



## SOURCES, NATURE, AND IMPACT OF UNCERTAINTIES ON CATASTROPHE MODELING

Patricia GROSSI<sup>1</sup>

### SUMMARY

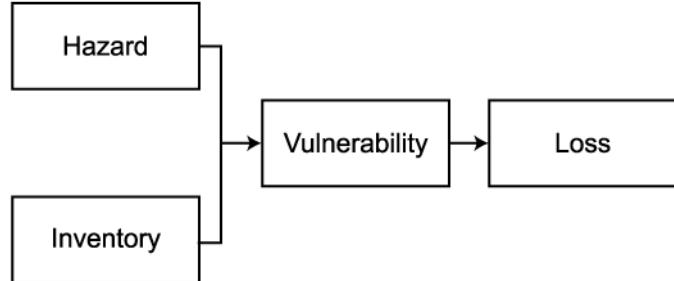
The information presented in this paper is one part of a larger discussion on the inner workings of catastrophe models and how they assess risk from natural hazards. In particular, this discussion focuses on the role of uncertainty in catastrophe models by examining the sources, nature, and impact of uncertainty on assessing natural hazard risk. An illustrative example of assessing earthquake risk in South Carolina enables one to understand how uncertainty in the modeling process affects the risk between stakeholders. The larger discussion, to be published in 2004, brings together researchers at the Wharton School of the University of Pennsylvania with the collective wisdom of three leading modeling firms (Risk Management Solutions, EQECAT, and AIR Worldwide). *Catastrophe Modeling: A New Approach to Managing Risk* will be available from Kluwer Academic Publishers.

### INTRODUCTION

Catastrophe modeling is a complex tool to assess the risk from natural hazards. The four components of hazard, inventory, vulnerability, and loss (Figure 1) require information from a range of sources and the expertise of an array of professionals. Natural hazard, engineering and economic data are the foundation of catastrophe models. Limitations in data and assumptions about the model's parameters, in the hazard, inventory, and vulnerability modules, affect a catastrophe model's loss estimates and the certainty associated with these estimates. This paper explores the sources, nature, and impact of uncertainties in a catastrophe model. Prevalent methods to represent and quantify uncertainty through the components of the catastrophe model are discussed. Finally, the impact of uncertainty on exceedance probability (EP) curves used by risk managers to quantify their catastrophe risk potential is illustrated by examining potential losses to residential property from earthquakes in Charleston, South Carolina. Quantification and classification of uncertainty provides opportunities to reduce risk. With accurate measures of uncertainty, stakeholders can potentially lower the cost of dealing with catastrophe risk. Furthermore, since the risk affects stakeholders in dissimilar ways, the robustness of a risk management strategy can be made clear to each stakeholder if uncertainty is delineated.

---

<sup>1</sup> Risk Management Solutions, Inc., Newark, CA, USA. Email: patricia.grossi@rms.com



**Figure 1: Catastrophe model components.**

## CLASSIFICATIONS OF UNCERTAINTY

There is a great deal of information needed to develop the hazard, inventory, vulnerability, and loss components of a catastrophe model. Therefore, all stakeholders in the management of risk value new information regarding these modules. For example, an insurer values additional information on the likelihood of disasters and potential damage to properties in its portfolio in order to more accurately manage the risk. Local government officials value a thorough understanding of the hazards in the region in order to plan for emergency response and recovery efforts following a disaster. Model developers value any additional information to validate and calibrate their catastrophe models.

Since catastrophe modeling is a fairly new field of application, there are no historical classifications of catastrophe modeling uncertainty, per se. However, building on the concepts from probabilistic hazard analyses, uncertainty can be characterized as either aleatory or epistemic in nature [1]. Aleatory uncertainty is the inherent randomness associated with natural hazard events, such as earthquakes, hurricanes, and floods. It cannot be reduced by the collection of additional data. In contrast, epistemic uncertainty is the uncertainty due to lack of information or knowledge of the hazard. Unlike aleatory uncertainty, epistemic uncertainty can be reduced by the collection of additional data.

While the advantage of differentiating between aleatory and epistemic uncertainty in an analysis is clear (only epistemic uncertainty can be reduced), the necessity of distinguishing between aleatory and epistemic uncertainty is not. “Epistemic and aleatory uncertainties are fixed neither in space...nor in time. What is aleatory uncertainty in one model can be epistemic uncertainty in another model, at least in part. And what appears to be aleatory uncertainty at the present time may be cast, at least in part, into epistemic uncertainty at a later date” [2]. Therefore, developers of catastrophe models do not necessarily distinguish between these two types of uncertainty; instead, model developers concentrate on not ignoring or double counting uncertainties and clearly documenting the process in which they represent and quantify uncertainties.

## SOURCES OF UNCERTAINTY

Limited scientific knowledge, coupled with a lack of historical data, leave open several possible and competing explanations for the parameters, data, and mathematical models underlying each of the components in a catastrophe model. Simply put, the science and impact of natural hazards are not completely understood; in addition, the cross-disciplinary nature of a catastrophe model leads to complexity. Experts in seismology or meteorology who model the hazard must interact with structural engineers who model the vulnerability; similarly structural engineers who model the vulnerability must interact with actuaries who model the loss. Basically, as each discipline’s modeling assumptions are added to the process, more uncertainty is added to the estimates.

In catastrophe modeling, epistemic and aleatory uncertainties are reflected in the four basic components of a model. Aleatory uncertainty is reflected via probability distributions. The frequency of a hazard occurrence and the fragility of a building are examples of aleatory uncertainty. Since the exact time of occurrence in the future and the precise level of structural damage cannot be known in advance of a hazard event, the recurrence rate and the vulnerability of the inventory exposed to the natural hazard are characterized using probability distributions. Similarly the capacity of individual structural elements of a building during a severe event and the resulting cost of repair cannot be determined beforehand. Probability distributions are also used to characterize these parameters in a catastrophe model.

A larger issue in quantifying uncertainty is the lack of data for characterizing the four components in a catastrophe model. For example, the recurrence of earthquake events on fault sources can be modeled using a magnitude-frequency model [3], a characteristic earthquake model [4], or a combination of both models. In California, estimates of ground shaking probabilities on certain fault segments are established by combining the two recurrence models for earthquake magnitude-frequency distributions [5]. Historical earthquake records are used to establish a recurrence curve, or the Gutenberg-Richter relationship, for the smaller magnitude events, while geologic data (most importantly, a fault's slip rate) is used to estimate the recurrence of the larger, characteristic events. The availability of seismological data describing earthquake occurrence in California for only a few hundred years makes the updating of the recurrence distributions problematic. When more data become available, in the form of fault slip rates or seismograph recordings, these relationships could potentially be improved.

The deficiency of information regarding repair costs and business interruption costs affect the accuracy of the loss component of a catastrophe model. For example, the increased cost to repair or rebuild after an event is often taken into account using a demand surge adjustment. This is simply the percentage increase in costs due to the limited supply of construction material and labor immediately following a disaster. Further, due to the growing understanding of indirect losses, estimates of business interruption costs to commercial property owners are continually validated and calibrated with the latest loss information.

Another source of epistemic uncertainty in a catastrophe model is the lack of available data to create the GIS databases within the modeling software. For any model, recognizing the importance of input data is essential. The GIGO (garbage in, garbage out) principle holds irrespective of how advanced or state-of-the-art a model may be. GIS maps of hazard sources and geologic features characterize hazard and GIS maps of structure locations characterize inventory for an earthquake model.

An incomplete description of a hazard source or the geology (e.g., underlying soil) can cause erroneous results. For example, having accurate information on the underlying soil in a region is very important. A structure built on rock-like material is likely to sustain much lower losses compared to a structure built on soft clay-like material. Inaccurate information on soil conditions can lead to large errors in estimation of loss due to an earthquake. In fact, past observations from earthquakes confirm that soil condition plays a very important role in building performance. As expected, buildings on soft ground or steep slopes usually suffer more significant damage in comparison to those on firm and flat ground. Since soil condition may vary dramatically within a small area, such as the Marina District in San Francisco (where soil conditions vary from bay mud to rock site), using zip code to identify a location may not be sufficiently accurate. At a particular location, high-resolution geocoding should be used as it can pin down the accurate soil condition more accurately.

Partial information on a structure's characteristics can also result in an inaccurate estimate of future damage. For example, most structural engineers would agree that the construction type, age, height, occupancy, assessed value and the location of a structure are needed – at a minimum – for the inventory

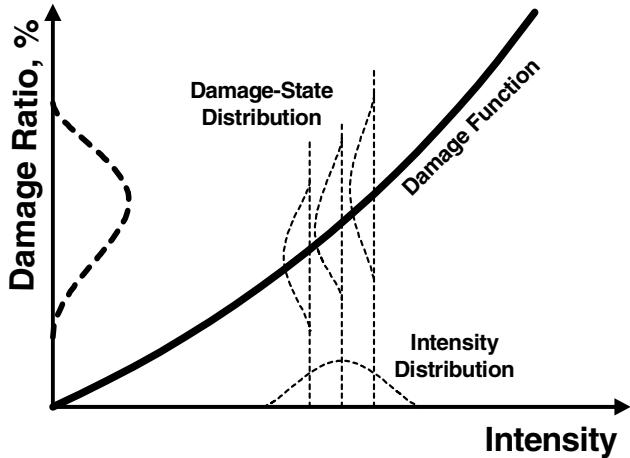
component of a catastrophe model. If more specific information regarding the structure such as its location relative to other structures and previous damage to the structure were available, a more accurate estimate of damage or vulnerability would result.

An additional source of epistemic uncertainty in the modeling process is lack of accurate information on true market value of the properties under consideration. For determining the appropriate coverage limit, many residential policies use property tax assessment data, which are generally outdated and under-valued. Under-valued exposures will result in under-estimating potential loss. For example, suppose a home's property value is assessed at \$600,000 when its true worth is \$1 million. Furthermore, suppose it is insured with a 15% deductible and full coverage based on the lower assessed value. If an earthquake occurs and causes major damage and the cost to repair the structure is 35% of true value of the home, the resulting monetary loss is \$350,000. A \$600,000 insurance policy with a 15% deductible translates to the homeowner being responsible for \$90,000, with the insurer covering the remaining \$260,000 of the loss. If the insurance coverage had been based on the home's true worth of \$1 million, the homeowner would have to cover the first \$150,000 of the loss and the insurer would only have claim payments of \$200,000.

Incomplete or inaccurate information on an inventory's description is a concern not only to insurers but also to all risk management stakeholders. To improve on the amount of such information available, an effort to document the types of housing structures worldwide was initiated in 2000 to assess the vulnerability of the world's population to earthquake hazard. Under the guidance of the Earthquake Engineering Research Institute (EERI) and the International Association of Earthquake Engineering (IAEE), the World Housing Encyclopedia has a web-based listing of housing construction types from earthquake-prone countries around the world [6]. In addition, the Institute for Business and Home Safety relies on INCAST, a data inventory tool used in conjunction with the HAZUS catastrophe model, to store inventory information on the homes that are a part of their "Fortified...for safer living" program. These homes are reinforced to withstand many natural hazards, including high winds, wildfire, flood, hail, and earthquake.

Epistemic uncertainty is also found in the use of laboratory testing (e.g., shake table tests) and expert opinion to develop the vulnerability component of a catastrophe model. For a portfolio risk assessment, damage functions (e.g., Figure 2) have traditionally been constructed using these sources along with damage surveys of actual structures. Given that laboratory testing has been restricted to certain types of structural materials, there is a limited understanding of how other materials withstand lateral loading.

In the earliest versions of catastrophe models, damage ratios, or the ratios of repair costs to the replacement cost of a building, were estimated using the ATC-13 report of Earthquake Damage Evaluation Data for California [7]. This report was generated using the Delphi method of collecting information from a group of experts [8]. In this method, a series of questionnaires interspersed with controlled opinion feedback resulted in a group judgment. In the ATC-13 study, 71 earthquake engineering experts were asked to indicate their low, best, and high estimates of damage ratios for 78 types of structures subject to earthquakes with Modified Mercalli Intensity (MMI) levels of VI through XII. Catastrophe model developers used these estimates in their earliest versions of their earthquake loss software, skewing estimates of damage due to the use of the Delphi Method. More recent models employ cost models that translate estimates of physical damage into direct monetary loss rather than depending on damage ratios.



**Figure 2: Illustration of a typical damage function.**

### REPRESENTING AND QUANTIFYING UNCERTAINTY

Guidelines do exist for identifying the sources of uncertainty and incorporating them into catastrophe models. The Senior Seismic Hazard Analysis Committee (SSHAC) Report is a comprehensive study addressing this issue and the use of expert opinion in a probabilistic seismic hazard analysis [1]. This report can also be used for the incorporation of uncertainty of other natural hazards. Additionally, guidelines set forth by the Environmental Protection Agency (EPA), requiring that all risk assessments possess the core values of "transparency, clarity, consistency, and reasonableness," are relevant for the modeling of natural hazards [9].

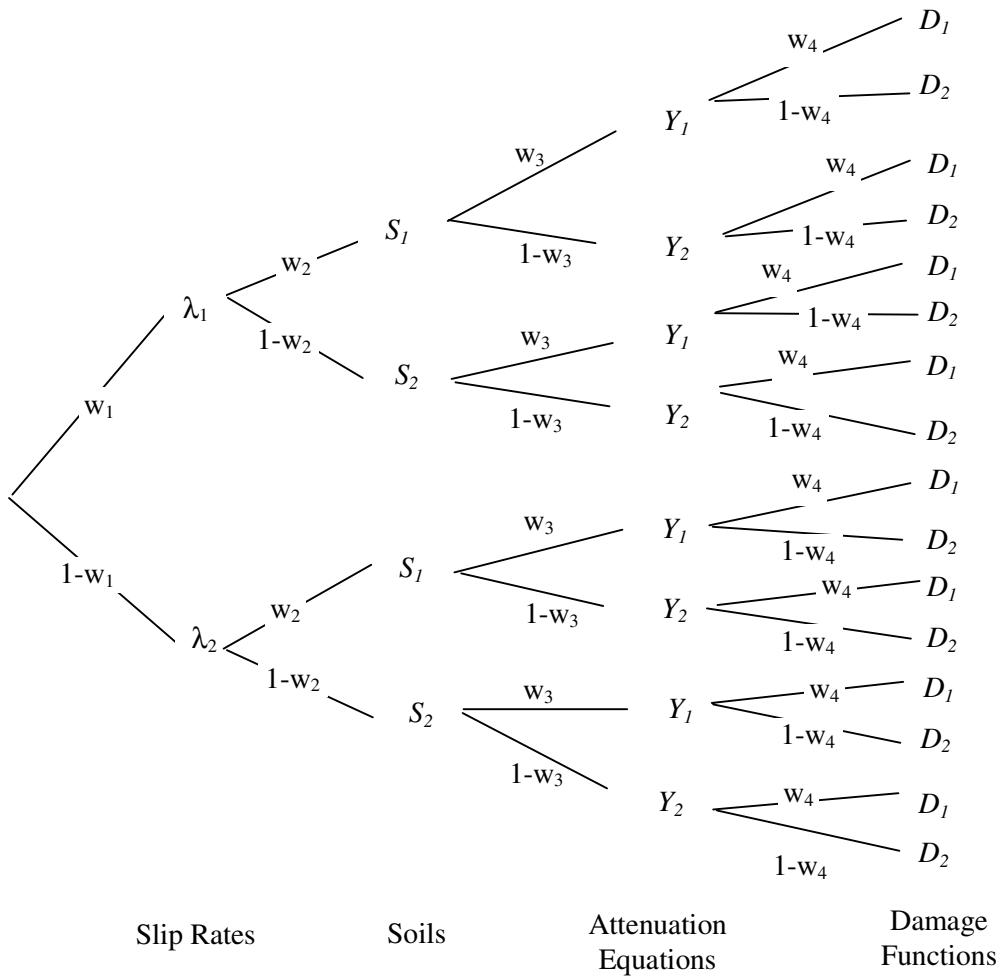
The most common methods for incorporating uncertainty into catastrophe modeling are logic trees and simulation techniques. These two methods are standard approaches for quantifying and propagating uncertainty when there is intrinsic aleatory uncertainty, lack of consensus among experts, and lack of data used to estimate parameters.

#### Logic Trees

In the logic tree approach, alternative parameter values or mathematical relationships are identified within the catastrophe model, relative weighting schemes are assigned to each alternative, and estimates of parameters or relationships are calculated using a weighted, linear combination of the outcomes. Weighting schemes are numerous, with the weights representing the credibility of that alternative in relation to the available data. For example, one can use equal weights, weights proportional to the ranking of alternatives, or weights based on some comparison of previously assessed estimates with actual outcomes. Weights are often established through the use of expert opinion, and therefore, are biased towards an expert's judgment.

Figure 3 depicts a simple example of how a logic tree can be used in a catastrophe model. Suppose that there is an earthquake fault that generates a characteristic magnitude event. This event is estimated using a recurrence model with two alternatives for the fault's slip rate,  $\lambda_1$  and  $\lambda_2$ , weighted  $w_1$  and  $1-w_1$ , respectively. Next, suppose a single family residential structure is the only structure to be assessed in the inventory. However, there is a lack of consensus regarding the type of underlying soil at the site. Thus, there are two alternatives for the soil parameter, denoted  $S_1$  and  $S_2$  with respective weights  $w_2$  and  $1-w_2$  in Figure 3.

In the next branch of the logic tree, the two estimates of recurrence for a characteristic magnitude event and the two alternatives for site-specific soils are combined with two competing attenuation equations describing the rate at which the amplitude of the seismic waves decreases as the waves propagate outward from the source of the rupture. For example, the Frankel et al [10] attenuation relationship and the Toro et al [11] relationship can be used as two competing models of strong ground motion in the Central and Eastern United States. These two models, denoted  $Y_1$  and  $Y_2$  in Figure 3, are weighted  $w_3$  and  $1-w_3$ , respectively. This combination results in estimates of earthquake ground motion for certain magnitude events at a certain frequency of occurrence, under certain site conditions, and at certain distances from the event's epicenter.



**Figure 3: Logic tree approach to catastrophe modeling.**

Finally, these ground motion estimates are combined with two competing models for damage functions, one created using expert opinion and one based on laboratory testing. These functions,  $D_1$  and  $D_2$ , relate the expected damage state of the residential building (minor, moderate, severe damage, or collapse) to the level of ground motion at the site. Each is weighted accordingly, denoted  $w_4$  and  $1-w_4$  in Figure 3. The final results of this simple example are sixteen calculations of structural damage to a single-family dwelling based on alternative assumptions of characteristic fault slip rates, underlying soils, and empirical attenuation models. As is evident, the costs of repair have not yet been incorporated.

The logic tree approach to incorporating uncertainty is utilized often in practice because of its tractability and its usefulness as a tool to communicate risk to stakeholders. While a set of results grows with each alternative assumption added to the analysis, advances in computing power allow the handling of large databases; therefore, both parameter and model alternatives can be identified within this type of approach. Although the above example shows two alternatives at each branch, a larger (yet finite) number of alternatives can be considered, as is typically the case in a catastrophe model.

### **Simulation Techniques**

Simulation is a method for learning about a real system by experimenting with a model that duplicates the essential behavior of the system. It is one of the most widely used quantitative approaches to decision making. In contrast to a logic tree, which requires a set of simplifying assumptions, simulation can model extremely complex processes. An uncertain parameter is represented by a discrete or continuous probability distribution, multiple simulations are run which sample from the distribution, and the analyses are completed using these sample values. The results are statistically analyzed to estimate important performance measures of the system. In the case of catastrophe modeling, a performance measure is, for example, exceedance probability loss.

Although most distributions in catastrophe modeling are continuous, a simulation using a discrete distribution is presented here for simplicity. Suppose that a single-family residential structure is subject to an earthquake hazard and five levels of damage states are defined (none, minor, moderate, severe, or collapse) in a catastrophe model. Suppose further that damage functions are available that represent the probability of being in, or exceeding, a certain damage state level given a certain level of ground shaking. Now suppose that the residential insurer wants a probabilistic estimate of being in a certain damage state given that the peak ground acceleration is 0.25 g.

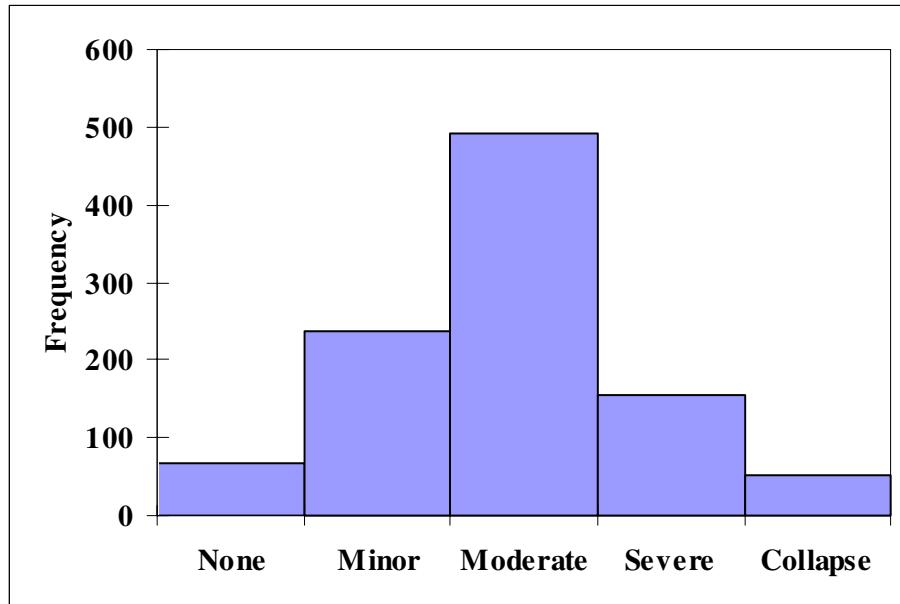
Simulation can be used to generate this probability distribution. First, the probability of being in one of the five damage states is calculated based on the given set of damage functions, indicated by damage state probability in Table 1. For example, there is a 5% probability that there will be no damage and a 7% probability that the building will collapse. In this case, an arbitrary range from 00-99 (100 digits) is used, with 5% representing the probability of having no damage (00-04), 24% representing minor damage (05-28), 48% representing moderate damage (29-76), 16% representing severe damage (77-92), and 7% representing collapse of the structure (93-99). Then the cumulative probabilities are calculated for the ordered damage states and random numbers are assigned in proportion to these cumulative probabilities as shown in Table 1.

**Table 1: Simulation example in catastrophe modeling**

| <b>Damage State</b> | <b>Damage State Probability</b> | <b>Cumulative Probability</b> | <b>Random Number Lower Bound</b> | <b>Random Number Upper Bound</b> |
|---------------------|---------------------------------|-------------------------------|----------------------------------|----------------------------------|
| None                | 0.05                            | 0.05                          | 00                               | 04                               |
| Minor               | 0.24                            | 0.29                          | 05                               | 28                               |
| Moderate            | 0.48                            | 0.77                          | 29                               | 76                               |
| Severe              | 0.16                            | 0.93                          | 77                               | 92                               |
| Collapse            | 0.07                            | 1.00                          | 93                               | 99                               |

To start the simulation, a random number between 00 and 99 is generated. Based on the resulting value, a damage state is projected. For example, if the random number is 36, the structure has moderate damage; if the random number is 21, the structure sustains minor damage. This random number generation is

repeated, for example, 1,000 times and the levels of damage are stored. At the end of the 1,000 sample runs, a histogram of the sample damage state frequencies is created (Figure 4). This histogram is an approximation to the distribution of damage, given a level of peak ground acceleration.



**Figure 4: Histogram of damage state frequency for 1000 simulation runs.**

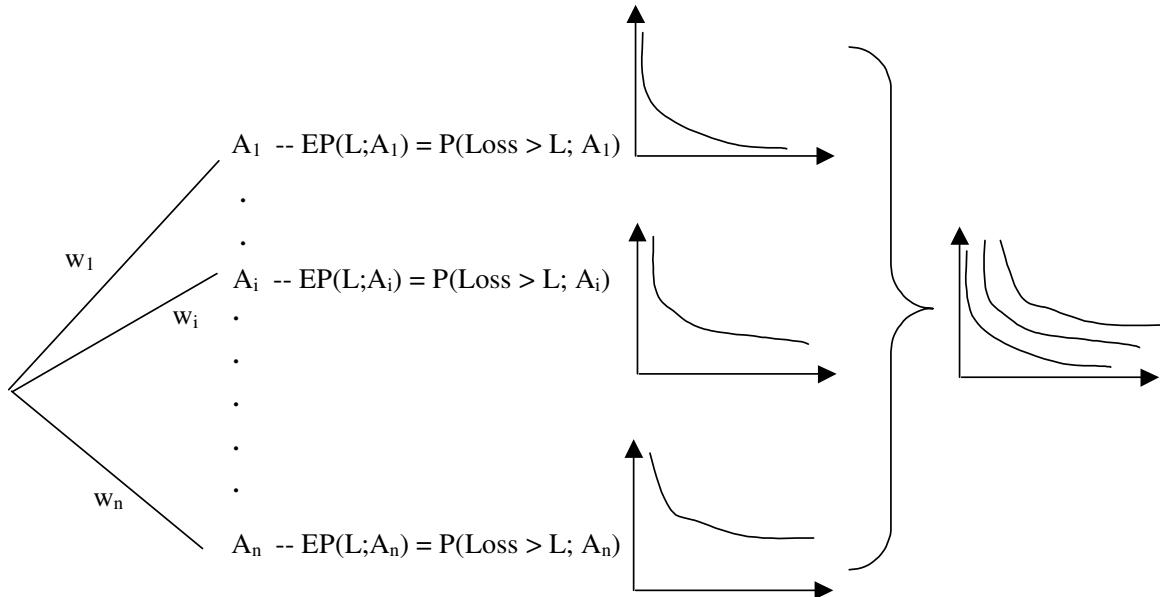
While this is a simple example of a Monte Carlo simulation (and actual simulations in catastrophe modeling are much more complicated), it should be noted that this type of modeling is computationally intensive and requires a large number of samples. If the time and computer resources required to run a full-blown simulation is prohibitively expensive, a degree of computational efficiency can be found through the use of Modified Monte Carlo methods, such as Latin Hypercube Sampling, that sample from the input distribution in a more efficient manner [12]. In this way, the number of necessary runs compared to the Monte Carlo method is significantly reduced.

### Uncertainty and the Exceedance Probability Curve

An exceedance probability curve is a graphical representation of the probability that a certain level of loss will be exceeded over a future time period. A widely used technique to create an exceedance probability curve in a catastrophe model is a combination of a logic tree with Monte Carlo simulation. Building on the simple examples presented earlier, each branch of the logic tree represents an alternative that samples from a probability distribution rather than assuming a simple point estimate alternative. For example, consider the competing attenuation equations for ground motion presented earlier, denoted  $Y_1 = F_1(f, M, r, Source, Site)$  and  $Y_2 = F_2(f, M, r, Source, Site)$ . Instead of using the mean estimates of ground motion amplitude based on these functions for each branch of the logic tree, Monte Carlo methods can be used to sample from the attenuation functions along the branches of the tree.

This blended approach allows the creation, in a systematic way, of a set of curves that represent various confidence levels in exceedance probabilities. For example, suppose that there are a set of assumptions,  $A_1, A_2 \dots A_n$ , which represent an exhaustive set of all possible assumptions about the parameters, data, and mathematical models needed to generate an exceedance probability curve in a catastrophe model. Further, suppose that each set of assumptions is an alternative on one branch of a logic tree and each logic tree branch results in an EP curve that is generated when the assumptions  $A_i$  are made, characterizing the loss  $L$ , as shown in Figure 5 (i.e.,  $EP(L, A_i) = P(Loss > L, A_i)$ ). If each of the sets of assumptions are weighted

with subjective probabilities,  $w_1, w_2 \dots w_n$ , that add up to one and the assumptions,  $A_1, A_2 \dots A_n$ , give rise to a monotonic ordering of their respective EP curves, the mean, median and a confidence interval for the resulting collection of EP curves can be defined.



**Figure 5: Logic tree and simulation to create a set of exceedance probability curves.**

Given the complexity of catastrophe modeling and this discussion of the sources and techniques to incorporate uncertainty in a model, it is not surprising that competing catastrophe models will generate different EP curves for the same portfolio of structures. When first used in practice, the degree to which these curves could differ was surprising to the users of catastrophe models. With more experience, a user expects a range of possible EP curves.

### A CASE STUDY IN UNCERTAINTY

In the summer of 1999, a meeting was held among representatives of Risk Management Solutions, EQECAT, AIR Worldwide and The Wharton School to discuss a sensitivity analysis regarding catastrophe models' estimates of earthquake loss [13]. In this section, a case study of earthquake hazard in Charleston, South Carolina is presented using data from four catastrophe models: three models developed by the three modeling firms involved in this study (denoted Model A, Model B, and Model C), along with the U.S. federal government's catastrophe model, HAZUS. A list of the common assumptions were specified for each modeling firm to conduct an assessment of the Charleston region, along with the key elements of uncertainty for the Wharton team to consider in an analysis they would undertake using the HAZUS model.

#### Composite Model Curves

The first goal of this case study was to discover not only the range of differences between results generated by the three competing catastrophe models, but also to compare a set of exceedance probability curves that represent the 5th percentile, mean, and 95th percentile level of loss. With these curves, a 90% confidence interval on loss is created. In other words, each model created three EP curves for comparison: a best estimate of loss, defined by its mean exceedance probability curve, and two additional curves representing a symmetric 90% confidence level about the mean loss.

As previously mentioned, the exceedance probability curves produced were expected to be dissimilar, given the degree of uncertainty associated with earthquake recurrence in the Charleston, South Carolina region. In fact, the degree of uncertainty amongst the models was expected to be quite great due to the lack of understanding of the seismic sources in this region. Charleston region is a low earthquake hazard area and the moment magnitude 7.3 earthquake event in 1886 is the only known historical event of note.

The assumptions for the analysis are summarized in Table 2. Four counties in the southeastern region of South Carolina, which surround the city of Charleston, comprised the study region. One hundred and thirty four census tracts are contained within the counties of Berkeley, Dorchester, Charleston, and Colleton. The HAZUS database of structures, as defined by the HAZUS97 release [14], was assumed for the inventory at risk. This database consists of seven occupancy classes of structures, namely residential, commercial, industrial, agricultural, religious, government, and educational occupancies. There were roughly 170,000 buildings in the data set, with approximately 97% of them classified as residential structures.

Using this common inventory database, each catastrophe model was run unaltered. In other words, no additional common information was used to define the hazard component, the vulnerability component, and the loss component of each model; the proprietary portion of each model remained as such for the study. The generated exceedance probability curves with the relevant confidence intervals were constructed by each of the modeling firms for the loss associated with building damage only (i.e., ground up loss); no insurance parameters were considered in the analysis.

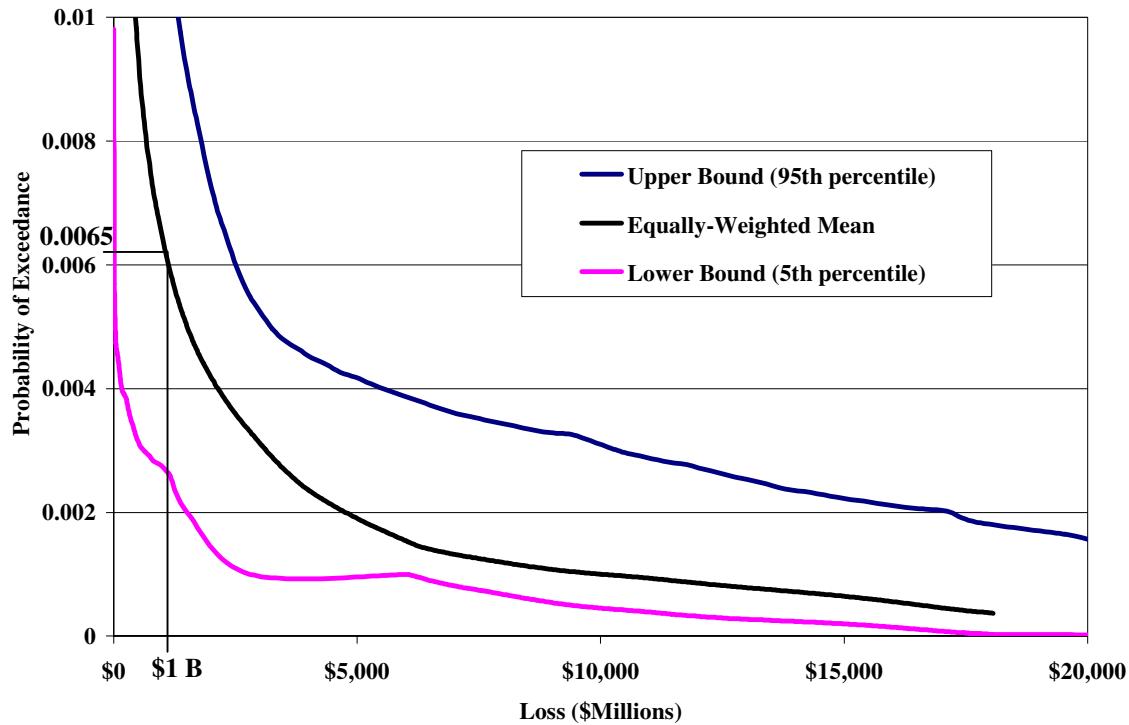
**Table 2: Charleston, South Carolina earthquake hazard analysis assumptions**

| Component     | Assumptions  |
|---------------|--|
| Hazard        | <ul style="list-style-type: none"> <li>Fault and area sources defined by model</li> <li>Recurrence defined by model</li> <li>Site specific characteristics defined by model</li> </ul> |
| Inventory     | <ul style="list-style-type: none"> <li>134 census tracts containing 170,000 structures</li> <li>97% residential structures</li> </ul>  |
| Vulnerability | <ul style="list-style-type: none"> <li>Damage functions/fragility curves defined by model</li> </ul>   |
| Loss          | <ul style="list-style-type: none"> <li>Repair costs defined by model</li> <li>Building damage loss only</li> </ul>   |

Given the proprietary nature of the competing models, each model's set of curves is not presented here. Instead, composite curves developed by the Wharton research team are shown. (The individual and composite curves were reviewed by Professor Robert Whitman of the Massachusetts Institute of Technology, as part of the Technical Advisory Committee input to the project.) In Figure 6, a composite EP curve for the mean loss is shown that represents an equally weighted linear combination of the data (1/3 of each). For example, suppose an estimate of the probability of exceeding a loss of \$1 billion ( $EP(L) = P(Loss > \$1 \text{ billion})$ ) is needed for the study area. Model A's probability of exceedance of 0.0091 is combined with Model B's exceedance probability of 0.0051 and Model C's exceedance probability of 0.0053 to estimate:  $P(Loss > \$1 \text{ billion}) = (0.0091 + 0.0053 + 0.0051)/3 = 0.0065$  or 0.65% probability of exceedance (a 1-in-154 year return period), as seen in Figure 6.

Bounding the composite mean EP curve are composite symmetric 90% confidence interval curves: a lower bound on loss, representing the 5th percentile loss, and an upper bound on loss, representing the 95th percentile loss. Since the range of exceedance probabilities varied greatly for a particular loss level for

these bounded curves, an equally weighted linear combination was not used (as it was in the mean case). Instead, the extreme value points across the three models were utilized in constructing the confidence intervals. Thus, the tendency to favor one model over the other two models was avoided.



**Figure 6: Composite exceedance probability curves for Charleston region.**

To illustrate the difference between the two approaches, consider the following example. Suppose that the 5th percentile of exceeding a loss of \$15 billion ( $EP(L) = P(\text{Loss} > \$15 \text{ billion})$ ) is required to determine a risk management strategy, which represents a large loss on the right-hand tail of the EP curve. Model A estimates a return period of 5,080 years, Model B estimates a return period of 1,730 years, but Model C's curve does not extend beyond 1,000 years because there is too much modeling uncertainty beyond this point. If the weighted linear combination of these two estimates were calculated equally, ignoring Model C, the result would be a return period of 3,405 years or 0.029% probability of exceedance.

Using the extreme value points for the lower and upper bound curves, the 5th percentile loss of \$15 billion has a return period of 5,080 years or approximately 0.02% probability of exceedance rather than the average of 0.029%. In this way, the 90% confidence level on the mean curve is an envelope of the three model curves, capturing the true bounds on the uncertainty across the three models.

Reconsidering the loss levels presented earlier for these curves, the probability that the loss to the inventory of structures in the Charleston region will exceed \$1 billion or  $EP(L) = P(\text{Loss} > \$1 \text{ billion})$  is, on average, 0.0065 or 0.65% with lower and upper bounds of 0.27% and 1.17%, respectively. The mean probability that the loss to the inventory of structures will exceed \$15 billion =  $P(\text{Loss} > \$15 \text{ billion}) = 0.00064$  or 0.064% with a lower bound of 0.02% and an upper bound of 0.22%.

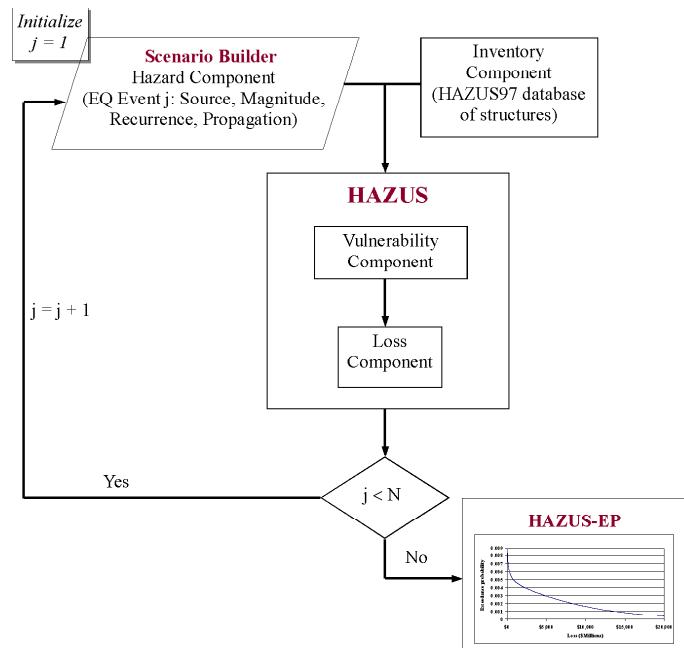
A specific loss level for the region could be determined, given a probability of exceedance, using the same data. Using the example of the range of losses for the 0.2% probability of exceedance or the 1-in-500 year

event, it can be determined from Figure 6 that the mean loss to these structures is \$4.6 billion with a lower bound of \$1.5 billion and an upper bound of \$17.1 billion. It should be clear that in dealing with catastrophe modeling, there is a wide variation in the probability of exceedance given a level of monetary loss and a wide variation in loss given a probability of exceedance.

### HAZUS Analysis

A related objective of the Charleston analysis was to generate an exceedance probability curve utilizing the HAZUS model and to test the sensitivity of the loss output to a few key assumptions in the model. For more details and the complete analysis, see Grossi and Windeler [15]. While the HAZUS methodology is more transparent than the approaches used in the three competing models (Model A, Model B, and Model C), it requires the development of additional software to create an EP curve [16]. The 1997 HAZUS earthquake model in its basic form was not designed to create an exceedance probability curve [14]. It could create either an estimate of loss based on one scenario event or based on a probabilistic seismic hazard map, such as the ones created by the USGS team of researchers [10].

The software tools that enable the creation of an exceedance probability curve using the HAZUS model consist of a pre-processor, designated Scenario Builder and a post-processor, designated HAZUS-EP. As shown in Figure 7, Scenario Builder defines a finite set of earthquake events,  $j = 1, 2, \dots, N$ , which represent a minimum set of data points needed to create an EP curve. Each event  $j$  is defined by its source, magnitude, rupture location, recurrence and attenuation (the hazard component of a catastrophe model). The data and assumptions used to develop the stochastic event set generally follow those described in the USGS National Seismic Hazard Map project [10]. Notably, the attenuation relationship to describe the rate at which ground motion decays from source to site is an equally weighted linear combination of the Frankel et al [10] and the Toro et al [11] empirical equations. In this way, all information available is incorporated into the model.



**Figure 7: Scenario Builder-HAZUS-HAZUS-EP to create an exceedance probability curve.**

The total set of events,  $N = 156$ , was chosen so that there was a wide enough spectrum of events capable of affecting the Charleston study region. The operative assumption in defining events was that variation in

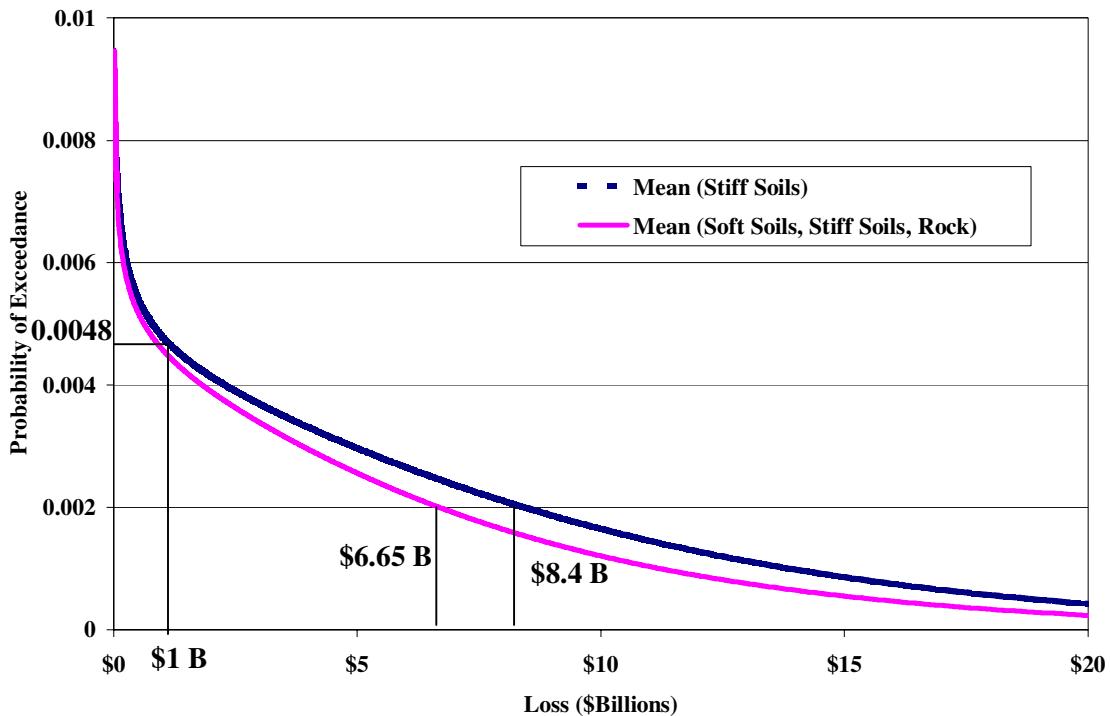
losses would decrease with distance from the study area. Therefore, the greatest number of events would be required within the study counties; the seismicity of progressively larger areas outside these counties could be represented by single events. Similarly, smaller magnitude events were eliminated with increasing distance. As in the earlier analysis to create the composite set of EP curves from Models A, B, and C, the database of inventory structures are defined by the HAZUS97 release [14], consisting of approximately 170,000 buildings of various occupancy classes.

With the portfolio of structures in Charleston, South Carolina, the HAZUS model is run for each event  $j$  with  $j = 1, 2, \dots, 156$ . The model calculates the damage to the structural and nonstructural building components and the resulting direct economic losses, as defined by the HAZUS methodology (the vulnerability and loss components of a catastrophe model). The results of each run, including losses by census tract and by occupancy type, are stored in a database file for input into the post-processor, HAZUS-EP. HAZUS-EP consolidates the losses to form an exceedance probability curve for the region.

In the complete analysis of the Charleston region using HAZUS, a collection of exceedance probability curves was generated under various assumptions in the hazard and inventory components of the model [15]. In this sensitivity analysis, such things as the occupancy mapping of structures, the attenuation relationships, the earthquake duration, and the soils mapping schemes were analyzed. Since a sensitivity analysis of every assumption in a catastrophe model cannot be presented here due to the large number of parameters, a single example demonstrating the sensitivity of loss to a site's underlying soil conditions is discussed here. The underlying soils across the entire region are classified as stiff soils or soil class D, as defined by the NEHRP provisions [17] and assumed in the default mode of HAZUS. To test the sensitivity of this assumption, a different GIS map was used which showed underlying soils in the region to be rock, stiff soils, and soft soils (soil classes B through E in the NEHRP provisions). This latter scheme, which considers varying soil classes, can be considered as a reduction in epistemic uncertainty due to the addition of new data on the geology of the region.

The two curves presented in Figure 8 are the mean exceedance probability curves assuming stiff soils and assuming a range of soil types (rock, stiff soils, and soft soils). Interestingly, for a given probability of exceedance, the loss assuming all stiff soils in the region is greater than the loss assuming a range of soil types. It is therefore a conservative assumption in the default mode of HAZUS. For example, at the 0.2% probability of exceedance or the 1-in-500 event, the stiff soils mean loss is \$8.4 billion and the mean loss assuming other soil types is \$6.65 billion. Therefore, the assumption of stiff soils everywhere in the region serves to establish a conservative estimate of loss. These curves show little expected loss above the 1% probability of exceedance level.

Finally, the probability of exceeding a loss of \$1 billion using the HAZUS model can be compared with the probability of exceeding this same loss calculated from the equally weighted linear combination of the three competing catastrophe models. The HAZUS analysis, assuming stiff soils everywhere in the region, estimates  $P(\text{Loss} > \$1 \text{ billion}) = 0.0048$  or 0.48% or 1-in-208 year event. As noted earlier and shown on Figure 6, the composite mean EP curve has  $P(\text{Loss} > \$1 \text{ billion}) = 0.0065$  or 0.65% probability of exceedance or a 1-in-154 year return period. These two return periods are not very different, a surprising result given the uncertainty in the seismicity of the Charleston region.



**Figure 8: HAZUS mean exceedance probability curves for Charleston region.**

## SUMMARY AND CONCLUSIONS

This paper examined the complexities of catastrophe modeling, mathematical constructs that allow the generation of exceedance probability curves, and the uncertainties inherent in the modeling process. By introducing the concepts of epistemic and aleatory uncertainty, the paper explored how to quantify uncertainty through the use of logic trees and simulation techniques. A case study in South Carolina indicated the importance of understanding where uncertainty lies in a catastrophe model and how it can be captured and utilized in the risk assessment process. By constructing exceedance probability curves with confidence intervals, the degree of uncertainty associated with natural hazard events, such as an earthquake in Charleston, can be appreciated.

## REFERENCES

1. Budnitz, RJ, Apostolakis, G, Boore, DM, Cluff, LS, Coppersmith, KJ, Cornell, CA, Morris, PA. "Recommendations for probabilistic seismic hazard analysis: guidance on uncertainty and use of experts." NUREG/CR-6372, Washington, DC: U.S. Nuclear Regulatory Commission, 1997.
2. Hanks, TC, Cornell, CA. "Probabilistic seismic hazard analysis: a beginner's guide." Proceedings of the Fifth Symposium on Current Issues Related to Nuclear Power Plant Structures, Equipment and Piping, I/1-1 to I/1-17, North Carolina State University, Raleigh, North Carolina, 1994.
3. Richter, CF. "Elementary seismology". San Francisco, California: W.H. Freeman and Company, 1958.
4. Youngs, RR, Coppersmith, KJ. "Implications of fault slip rates and earthquake recurrence models to probabilistic seismic hazard estimates." Bulletin of the Seismological Society of America 1985; 75(4): 939-964.

5. Peterson, MD, Bryant, WA, Cramer, CH, Cao, T, Reichle, MS, Frankel, AD, Lienkaemper, JL, McCrory, PA, Schwartz, DP. "Probabilistic seismic hazard assessment for the state of California". USGS Open-File Report 96-706, Menlo Park, California: United States Geological Survey, 1996.
6. Earthquake Engineering Research Institute. "World housing encyclopedia". <http://www.world-housing.net/>.
7. ATC-13. "Earthquake damage evaluation data for California." Redwood City, California: Applied Technology Council, 1985.
8. Dalkey, NC. "The delphi method." Santa Monica, California: Rand Corporation, 1969.
9. Browner, C. "Guidance for risk characterization." Environmental Protection Agency, February 1995.
10. Frankel, A, Mueller, C, Barnhard, T, Perkins, D, Leyendecker, EV, Dickman, N, Hanson, S, Hopper, M. "National seismic hazards maps: documentation." USGS Open-File Report 96-532, United States Geological Survey, June 1996.
11. Toro, GR, Abrahamson, N, and Schneider, J. "Model of strong ground motions from earthquakes in the Central and Eastern North America: best estimates and uncertainties." Seismological Research Letters 1997; 68: 41-57.
12. Iman, RL, Conover, WJ. "Small sample sensitivity analysis techniques for computer models, with an application to risk assessment." Communications in Statistics, Part A. Theory and Methods 1980; 17:1749-1842.
13. Grossi, P, Kleindorfer, P, Kunreuther, H. "The impact of uncertainty in managing seismic risk: the case of earthquake frequency and structural vulnerability." Risk Management and Decision Processes Working Paper 99-03-26, Philadelphia, Pennsylvania: The Wharton School, 1999.
14. NIBS. "HAZUS: Hazards U.S.: earthquake loss estimation methodology." NIBS Document Number 5200, Washington, DC: National Institute of Building Sciences, 1997.
15. Grossi, P, Windeler, D. "Sensitivity analysis of earthquake risk in the Charleston, South Carolina region," EERI's Sixth International Conference on Seismic Zonation 2000, November 12-15.
16. Grossi, P. "Quantifying the uncertainty in seismic risk and loss estimation." Doctoral Dissertation, Philadelphia: University of Pennsylvania, 2000.
17. Federal Emergency Management Agency. "FEMA 303: NEHRP recommended provisions for seismic regulations for new buildings and other structures." Building Seismic Safety Council, 1997.