CAPRA - Comprehensive Approach to Probabilistic Risk Assessment: International Initiative for Risk Management Effectiveness

O.D. Cardona
Instituto de Estudios Ambientales, Universidad Nacional de Colombia, Manizales, Colombia

M.G. Ordaz & E. Reinoso
Instituto de Ingeniería, Universidad Nacional Autónoma de México (UNAM), México, D.F.

L.E. Yamín
Universidad de los Andes, Bogotá, Colombia

A.H. Barbat
Centro Internacional de Métodos Numéricos en Ingeniería, CIMNE, UPC., Barcelona, Spain

SUMMARY:
Understanding disaster risk due to hazard events, such as earthquakes, creates powerful incentives for countries to develop planning options and tools to reduce potential damages. This has been the reason why CAPRA, the risk evaluation model described in this paper, was developed with the technical and financial support of the World Bank, the Inter-American Development Bank (IDB) and the International Strategy of United Nations for Disaster Reduction (ISDR). CAPRA is a techno-scientific methodology and information platform, composed of tools for the evaluation and communication of risk at various territorial levels. This model allows the evaluation of probabilistic losses on exposed elements using probabilistic metrics, such as the exceedance probability curve, expected annual loss and probable maximum loss, useful for multi-hazard/risk analyses. The platform is conceptually oriented to facilitate decision making. Using CAPRA it is possible to design risk transfer instruments, the evaluation of probabilistic cost-benefit ratio, providing an innovative tool for decision makers to analyze the net benefits of the risk mitigation strategies, such as building retrofitting. This model is useful for land use planning, loss scenarios for emergency response, early warning, on-line loss assessment mechanisms, and for the holistic evaluation of disaster risk based on indicators that facilitates the integrated risk management by the different stakeholders involved in risk reduction decision-making. CAPRA has been used in Central and South America and in some countries of Europe and Asia. It has been the base for the evaluation of the country’s risk profile for Colombia, Mexico and Nepal in the framework of the United Nations Global Assessment Report GAR-2011 and it is a potential contribution for the Global Earthquake Model (GEM). Examples of application of the model in the different countries, descriptions of the wiki and visualization tools available are made to illustrate the capabilities of this innovative open architecture ad open source platform using as example earthquakes, nevertheless similar application can be made for hurricanes, floods, landslides and volcanoes.

Keywords: Seismic risk, building damage, insurance, risk reduction, loss scenarios.

1. PROBABILISTIC RISK MODEL

The frequency of catastrophic seismic events is particularly low; this is one of the reasons why very limited historical data are available. Considering the possibility of future highly destructive events, risk estimation has to focus on probabilistic models which can use the limited available information to best predict future scenarios and consider the high uncertainty involved in the analysis. Therefore, risk assessments need to be prospective, anticipating scientifically credible events that might happen in the future. Seismological and engineering bases are used to develop earthquake prediction models which permit to assess the risk of loss as a result of a catastrophic event. Since large uncertainties are inherent in models with regard to event severity and frequency characteristics, in addition to consequent losses caused by such events, the earthquake risk model is based on probabilistic formulations that incorporate this uncertainty into the risk assessment. The probabilistic risk model built upon a sequence of modules, quantifies potential losses arising from earthquake events as shown in the Fig. 1.1 (Cardona et al 2006, 2009, ERN-AL 2010).
2. SEISMIC HAZARD MODULE

The hazard module defines the frequency and severity of a peril, at a specific location. This is completed by analyzing the historical event frequencies and reviewing scientific studies performed on the severity and frequencies in the region of interest. Once the hazard parameters are established, stochastic event sets are generated which define the frequency and severity of thousands of stochastic events. In the case of earthquakes this module can analyze the intensity at a location once an event in the stochastic set has occurred, by modeling the attenuation of the event from its location to the site under consideration, and evaluates the propensity of local site conditions to either amplify or reduce the impact. The seismic hazard is expressed in terms of the exceedance rates of given values of seismic intensity \( a \). Its calculation includes the contribution of the effects of all seismic sources located in a certain influence area. Once these seismic sources are identified, a certain occurrence model is assigned to the earthquakes that take place there. In the most cases all seismic sources are modeled to follow a Poisson process in which \( \lambda(M) \) represents the activity rates for each faulting system. Since the seismic sources are volumes and the methodology considers a point source approach, the epicenters cannot only occur in the centers of the sources, but can also occur, with equal probability, in any point inside the corresponding volume. Therefore, for the simulation of event sets, sub-sources are defined by subdividing the seismic sources, depending on hypocentral distance \( R_0 \), in diverse geometric shapes. For each subdivision the seismicity of the source is considered to be concentrated in its center of gravity. In addition the model considers the attenuation effects of the seismic waves by means of probabilistic spectral attenuation laws that include different source types and the local amplification effects based on microzonation studies and other available complementary information. Since the computed intensity is regarded as a random variable with lognormal distribution, its corresponding uncertainty value \( \sigma_{ln(a)} \) is considered to include the associated variability. Assuming that the intensity variable has a lognormal distribution given the magnitude \( M \) and distance \( R_0 \), the probability of a given seismic intensity \( a \), \( \Pr(A > a|M, R_0) \) is calculated as follows:

\[
\Pr(A > a|M, R_0) = \Phi \left( \frac{1}{\sigma_{ln(a)}} \ln \frac{MED(A|M, R_0)}{a} \right)
\]

(1.1)

where \( \Phi(\cdot) \) is the standard normal distribution, \( MED(A|M, R_0) \) is the median value of the intensity variable (given by the corresponding attenuation law) and \( \sigma_{ln(a)} \) the standard deviation of the natural logarithm of the intensity \( a \). This methodology based on (Esteva 1970, Ordaz 2000), generates stochastic seismic events at random locations within the modeled seismic sources, calculates the probability density function (PDF) of the seismic intensity \( a \) for a specific location, and, if required, adds up the contributions of all sources and magnitudes in order to compute intensity exceedance rates, as those depicted in Fig. 2.1. From these intensity exceedance rates, it is possible to determine uniform hazard spectra (UHS) for a specific site, based on the calculated intensity value (e.g. PGA, spectral acceleration, etc.) associated to a fixed return period. Therefore UHS can be determined by connecting the intensity points calculated from Fig. 2.1 for a given exceedance rate (inverse of the return period).
If the procedure described is followed for different locations within the city, and the selected intensity variable is calculated for a return period with site effects, it is possible to build a map for different seismic intensities at ground level including the site effects of the seismic microzonation of the city (Fig. 2.2.).

3. EXPOSURE MODULE

The exposure values of “assets at risk” are estimated either from available secondary data sources such as existing databases or they are derived from simplified procedures based on general macroeconomic and social information such as population density, construction statistics or more specific information. This “proxy” approach is used when the preferred specific site by site data are not available.

Based on the information available, an input database is constructed based on GIS and specific required information is completed, using for example internet data gathering tools. The exposure can be developed also using remote sensing images and the digitalization of polygons, lines, points using drawing tools as the Fig. 3.1 illustrates. In addition, the exposure database can be developed using cadastral information when it is available. Special routines allow for the visualization of the database information. Fig. 3.2 presents example maps of a Colombian city’s database used for analyzing all building constructions in the urban area, building a model of up several hundred thousand items.
In order to calculate the social impact, general information related to building occupation is also estimated. Maximum occupancy and occupancy percentage at different hours are also defined in order to allow different time scenarios of the event’s occurrence. When no specific occupation information is available, approximate density occupation by construction class can be used in order to complete such information.

4. VULNERABILITY MODULE

Defining loss (L) as a random variable, vulnerability functions describe the loss statistical moments variation to different values of seismic demand. Loss probability distribution is usually assumed Beta, where statistical moments correspond to mean (usually referred to as Mean Damage Ratio, MDR) and standard deviation. Beta distribution \( p_L|S(L) \) is defined as follows:

\[
Pr_{L|S}(L) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} L^{a-1} (a - L)^{b-1}
\]  

(4.1)

where \( \Gamma \) is the Gamma function and parameters \( a \) and \( b \) are:

\[
a = \frac{1-c^2(L|S)E(L|S)}{c^2(L|S)}
\]  

(4.2)

\[
b = \frac{1-E(L|S)}{E(L|S)}
\]  

(4.3)
where $E(L|S)$ is the loss mean or expected value and $c(L|S)$ is the loss variation coefficient, given a seismic demand $S$ (note that $c(L|S) = SD(L|S)/E(L|S)$ where $SD(L|S)$ is the loss standard deviation given a seismic demand $S$).

Vulnerability functions provide all the necessary information to calculate the probability of reaching or exceeding a loss value, given as seismic demand. Loss is defined using numerical scales instead of qualitative scales as for damage states (for example the ratio of repair cost to the asset replacement value), which allows its direct use in probabilistic risk and loss calculations. The probability of reaching or exceeding a loss value is calculated as follows:

$$Pr(L \geq l|S) = \int_{l}^{\infty} p_{L|S}(L) dL$$

where $l$ is a loss value in the random variable $L$ dominium, and $S$ is the seismic demand.

This module quantifies the damage caused to each asset class by the intensity of a given event at a site (Miranda 1999). The development of asset classification is based on a combination of construction material, construction type (say, wall & roof combination), building usage, number of stories and age. Estimation of damage is measured in terms of the Mean Damage Ratio (MDR). The MDR is defined as the ratio of the expected repair cost to the replacement cost of the structure. A vulnerability curve is defined relating the MDR to the earthquake intensity which can be expressed in terms of maximum acceleration (e.g. useful for 1-2 story buildings), spectral acceleration, velocity, drift or displacement (e.g. useful for multi-story buildings) at each location. Given a value of seismic intensity for a certain building type, MDR can be calculated using Eqn. 4.1 (Miranda 1999, Ordaz 2000).

$$E(\beta | \gamma_i) = 1 - \exp \left( \ln 0.5 \left( \frac{\gamma_i}{\gamma_0} \right)^{\frac{E(\cdot)}{E(\cdot)}} \right)$$

where $\beta$ is the loss, $\gamma_0$ and $\gamma_i$ are structural vulnerability parameters that depend of the building typology and construction date, and $E(.)$ is the expected value. By definition, $\beta$ is the ration between the repairing cost and the total cost of the building; a value from 0 to 1. Using the spectral acceleration is possible to determine the maximum nonlinear drift as follows (Miranda, 1997):

$$\gamma_i = \frac{\beta_1 \beta_2 \beta_3 \beta_4 (\eta N)^2}{4 \pi^2 \rho n h} S_a(T) \quad T = \eta N \rho$$

where $\beta_1$ is the ration between the maximum lateral displacement in the upper level of the structure and the spectral displacement; $\beta_2$ is the ration between the maximum interstory distortion and the global distortion of the structure; $\beta_3$ is the ratio between the maximum inelastic lateral displacement and the maximum displacement of the elastic model $\beta_4$ is the ratio between the elastic and inelastic $\beta_2$ factors; $\rho$ and $n$ are factors to estimate the fundamental period of the structure from the number of stories, $N$; $h$ is the height of each structural story, that depends of the structural typology, the geographic location and the construction date; $S_a(T)$ is the spectral acceleration, that depends of the vibration fundamental period of the structure, the structural damping and the seismic hazard in the place. Several construction classes are included in the system for different type of intensities. Fig. 4.1 shows a vulnerability curve.

Structural damage on buildings and infrastructure due to earthquakes is often represented using fragility functions, which relate the probability of reaching or exceeding a given damage state, versus the seismic demand in terms of peak or spectral intensities. The development of fragility functions or curves for a given structural system requires the definition of the critical ground motion parameter and the identification of the expected damage states.
Suppose there are $N$ identified damage states ($ds$) for a structural system. Then the probability of reaching or exceeding the $i$th damage state ($P_i$), given ground motion intensity ($S$) is:

$$P_i = \Pr(DS \geq ds_i | S)$$

(4.7)

where $DS$ is a damage random variable on the damage state vector $\{ds_0, ds_1, \ldots, ds_N\}$. In other words, fragility curves define the probability that the expected global damage $d$ of a structure equals or exceeds a given damage state, $ds_i$, as a function of a parameter quantifying the severity of the seismic action. Let be this parameter the spectral displacement $S_d$. Thus, fragility curves are completely defined by plotting $\Pr(d \geq ds_i)$ in ordinate and the spectral displacement $S_d$ in abscissa. If it is assumed that fragility curves follow a lognormal probability distribution, they can be completely defined by only two parameters which, in this case, are the mean spectral displacement $S_{d,ds}$ and the corresponding standard deviation $\beta_{ds}$. Typical damage states and fragility curves are shown in Fig. 4.2.

Vulnerability functions describe loss in a proper manner for risk calculation. However, fragility curves and damage probability matrices (DPMs) have been extensively used worldwide due to their comprehensive description of structural behavior to seismic demands. Relationships between these estimations can be addressed. Damage states in fragility curves and DPMs are often defined to characterize a physical state of the structure. Physical states are qualitative and merely descriptive of the expected damage. For rigorous loss calculations, a numerical damage cost scale, for example in terms of the ratio of repair cost to replacement value must be related to the defined damage states.

Suppose there are $N$ identified damage states ($ds$) for a structural system. For each damage state, a loss value ($L$) is assigned. In other words, when the structure reaches a damage state $ds_i$, its owner will have to pay a repair cost $L_i$. Therefore, the loss statistical moments can be calculated as follows:
\[
E(L|S) = \sum_{i=0}^{N} L_i \Pr(\text{DS} = d_{si}|S) \quad (4.8)
\]
\[
SD^2(L|S) = \sum_{i=0}^{N} (L_i - E(L|S))^2 \Pr(\text{DS} = d_{si}|S) \quad (4.9)
\]

where \( E(L|S) \) is the loss mean and \( SD^2(L|S) \) is the loss standard deviation given a seismic demand \( S \), and \( \Pr(\text{DS}=d_{si}|S) \) is the discrete probability of reaching a damage state \( d_{si} \).

The system also allows for the use of customized vulnerability models in different formats (Fig. 4.3). Specific vulnerability curves can be defined for building contents and for business interruption costs.

Figure 4.3. Different format used by CAPRA include capacity curves (pushover), fragility curves and vulnerability functions.

5. DAMAGE AND LOSS MODULE

As it is well known, risk is normally measured using the exceedance rate of loss values, \( \nu(p) \). This quantity is the expected number of earthquakes, per unit time, that will produce losses equal or larger than \( p \). It is computed using the total probability theorem:

\[
\nu(p) = \sum_{i=1}^{\text{Events}} \Pr(P > p|\text{Event } i) \cdot F_A(\text{Event } i) \quad (5.1)
\]

where \( \Pr(P>p|\text{Event } i) \) is the probability of exceedance of loss \( p \) given that event \( i \) took place, and \( F_A(\text{Event } i) \) is the annual occurrence frequency of event \( i \). Vulnerability functions are used to compute \( \Pr(P>p|\text{Event } i) \).

Normally, an event would be specified in terms of, at least, its magnitude and its hypocentral location. Hence, in order to compute \( \Pr(P>p|\text{Event } i) \) the following considerations are made.

It is assumed that, given the occurrence of event \( i \), with known magnitude and hypocentral location, the intensity at the site of the structure is a lognormal random variable with median and logarithmic standard deviation that, in general, depend on magnitude and source-site distance.

Under this assumption, the required probability \( \Pr(P>p|\text{Event } i) \) is computed chaining two conditional distributions:

\[
\Pr(P > p|\text{Event } i) = \int_0^{\infty} \Pr(P > p|Sa)p_{Sa}(Sa|M, R) \ dSa \quad (5.2)
\]
where \( p_{SA}(Sa|M,R) \) is the probability density function of intensity \( Sa \) given that a magnitude \( M \) earthquake took place at a source-site distance \( R \). As it was mentioned, \( Sa|M,R \) is often assumed to be lognormally distributed, with median and logarithmic standard deviation that depend on \( M \) and \( R \), which are computed using the ground-motion prediction model selected by the analyst. The first term of the integrand is, obviously, computed using the vulnerability relation that describes the behavior of the structure under analysis. The above equations give a clear indication of how uncertainties in vulnerability are propagated throughout the risk analysis.

In this module, then, to calculate losses, the damage ratio derived in the vulnerability module is translated into economic loss by multiplying the damage ratio by the value at risk. This is done for each asset class at each location. Losses are then aggregated as required (Ordaz et al. 1998, 2000). The loss module estimates the net losses. They can be useful for insurance information taking into account for example deductible, sum insured, etc. Risk metrics produced by the model provide risk managers and decision makers with essential information required to manage future risks. The main metrics for risk assessment are the following:

**Average Annual Loss.** AAL is the expected loss per year. Computationally, AAL is the sum of products of event expected losses and event annual occurrence probabilities for all stochastic events considered in the loss model. In probabilistic terms AAL is the mathematical expectation of the annual loss. The expected annual loss, also known as pure premium when it is express as a rate of the asset replacement value, is defined as the expected loss value that could occur in any year, supposing that the process of occurrence of hazard events is stationary and that damaged structures have their resistance immediately restored after an event. It can be calculated as follows (Ordaz et al. 1998, Ordaz 1999):

\[
AAL = \sum_{i=1}^{\text{Events}} E(P|Event \ i) \cdot F_A(\text{Event } i)
\]  

(5.3)

where \( AAL \) is the Average annual expected loss, \( E(P|Event \ i) \) is the expected loss value since event \( i \) occurred, and \( F_A(\text{Event } i) \) is annual occurrence frequency of event \( i \). Annual occurrence frequency of events depends on the results of hazard assessments. The loss expected value given the occurrence of a particular event depends on the vulnerability of the exposed element.

**Loss Exceedance Curve.** LEC represents the annual frequency with which a loss of any specified monetary amount will be exceeded. This is the most important catastrophe risk metric for risk managers, since it estimates the amount of funds required to meet risk management objectives. The LEC can be calculated for the largest event in one year or for all (cumulative) events in one year. For risk management purposes, the latter estimate is preferred, since it includes the possibility of one or more severe events resulting from earthquakes. Fig. 5.1 presents the PML curve for a portfolio of buildings in a city of Colombia (CEDERI, 2005, ERN-Colombia 2005).

**Probable Maximum Loss.** PML represents the loss amount for a given annual exceedance frequency, or its inverse, the return period. The PML curve on the other hand is generally specified as the PML in economic value or in percentage with regard to the return period. The PML of an exposed base is an appraiser of the size of maximum losses that could be reasonably expected in such set of elements exposed during the occurrence of a hazard event. It is typically used as fundamental data to determine the size of reserves insurance companies should maintain to avoid excessive losses that might surpass their adjustment capacity. It is defined in this model as the loss average that could occur for a given return period. Therefore, it is necessary to calculate excess rates of net losses from the portfolio. Such excess rates are not more than the number of times per year that is expected that a certain value of loss is even or exceeded. The excess rate of a given loss value \( p \) is calculated as:

\[
v(p) = \sum_{i=1}^{\text{Event}} \Pr(P > p|Event \ i) \cdot F_A(\text{Event } i)
\]  

(5.4)

where \( v(p) \) is the excess rate of \( p \) loss, \( \Pr(P > p|Event \ i) \) is the excess probability of \( p \) loss, since event \( i \) occurred, and \( F_A(\text{Event } i) \) is the annual occurrence frequency of event \( i \).
Depending on a stakeholder’s risk tolerance, the risk manager may decide to manage for losses up to a certain return period (e.g. 1 in 300 years). For that stakeholder (e.g. a public or private agency), the PML is the 300-year loss. For others, it may be 150 years, or for others 500 years. It is noteworthy that it is frequent to set program insolvency at the one in 150-year period to one in 200-year level, which roughly corresponds to the level of solvency required for BBB+ companies rated by S&P. However, other stakeholders (e.g. governments or regulation agencies) involved have chosen much longer return periods, such as the Mexican Insurance Commission, which uses a return period of 1500 years to fix solvency margins of insurance companies in Mexico.

Curves equivalent to those of loss excess can also be generated for other risk measuring parameters such as the probable maximum number of casualties or injured in function to the return period. On the other hand, in addition to the probabilistic economic figures it is also relevant for disaster management and vulnerability reduction to have the earthquake loss scenarios from a deterministic perspective, considering some historical earthquakes or future events. This is particularly useful for the formulation of a city’s emergency response plan and to identify the buildings and blocks with potential damage concentration. Fig. 5.2 shows the map of damage in Manizales City, in Colombia, using the damage and loss module.

Considering the possibility of future highly destructive events, risk estimation has to focus on models which can use the limited available information to best predict future scenarios and consider the high uncertainty involved in the analysis. As a conclusion, since large uncertainties are inherent in models with regard to event severity and frequency characteristics, in addition to consequent losses caused by such events, the risk model of CAPRA is based on formulations that incorporate this uncertainty into the hazard and risk assessment.
6. CONCLUSIONS

The CAPRA initiative provides different type of users with tools, capabilities, information and data to evaluate disaster risk. CAPRA applications include a set of different software modules for the different types of hazards considered, a standard format for exposure of different components of infrastructure, a vulnerability module with a library of vulnerability curves and an exposure, hazard and risk mapping geographic information system. Probabilistic techniques of CAPRA employ statistical analysis of historical datasets to simulate hazard intensities and frequencies across a country’s territory. This hazard information can then be combined with the data on exposure and vulnerability, and spatially analyzed to estimate the resulting potential damage. This measure can then be expressed in quantified in risk metrics such as a probable maximum loss for any given return period or as an average annual loss. Since this risk is quantified according to a rigorous methodology, users are enabled with a common language for measuring, and comparing or aggregating expected losses from various hazard, even in the case of future climate risks associated with climate change scenarios. The platform’s architecture has been developed to be modular, extensible and open, enabling the possibility of harnessing various inputs and contributions. This approach enables CAPRA to become a living instrument. CAPRA’s innovation extends beyond the risk modeling platform; a community of disaster risk users is now growing in the countries; training and workshops are now under development and a complete strategy for future development is under way.

ACKNOWLEDGEMENT

CAPRA is an ongoing initiative that has been developed in different phases with financial support of the World Bank, the Inter-American Development Bank and the UN-ISDR. The development of the platform and its application in different countries have been in charge of the Consortium Evaluación de Riesgos Naturales - America Latina, ERN-AL, composed by the International Center of Numerical Methods in Engineering, CIMNE, from Barcelona, INGENIAR LTDA. and ITEC S.A.S from Bogota, and ERN Ingenieros Consultores from Mexico. The authors acknowledge the technical contribution of the large number of engineers and professionals involved in the development of the software platform.

REFERENCES

Esteva, L. (1970). Regionalización sísmica para fines de ingeniería; serie azul, Instituto de Ingeniería, UNAM.