Proceedings of Ninth World Conference on Earthquake Engineering
August 2-9, 1988, Tokyo-Kyoto, JAPAN (Vol.V)

8-2-3

DEVELOPMENT OF AN EXPERT SYSTEM FOR SEISMIC HAZARD EVALUATION OF SITES IN CALIFORNIA

Michèle LAMARRE[1] and Haresh C. SHAH[2]

[1]Ammann & Whitney Consulting Engineers, New York, NY 10014
[2]Department of Civil Engineering, Stanford University, Stanford, CA

INTRODUCTION

In the last 25 years, seismic hazard analysis has received much attention leading to the development of a large number of models to describe earthquake recurrence, magnitude distribution, and ground motion attenuation. To assist engineers in selecting the appropriate models, gathering and manipulating the data for parameter estimation, and interpreting the results, a self-documenting tool has been developed. The tool, built using the IBM based expert system shell, INSIGHT 2+, to control the problem solving strategy and the C programming language for computational algorithms, provides a general framework that can accommodate any recurrence or attenuation model for evaluating the seismic hazard. The uncertainty associated with the models and data used for seismic hazard analyses is carried through to the final results using the bootstrap statistical method. The system is designed to use incomplete seismic records optimally and is general enough to be implemented for various seismic regions by providing the appropriate data. The features of the system and the different options offered for seismic hazard analysis will first be discussed. The handling of recurrence models and the data manipulations necessary to perform the bootstrap and obtain uncertainty on the results are described in the second part. The uncertainty measure takes into account earthquake catalogue incompleteness, magnitude measurement and conversion errors, and modeling uncertainty.

USER INTERFACE

The system allows the assessment of seismic hazard to be performed in two different fashions. In the first assessment, a characterization of the random sequence of future ground motion intensities at a site due to earthquakes on the surrounding seismic sources is established. The seismic hazard in this case can be represented by the probability that a given level of ground shaking will be exceeded in a given time period, by the ground shaking level associated with a given risk, by hazard curves, or by response spectra.

The second assessment consists of determining a general hazard index which gives a measure of the total hazard at a site obtained by combining the ground shaking hazard with the potentials for liquefaction, fault break, and landslide. The factors involved in estimating the ground rupture potential are often described linguistically. In order to model and combine vague concepts, fuzzy set theory is being used (Ref. 1). To combine the ground motion with the ground rupture potential into the total hazard, the value of the ground motion part of the hazard calculated in the first assessment is mapped into a fuzzy grade.

Currently, the system assesses the seismic hazard for any site in California. However, the system is designed to be general and it can be applied to any seismic region characterized by relatively well defined faults by providing the data for the region of interest. This data consists of the seismic sources, the earthquake record, and the available geological information regarding the soil conditions and the potentials for landslide and liquefaction. This information is stored in the system's database. The seismic sources can be composed of several faults with homogeneous seismicity and the faults can have multiple segments. For instance, the seismic source data for California has been obtained by digitizing a California fault map (Ref. 2).

In order to use old and large earthquakes to refine the modeling of the occurrence of these earthquakes in the future, the entire historical record is included in the database. The historical record is however incomplete, and the use of the entire earthquake catalogue to estimate the recurrence model parameters introduces a strong bias. To remove the bias, the degree of incompleteness of the earthquake data must be assessed. The degree of incompleteness can be described with the detection probabilities which depend on the population and its spread, on the extent of the seismograph network, and on the size of the earthquakes. Thus, detection probabilities vary with time, space, and magnitude. Completeness regions are used to map the degree of incompleteness of the earthquake data. A completeness region is defined as a geographical region, a time period, and a magnitude range with homogeneous earthquake recording or in other words with homogeneous probability of detection. These completeness regions are determined by studying the growth of the population and of the seismograph network over the years, and by comparing the variation in the earthquake counts for different magnitude ranges before and after the advent of full seismograph coverage. The system can assist the user in obtaining the detection probabilities since it can provide earthquake counts for any completeness region.

Assuming that the seismicity remained constant over the time period covered by the historical record, the detection probability in a completeness region is obtained by comparing the number of earthquakes in this region with the number of earthquakes in a completeness region with complete reporting for the same geographical region and the same magnitude range. The probability of unobserved events in any completeness region and for any time period can then be determined from the detection probabilities. The use of these probabilities to evaluate the uncertainty in the results of the recurrence models due to earthquake catalogue incompleteness is discussed in the next section.

Typically, an earthquake catalogue displays a certain number of measures in different magnitude scales for each earthquake. To form a uniform basis for comparison and calculations, the different scales are converted into a unique magnitude scale which is chosen to be the moment magnitude since it gives the best physical representation of the size of an earthquake and it does not saturate for large values. All the different magnitude measurements available for an earthquake are used to determine the moment magnitude and a range of uncertainty is assigned to the resulting magnitude taking into account the measurement and conversion errors.

Once the seismic source and the earthquake data are read into the system, each earthquake is automatically associated with a source in the database by finding the closest source to its epicenter. The hypocenter information is used when available.

Geological data such as the soil conditions and the potentials for landslide and liquefaction is used to obtain response spectra or general hazard indices at various sites. Obviously, the finer the grid for which the geological information is available, the better the hazard index and the effects of the local soil conditions can be assessed. Furthermore, the modelling of ground motion attenuation improves with greater accuracy in the representation of soil types along the travel path of the seismic waves. For the time being, a precise representation of geological conditions is not available and the use of fuzzy sets to describe the uncertainty in the general hazard index reflects the degree of uncertainty encountered in the database.

The recurrence and attenuation models must be set for all sources for the seismic hazard analysis part of the system to be operational. The recurrence models are divided into two parts: an inter-event time distribution and a magnitude distribution. An inter-event distribution can be combined with any magnitude distribution. For instance, the Poisson model can be coupled with a truncated exponential or with a Weibull magnitude distribution.

Two types of empirical attenuation relationships are available in the system: point rupture models and fault rupture models. The seismic sources can be attributed any recurrence or attenuation model available in the system. The system is designed, however, to be able to handle any recurrence or attenuation model. A new attenuation model can be added by providing its empirical relationship and a new recurrence model by providing its probability distribution functions. The recurrence models can be set manually or automatically. In the case the user wants to set the models manually, statistical tests and recommendations about the results of these tests are available to help the user. The system can also determine automatically which model should be attributed to a source. Rules of thumb extracted from experts or their publications and calculation of statistics such as coefficients of variation or correlation coefficients are used to select models for sources. The rules and statistical tests used in the algorithm for automatic selection are discussed in a report by the authors (Ref. 3).

RECURRENCE MODELS

Once the recurrence and attenuation models are set for each source, the seismic hazard can be evaluated. The goal of the system is to provide a measure of uncertainty along with the hazard estimates. The uncertainty that arises when dealing with incomplete seismic records, magnitude measurement and conversion errors, and recurrence models (modeling uncertainty) is used to determine the uncertainty attached to the final hazard estimates with the bootstrap statistical method (Ref. 4).

To determine the probability of exceeding an acceleration level, a_c, at a site due to a particular source, the probability of exceeding the magnitude, m_c, causing this acceleration level must first be determined. For each source having the potential of creating a_c at the site of interest, the following procedure, referred to as one bootstrap, is performed. An event history of the source is obtained from the historical record contained in the database. The duplicate records and clusters are then removed from the event history. The initial sample for the bootstrap is obtained by modifying the event history to account for the possibility for unobserved events along the source. Whenever a source covers one or more completeness regions either in time, space, or magnitude for which the detection probabilities are different from 1.0, there is a possibility for unreported events.

For each earthquake in the event history, all the possible scenarios for its previous earthquake are determined. A scenario consists of the magnitude range of an earthquake in the original event history, the magnitude range for the possible preceding earthquake, and the inter-arrival time separating these two earthquakes. As a reminder, each earthquake magnitude is stored as a range in the database. Each scenario is assigned a probability which depends on the uncertainty in magnitudes of the two earthquakes and on the probability of either unobserved or no unobserved events in the time period covered by the scenario. This probability represents the chance with which a particular scenario of an earthquake will be in the outcome of the sampling.

A possible event history is drawn from the initial sample which contains all the possible scenarios for the event history. The new sample consists of a list of two magnitude ranges and an inter-arrival time. The removal of the absolute time information allows the sampling to be performed randomly without introducing temporal inconsistencies. From these new samples, the model parameters of the model selected for the source are estimated with the maximum likelihood criterion. The probability of exceeding a magnitude level on the source is computed with the values of these parameters. A general function is used in the system to accommodate any recurrence model. This function is computed iteratively using a numerical approximation. All that needs to be provided are the complementary cumulative distribution functions fro the magnitude and the inter-arrival time distributions. The formulation of this function is shown in a report by the authors (Ref. 3).

The probability of exceeding a peak ground acceleration level a_c at a site is then computed from combining the attenuation relationships and the probabilities of exceeding the magnitude levels creating this acceleration for various source-to-site distances. This procedure consists of one bootstrap and is applied to each source having the potential of affecting the site in order to determine the total probability of exceeding the acceleration level a_c.

The bootstrap technique consists in this case of repeating this procedure a large number, N, of times from the sampling of a possible event history on each source to the combination of the effects from all sources to obtain the probability of exceeding a_c at the site. The N probabilities obtained are sorted from the smallest to the largest and the desired percentile confidence interval can be drawn to give uncertainty bounds on the results.

CONCLUSION

Expert system techniques provide the transparency and flexibility necessary to assist engineers in evaluating the seismic hazard. The system that has been developed is general and can be applied to other seismic regions by providing the data for the region. The system assists the user in gathering the appropriate data. The system also embodies the knowledge to decide automatically or recommend which recurrence model should be used for a particular source. The system provides a general framework that can accommodate any recurrence or attenuation model to assess the seismic hazard. The availability of several models in the system makes the comparison between the results and the uncertainty obtained with these models very easy and informative.

To deal with the incompleteness of the earthquake data, the method developed in the system makes use of earthquake detection probabilities varying with time, magnitude, and geographical region. The incompleteness information is used to build scenarios of the earthquake history that include the possibility for unobserved events. This information also permits the use of all the data in the historical record for unbiased estimation of recurrence model parameters.

A measure of the uncertainty due to earthquake catalogue incompleteness, magnitude measurement and conversion errors, and recurrence models is provided along with seismic hazard estimates. This uncertainty helps the user interpret the results since this measure gives an idea of the estimate's accuracy. The system developed in this work can systematically propagate the uncertainty for any recurrence model used. This general method for quantifying uncertainty represents a major improvement over the methods developed almost exclusively for the Poisson model in previous studies.

The results of the seismic hazard analyses supported by the system were compared to the results obtained in other studies. The latter results were almost always within the confidence bounds of the former, which indicates that the method developed in this work is appropriate (Ref. 3).

REFERENCES

1. L.A. Zadeh. Fuzzy sets. *Information and Control,* 8:338-353, 1965.
2. L. Mualchin and A.L. Jones. Peak acceleration from maximum credible earthquakes in California. Technical report, California Division of Mines and Geology, Jan. 1987.
3. M. Lamarre. Seismic hazard evaluation for sites in California: Development of an expert system. Technical Report 85, The John A. Blume Earthquake Engineering Center, Department of Civil Engineering, Stanford University, June 1988.
4. B. Efron and R. Tibshirani. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1(1):54-77, 1986.