf. on Earthquake Engineering, Santiago, Chile.

Cornell, C.A. and Merz, H.A. (1975), "Seismic Risk Analysis of Boston," J. of the Structural Division, ASCE, Vol. 101, No. ST10, pp. 2027-2043.

Cornell, C.A., Banon, H., and Shakal, A. (1979), "Seismic Motion and Response Prediction Alternatives," Earthquake Engineering and Structural Dynamics, Vol. 7, pp. 295-315.

Dalkey, N.C. (1975), "Group Consensus Distribution," in Turoff, M., et.al. (eds.), The Delphi Approach, Addison-Wesley.

Dames & Moore (1980), "Seismic Ground Motion Hazard at La Crosse Boiling Water Reactor, Wisconsin," Draft Report.

de Groot, M. (1974), "Reaching a Consensus," Journal of the American Statistical Association, 69, pp. 118-1121.

Der Kiureghian, A. and Ang, A. H.-S. (1975), "A Line Source Model for Seismic Risk Analysis," Technical Report, Structural Research Series, No. 419, Department of Civil Engineering, University of Illinois, Urbana, Illinois, October.

Der Kiureghian, A. and Ang, A. H-S. (1977), "A Fault-Rupture Model for Seismic Risk Analysis," Bull. Seism. Soc. Am., 67, pp. 1173-1194.

Donovan, N.C. and Bornstein, A.E. (1978), "Uncertainties in Seismic Risk Procedures," J. of the Geotechnical Engineering Division, ASCE, Vol. 104, No. GT7, pp. 869-887.

Draper, N. and Smith, H. (1966), Applied Regression Analysis, John Wiley & Sons, New York.

Esteva, L. (1970), "Seismic Risk and Seismic Decision," in Seismic Design for Nuclear Power Plants, R.J. Hansen, M.I.T. Press, Cambridge, Mass.

Esteva, L. (1976), "Seismicity," Chapter 6 of Seismic Risk and Engineering Decision, E. Rosenblueth and C. Lomnitz, Eds., Elsevier Scientific Publishing Co., Amsterdam.

Fine, T. L. (1973), Theories of Probability: An Examination of Foundations, Academic Press, New York.

Grigoriu, M. (1976), "A Decision Theoretic Approach to Model Selection for Structural Reliability," Ph.D. Thesis, Dept. of Civil Engineering, M.I.T., Cambridge, Mass.

Grigoriu, M., Veneziano, D., and Cornell, C.A. (1979), "Probabilistic Modeling as Decision Making," J. of the Engineering Mechanics Division, ASCE, Vol. 105, No. EM4, pp. 585-596.

Gupta, I.N. and Nuttli, O.W. (1976), "Spatial Attenuation of Intensities for Central U.S. Earthquakes," Bulletin of the Seismological Society of America, 66(3), pp. 743-751.

Guttman, I. (1970), Statistical Tolerance Regions: Classical and Bayesian, Griffin's Statistical Monographs & Courses, No. 26, Stuart Ed., Hafter Publ. Co., London.

Hasselman, T.K. and Simonian, S.S. (1980), "Structural Uncertainty in Seismic Risk Analysis," NUREG/CR-1560, UCRL-15218, Lawrence Livermore Laboratory, Livermore, California.

Hofstetter, E.M. (1964), "Random Processes," Chapter 3 of The Mathematics of Physics and Chemistry, Vol. 2, H. Margenau and G.M. Murphy, Eds., Van Nostrand, Princeton, New Jersey.

Kameda, H. and Ozaki, Y. (1979), "A Renewal Process Model for Use in Seismic Risk Analysis," Memoirs of the Faculty of Engineering, Vol. XLI, Part 1, Kyoto University.

Kaplan, S. (1980), "On the Method of Discrete Probability Distributions in Risk and Reliability Calculations," Pickard, Lowe and Garrick, Inc., Irvine, California, October.

Kaplan, S. and Garrick, B.J. (1980), "On the Quantitative Definition of Risk," Pickard, Lowe and Garrick, Inc., Irvine, California, June.

Kiremidjian, A., Anagnos, T., and Shah, H.C. (1980), "A Time and Space Dependent Model for Earthquake Occurrences," Seism. Soc. Am. Meeting, Seattle, Washington, April.

Knopoff, L. and Kagan, Y. (1977), "Analysis of the Theory of Extremes as Applied to Earthquake Problems," J. Geoph. Res., Vol. 82, No. 36, December.

Long, L.T. (1976), "Speculations Concerning Southeastern Earthquakes, Mafic Intrusions, Gravity Anomalies, and Stress Amplification," Earthquake Notes, 47, pp. 29-35.

Lucas, J.R. (1970) The Concept of Probability, Clarendon Press, Oxford.

Mann, O.C. (1974), "Regional Earthquake Risk Study; Appendix D2," Technical Report for MATCOG/MDDD, M & H Engineering and Memphis State University.

Matheron, G. (1971), The Theory of Regionalized Variables and Its Applications, Les Cahiers du Centre de Morphologie Mathematique de Fontainbleau, No. 5.

McGuire, R.K. and Shedlock, K.M. (1980), "Statistical Uncertainties in Seismic Hazard Evaluation in the United States," Unpublished manuscript.

McKeown, F.A. (1978), "Hypothesis: Many Earthquakes in the Central and Southeastern United States Are Causally Related to Mafic Intrusions Bodies," J. Research U.S. Geol. Survey, Vol. 6, No. 1, January-February.

Merz, H.A. and Cornell, C.A. (1973), "Aftershocks in Engineering Seismic Risk Analysis," Proc., Fifth World Conf. on Earthquake Engineering, Rome, Italy, June.

Morris, P.A. (1974), "Decision Analysis Expert Use," Management Science, 20, pp. 1233-41.

Newmark, N.M. (1974), "Comments on Conservatism in Earthquake Resistance Design," presented at the U.S. Atomic Energy Commission, September.

Patwardhan, A.S., Kulkarni, R.B., and Tocher, D. (1978), "A Semi-Markov Model for Characterizing Recurrence of Great Earthquakes," EHRP Conf. on Methodology for Defining Seismic Gaps and Soon-to-Break Gaps, MIT, Cambridge, Mass., May 25-27.

Philips, L.D., Hays, W.L., and Edwards, W. (1966), "Conservatism in Complex Probabilistic Inference," Institute of Electrical and Electronics Engineers, Transactions, Vol. HFE-7, pp. 7-18.

Proceedings of International Conferences on Structural Mechanics in Reactor Technology, Berlin, 1971, 1973; London, 1975; San Francisco, 1977.

Proshan, F. (1953), Confidence and Tolerance Intervals for the Normal Distribution, J. Am. Statist. Assoc., Vol. 48, pp. 550-564.

Raiffa, H. and Schlaifer, R. (1961), Applied Statistical Decision Theory, Graduate School of Business Administration, Harvard University, Cambridge, Mass.

Rapoport, A., and Wallsten, T.S. (1972), "Individual Decision Behavior," Annual Review in Psychology, Vol. 23, pp. 131-176.

Schlien, S. and Toksoz, M. (1970), "A Clustering Model for Earthquake Occurrences," Bull. Seism. Soc. of Am., Vol. 60, No. 6, pp. 1765-1787.

Schumaker, J. (1977), "Statistical Methodologies in Seismic Risk Analysis," Research Report R 77-18, Dept. of Civil Engineering, MIT, Cambridge, Mass.

Staël Von Holstein, C.S. (1971), "Two Techniques for Assessment of Subjective Probability Distributions--An Experimental Study," Acta Psychologica, Vol. 35, pp. 478-494.

TERA Corp. (1979, 1980), "Seismic Hazard Analysis; I: A Methodology for Eastern United States; II: Solicitation of Expert Opinion; III: Site Specific Spectra Results," Berkeley, California.

TERA Corp. (1980), "Seismic Hazard Analysis: Site-Specific Response Spectra-Sensitivity Results," Berkeley, California, May.

Varmarcke, E.H. (1971), "Example of Expected Discounted Future Cost Computation," Internal Study Report No. 2, Seismic Design Decision Analysis, Dept. of Civil Engineering, MIT, Cambridge, Mass.

Veneziano, D. (1974), "Statistical Estimation and Prediction in Probabilistic Models, with Application to Structural Reliability," Ph.D. Thesis, Dept. of Civil Engineering, MIT, Cambridge, Mass.

Veneziano, D. (1975), "Probabilistic and Statistical Models for Seismic Risk Analysis," Research Report R75-34, Dept. of Civil Engineering, MIT, Cambridge, Mass.

Veneziano, D., Schumacher, J. and Cornell, C.A. (1977), "Statistical Errors in Seismic Risk Analysis," ASCE Spring Convention, Pittsburgh, April.

Veneziano, D. and Chung, D.H. (1980), "Interpretation of Seismic Hazard," Unpublished manuscript.

Vere-Jones, D., and Davies, R.B. (1966), "A Statistical Survey of Earthquakes in the Main Seismic Region of New Zealand: Part 2-Time Series Analyses," N.Z.J. Geol. Geophys., Vol. 9, pp. 251-284.

Vlek, C.A.J. (1973), "Coherence of Human Judgement in a Limited Probabilistic Environment," Organization Behavior and Human Performance, Vol. 9, pp. 460-481.

Walley, P. (1976), "The Estimation of Earthquake Risk in New Zealand," Technical Report No. 49, Dept. of Scientific and Industrial Research, Wellington, New Zealand.

Whitman, R.V. (1973), "Damage Probability Matrices for Prototype Buildings," Research Report R73-57, Dept. of Civil Engineering, MIT, Cambridge, Mass.

Whitman, R.V. and Hong, S.-T. (1973), "Data for Analysis of Damage to High Rise Buildings in Los Angeles," Optimum Seismic Protection for New Building Construction in Eastern Metropolitan Areas; Internal Study Report No. 32, Dept. of Civil Engineering, MIT, Cambridge.

Winkler, R.L. (1967), "The Assessment of Prior Distributions in Bayesian Analysis," Journal of the American Statistics Association, Vol. 62, pp. 776-800.

Winkler, R.L. (1968), "The Consensus of Subjective Probability Distributions," Journal of the American Statistical Association, 62, pp. 776-800.

Wood, E.S., Rodriguez-Iturbe, I., and Schaake, J.C. (1974), "The Methodology of Bayesian Inference and Decision Making Applied to Extreme Hydrologic Events," Technical Report No. R 74-8, Dept. of Civil Engineering, MIT, Cambridge, Mass.