The 7th Balkan Conference on Operational Research
"BACOR 05"
Constanta, May 2005, Romania

SEISMICITY AS A MULTIDIMENSIONAL STOCHASTIC PROCESS

MILICA DELIC
ZORAN RADOJICIC

Faculty of Organizational Sciences, Belgrade, Serbia and Montenegro

Abstract
One of the main tasks in engineering seismology is the analysis of potential effects of future earthquakes on the objects and systems of objects built in the region with high seismic activity.
Occurrence of earthquakes is stochastic in many ways and the approaches that are developed and often used are based on the fact that seismicity can be observed as multidimensional stochastic point process.
The main components that are most often used in modeling are place, magnitude and time of earthquake occurrence. From the history of seismicity and geological analysis of the region with high seismicity, potential earthquake sources are identified and parameters needed for the analysis can be defined.
In this paper will be described and discussed some ways of modeling seismicity as a multidimensional stochastic process.

***Keywords**: Earthquake, seismicity, seismic hazard, seismic risk, confidence interval*

1. INTRODUCTION

Seismicity is a relatively new segment of geophysics. It studies complex processes of earthquakes occurrence, movements of seismic waves created in earth and beneath its surface together with its mechanical effects on the ground and on different objects.

Fast progress and development in all spheres of human civilization impose the need for regulation and suppression of all negative effects of such natural phenomena like earthquakes. Therefore, seismology, together with similar sciences, is undergoing constant advancement. Based on the seismic data of a particular region, as of detailed and systematic studies of relevant facts, it is possible to predict what consequences future earthquakes would cause. Then it would be possible to take precaution measures. Engineering seismology and earthquake engineering study how to deal with the negative effects of earthquakes and how to repress them. Primary task of these scientific studies is to find the best way and place to build safe objects and systems of objects in the regions with high seismic activity.

2. EARTHQUAKES AND SEISMICITY

Earthquake is a complex, dynamic process that manifests as a sudden brake within rocks of the Earth's crust, which burst under the release of accumulated stress. According to the way they occurred, earthquakes can be natural (tectonic, volcanic) and artificial (artificially accumulated lakes, etc).

Under severe pressure on the underground rocks of the Earths' crust, most of the time caused by sudden movement of bigger blocks, a sudden brake within rocks crops up, followed by elastic deformations of surrounding rocks, then transformed into waves that propagate through the earth. Permanent deformations (brakes) of rocks manifest as cracks in the rocks. This complex, tectonic process is all-together called earthquake, to be more precise- tectonic earthquake.

The exact point of rock breaking (earthquake genesis) is called a focal point of the earthquake. The focal point represents the place with maximum stress within rocks of the earth's crust. This is where the rock breaking starts. The focal point of the earthquake is also called a hypocenter, and its vertical projection on the earth's surface is called epicenter [1].

Magnitude is a measure of the energy released during the earthquake. In 1935, Charles Richter defined the magnitude as a power of an earthquake. The magnitude has come to be called the Richters' magnitude. It does not have an upper or lower limit and it is expressed by unnamed number. Earthquake intensity represents the effect of the earthquake on the earths' surface, people, natural and artificial objects. Mercalli Intensity Scale (MSC) also provides information on the earthquake effects on the surface. Intensity ratings are expressed as Roman numerals between I and XII.

Seismic hazard is the study of expected earthquake ground motions at any point on the earth. Seismic hazard is expressed by three elements: ground movement amplitude, return period and realization probability of such event.

Seismic risk represents the level of possible loss of material goods in the case of earthquake and relative numbers usually expresses it. Mathematically it is defined as convolution of seismic hazard and the function of object vulnerability (the quality or seismic resistance of the object).

The question of earthquake prediction remains the most frequently discussed question, especially in the areas with high seismic activity. Earthquake is a very complex phenomenon. Therefore, seismologists need to apply complex and multidisciplinary scientific tools in order to understand this phenomenon, especially how it can be successfully predicted. Nevertheless, each earthquake is followed by different geological phenomena and different information that manifest in various geophysical fields. These circumstances stimulate and motivate seismologists to explore further into the meter, and one day discover reliable tool for earthquake prediction. The earthquake phenomenon is studied by thousands of scientists throughout the world, but the results so far are modest and sporadical. However, there is a great hope that the reliable tool for earthquake prediction will be discovered soon enough.

When we say 'earthquake prediction' we usually mean so called short-term forecast. Short-term forecast expresses the time, place and magnitude of the earthquake. Time of the earthquake is expressed in terms of days. Short-term forecast is still in experimental phase, but many world wide seismic institutions conduct various researches on this issue. Haicheng earthquake in China, that happened on February 4th 1975. in Liaoning province with the magnitude of 7.3 and surface intensity of IX-X degrees MSC scale, was the only successful short-term forecast. Thanks to the expertise of Chinese seismologists, the earthquake was successfully detected. The whole region was evacuated and more than 100.000 lives were saved (some reports say 400.000). Unfortunately, in 1976, much stronger earthquake (with the magnitude of 7.6) had hit the neighboring province called Tangshan and took over 250.000 lives (some reports say 600.000). This shows that predicting this phenomenon is very complex and responsible task [1].

Beside short-term, there is long-term and mid-term earthquake prediction. Long-term prediction defines the area and return period (statistical period of the earthquake happening again). Long-term prediction is used for prediction of hazardous and catastrophic earthquakes in larger regions Seismicity maps show zones or regions of earthquake activity (of different intensity) which are suppose to happen in the coming period (50, even 100 years from now). These maps show exact region where the earthquake will happen (close to 70 per cent accurate prediction).

Mid-term forecast is connected to smaller areas- like tectonic plates. In these areas there is a greater chance for strong earthquake occurrence and its time frame is expressed in terms of years (or decades). As mentioned earlier, all earthquake predictions are result of complex multidisciplinary studies, and therefore, very expensive and realizable by a few, developed countries [1].

One of the main tasks of engineering seismology is the analysis of potential implication of future earthquakes on planed object or system of objects. Seismic risk must be taken into account when projecting, building or using the objects. In order to calculate the risk, we need to know the level of seismic hazard for the region where we plan to build.

Maximal amplitude of ground movement for the location of object is important factor that allows making planes for safe construction works in the areas with high seismic activity. Therefore, seismic modeling can be viewed as defining and forecasting confidence intervals for maximal ground motion amplitude or for return periods of specific amplitude values. This analysis represents defining seismic hazard for particular region. Observed region can be divided into smaller, homogeneous parts of the potential locations where we plan to build. For each location it is possible to calculate maximal amplitude of movement and their return periods. Later in the paper we will describe one of the methods used to solve the problem of that kind.

3. SEISMIC HAZARD ANALYSIS

Because of the unpredictable character of the earthquakes it is impossible to use deterministic methods. To define seismic hazard we need to use mathematical statistics and probability theory methods. Analysis of seismic hazard is based on empirical facts on seismic characteristics and geological texture of observed region. In this paper, we will describe the application of probability theory and statistics on simulation of future seismic activity in the observed region. Thanks to such simulations, seismologists gather data on how simulated earthquakes effect the locations in the observed region. Those data are then processed and hazard indicators for a given earthquake size and distance are defined. The level of data reliability is also examined in this analysis. The results are usually expressed as appropriate confidence intervals.

Following steps need to be taken before the simulation:

- *Analysis of gathered data on past earthquakes which happened in the same area (region) in order to define parameters for further analysis.*
- *Defining of the seismic sources*
- *Defining of stochastic model for seismicity*
- *Defining of the earthquake frequency for each source separately*
- *Selection of analytical expressions for attenuation of earthquake intensity as it moves away from the focal point*
- *Defining the expected effect of the earthquake i.e. distribution of parameters of ground movement in place and time*

At the beginning of the analysis it is necessary to specify the position and seismic characteristics of active geological faults which were sources of earthquakes in the past. Faults are cracks in the earth's crust resulting from the displacement of one side with respect to the other. They represent seismic sources of earthquakes. For each source it is necessary to specify maximal magnitude and earthquake frequency with different magnitudes. It is necessary to locate the source precisely, specify and model its location and size. Model needs to be simple enough to allow analysis, but complex enough to give an insight into real situation.

The source can be modeled as spotted, linear, or surfaced. What type of the modeled source will be used depends on the available information on geological ground texture and on earthquakes, which appeared in the past on the observed source. If there are enough data on geological fault texture, by chopping it on small pieces, we can conduct precise approximation. If there is not enough data available, approximation must be rougher. Because of the automatic data processing, it is necessary to adjust data on seismic source to fit the needs of the analysis [2]. For each defined source that is independent of others, simulation of seismic future can be created. Simulation of seismic future is based on the previously gathered data and on assumption that process of earthquake occurrence on the source is stationary.

Poisson's random process is used for generating earthquake occurrence. It is based on following assumptions:

- *Earthquakes are independent of time, which means that the number of earthquakes in one period does not depend on the number of earthquakes during the previous or following period*
- *Expected number of earthquakes for specific time period is constant measure (the process is stationary)*
- *Probability that two earthquakes are going to happen at the same time and on the same place is equal to zero*

Disregarding several shortcomings, these assumptions reflect the true nature of earthquake occurrence. Probability that there will be no earthquakes in the time period t is:

$$P_x(x=0|t) = \exp(-\mathbf{u} \bullet t),$$

where **u** is expected frequency of earthquake with a magnitude greater than lowest relevant for analysis and for the observed source. Probability of at least one earthquake occurrence in the t period of time is:

$$F_t(t) = P_x(x \geq l \mid t) = 1-\exp(-\mathbf{u} \bullet t)$$

$$f_t(t) = \frac{d}{dt} F_t(t) = \mathbf{u} \bullet \exp(-\mathbf{u} \bullet t) \qquad (3.1)$$

is the of probability density function for a period t between two earthquakes.

All these facts make the process of modeling earthquake occurrence possible (like Poisson's random process). The only parameter which appears in that case is **u**. First we need to specify minimal magnitude relevant for the analysis, i.e. the minimal magnitude of an earthquake which causes ground movement on the observed locations. Each earthquake with the magnitude greater than minimal is being analyzed. Parameter **u** represents the earthquake frequency - for earthquakes we know have the magnitude greater than the minimum relevant for the analysis. Maximal possible magnitude can be determined for each source, according to geological evidences. For computer processing we need:

M_{min}, M_{max} – *minimal relevant and maximal magnitude for the source*

3.1. RELATIONS CONNECTING SEISMIC PARAMETERS

Seismic activity of the source is described in analytic terms which connect the frequency and the magnitude of earthquake, and are determined on the basis of statistical processing of empirical evidences. According to data on the earthquakes which occurred in the past, B. Gutenberg and C.F.Richter in 1935 derived the dependence of number of earthquakes and the magnitude [3] as:

$$\mathbf{N(M) = f(M,A,t)}$$, where is

N(M) – *number of earthquakes with magnitude bigger than* **M**,

M- *Richter's magnitude*

A – *surface of the source*,

t – *period of observation*.

There are many terms in the literature which are recommended to describe this dependence. One of the models which are used is logarithmic linear dependence described as:

$$\mathbf{\ln N(M) = \alpha + \beta \bullet M}$$

Parameters α and β are determined in regression analysis of data about earthquakes which have already had focus in the observed source. α depends on the period of observation **t**, and β on the relative frequency of strong earthquakes compared to weak ones.

Number of earthquakes, normalized by time, would be

$\mathbf{N'(M)} = \dfrac{N(M)}{t}$, and

$$\ln N'(M) = \alpha - \ln t + \beta \bullet M = \alpha' + \beta \bullet M \qquad (3.1.1)$$

Part of fault along which movement of the earth occurs (during earthquake) is the focus of the earthquake. From it the waves of released energy spread out. Richter's magnitude is the measure for the quantity of released energy.

For the simulation of earthquake 3 characteristics of the focus are important: the type of focus, the dimension of focus and the position of the focus on the source.

The type of focus directly depends on the type of the source. All focuses of earthquake in the spotted source are spotted. Line focuses occur in the sources whose width is not important in comparison to length, i.e. in line sources. Focuses whose width cannot be ignored occur in polyhedral or level sources and are of the same type as the source [2].

The dimensions of the focus depend on the magnitude of earthquakes which occurred in the past. Many terms of connection between the length of the focus and the magnitude are known in the literature, much more than in the case of the width of focus.

Assumptions that are adopted in modeling of the position of focus are [4]:

- *There is equal opportunity for earthquake occurrence at whole source*
- *Position of earthquake focus is not correlated with position of the previous earthquakes*
- *Axis focus are parallel to correspondent axis of the source*

The position of earthquake focus (hypocenter) is determined by coordinates of its geometric center H, the most usually 3: geographic width, length and depth of focus.

3.2. ATTENUATION OF THE EARTHQUAKE INTENSITY

Energy released in earthquake becomes weaker with removing from focus. In this phase of the analysis it is necessary to determine parameters which would describe characteristics of earthquake on the observed location (maximal amplitude of movement, speed or, most usually, acceleration of the earth, the duration of earthquake and so on), and analytical term for relation among characteristics of earthquake in the focus, the distance of focus and chosen parameters. Which parameters and terms will be chosen depend on the goal of the analysis, but in the case when objective of the analysis is to define seismic hazard, parameter which is observed is often maximal amplitude of acceleration of the earth on the location. Among formulas which are suggested by many authors [5] we should choose the ones that describe this aspect in best way for the midpoint through which seismic waves moves from focus of the earthquake to the location of the object.

The example of such term (L.Esteva) is:

$$a = 5000 \bullet \exp(0.8 \bullet M) / (R + 40)^2 \qquad (3.2.1)$$

where is:

a – *maximal amplitude of the acceleration on the location*,

M – *Richter's amplitude*;

R – *the shortest distance from the focus of earthquake to the observed location.*

By increasing the distance from the focus of earthquake, value of maximal amplitude of acceleration and the effects of earthquake on the objects which have distant locations are decreasing. According to this, at certain distance, earthquake will not have any effect, regardless of its intensity in the focus. Thus for every source in the seismic model, the area of its influence can be determined. Out of the boundaries of that area, earthquakes, even the strongest earthquake with the focus in the source will not cause any damage to objects.

The boundary of influence area of the source is determined in this way:
we define the minimal acceleration of the earth, which is of the interest to analysis and the maximal magnitude of earthquake for the given source, and put them into the term for attenuation (3.2.1). By solving equation for **R**, we get value $\mathbf{R_0}$ **-** the distance in which the most intensive earthquake with the focus in this source has no effect. Because the location of the objects and source are in the three-dimensional area, point that are on the distance $\mathbf{R_0}$ from the source, form the surface which shape depends on the type of the source. Dissection of this surface and the earth surface is the wanted boundary of influence area of the source.

3.3. SIMULATION OF THE SEISMICITY

The next step in the analysis is the simulation of earthquakes, i.e. the simulation of occurrence time, magnitude, position and dimension of the focus for the earthquakes on the source and for each observed source in the region. In the analysis which is described here, it is supposed that each parameter of earthquake follows specific probability distribution function. The problem of seismicity simulation becomes the problem of modeling seismicity as multidimensional stochastic point process, i.e. to generate random numbers with corresponding probability distributions.

A process of earthquake occurrence is very complex, random in many components. Statistical analysis of data on earthquakes that had happened showed that some simplifications can be made. In this way we can model a process of earthquake occurrence as process random by next three components [3]:

- *The time of earthquake occurrence;*
- *The earthquake magnitude;*
- *The focus position at seismic source.*

It is supposed that these three parameters are not interdependent and that probability distribution of the components is determinate on the basis of empirical data from the past.

The simulation of earthquake is performed by random number generator. The simulation is performed for each seismic source separately, and synthesis of earthquakes of all seismic sources gives 'continuation' of seismic history for area that we are observing. At that, we are presuming that random process of earthquake occurrence is stationary.

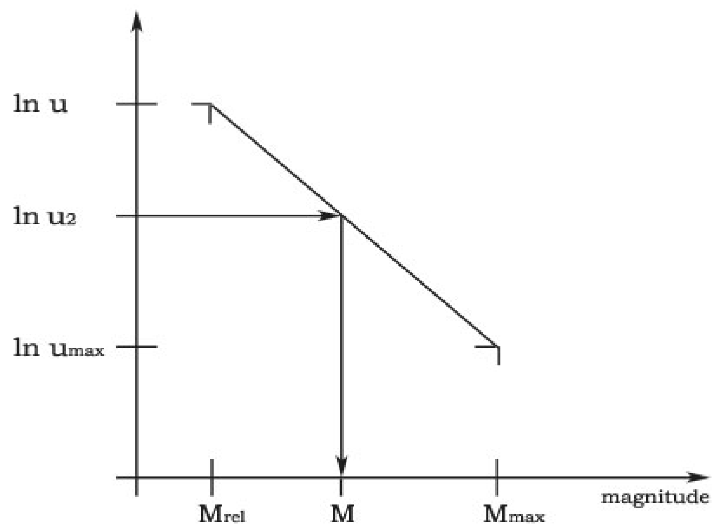
At the beginning of simulation, the time must be set to zero. After that, time is changing into random intervals that are corresponding to periods between earthquakes. A process of earthquake occurrence is modeled as Poisson's random process, and the period between earthquakes must be simulated adequately, according to the density function given in (3.1) [2].

Random value $\mathbf{u_2}$ is simulated in the interval
$(\mathbf{u_{max}}, \mathbf{u})$, where is:

$\mathbf{u_{max}}$ – *frequency of earthquake with the maximal magnitude for the source;*
u – *frequency of earthquake with the magnitude bigger than minimal relevance for analysis*.

For the simulated time of earthquake occurrence, the corresponding magnitude **M** is obtained from the relationship (3.1.1), represented in the picture 3.3.1.

The third component to be simulated is location of the earthquake focus center on seismic source. For the known simulated magnitude of earthquake, dimension of focus are determined on the basis of data evidence for the earthquakes which occurred in the past on the observed source. Than the location of the centre of the focus is simulated on the seismic source as random value with uniform distribution. For each focus three numbers are simulated: longitude, amplitude and depth of focus. Boundaries for intervals in which these numbers can be found can be easily determined when characteristic of seismic source and magnitude of earthquake are known.



Picture 3.3.1 Simulation of earthquake magnitude

By the simulation of these three components, the earthquake is completely defined at its focus, and it is possible to observe its effects in the area of influence. The procedure of earthquake simulation is repeated for the same period and for all seismic sources where earthquakes can occur and whose effects are felt in area for which seismic hazard is to be defined.

3.4. EARTHQUAKE CHARACTERISTICS ON THE LOCATION

Parallel with the simulation, data on characteristics of earthquake on the location are collected. To collect enough data for the statistical analysis, simulation should last more than a thousand years. According to small numbers of data from the past, we cannot be sure that the process of earthquake occurrence should be still stationary. That is why 'seismic history' obtained in the process of simulation, does not mean long-term forecast for future earthquakes but only gives data for statistical analysis. The results of the analysis can be treated as a good prediction for the period much shorter than the period of the simulation itself.

When we know time, place and magnitude of the earthquake, by the application of equations for attenuation, we might define consequences of the earthquake at location. Due to the characteristics of simulated earthquake in the focus, it is possible to determinate:

$\mathbf{a_o = F_a(M,R,K,L)}$

$\mathbf{a_o}$ - *maximal magnitude of acceleration of ground motion on the location*,

$\mathbf{F_a}$- *expression for attenuation of acceleration*,

M - *magnitude of simulated earthquake,*

R - *distance from location to focus,*

K - *source type,*

L - *ground type on location*.

By statistical processing of data collected during the simulation, we can obtain data which will represent parameters of seismic hazard at as much location as it is observed, needed and determinated by analyze goal. If the area included in the analysis is extensive and heterogeneous, the number of locations for which the data are collected must be bigger than in the case when the estimation of hazard is done for limited and homogeneous area.

Return periods for specific levels of amplitudes of ground acceleration are approximated by the first moments ('sample means' **RP**) and their dispersions by the second moments ('corrected samples dispersion' $\mathbf{S^2}$) of the corresponding data collected during the simulation of earthquakes. Thus calculated values are also random variables with specific probability distributions.

3.5. CONFIDENCE INTERVAL FOR RETURN PERIOD OF THE ACCELERATION LEVEL

Return period of ground acceleration level could be determinate by adequate confidence interval. It is possible to determinate left and right limits of confidence intervals for different acceleration levels. The 'real' return period **RRP** belongs to this interval with probability ν. ν is confidence level and is inversely proportional to interval length.

Statistics:

$$t_{n-1} = \frac{RP - RRP}{S} * \sqrt{n} \qquad (3.5.1)$$

has Student’s t distribution with n-1 degrees of freedom [6]. **RP** is mean of return periods for the given acceleration level and location, obtained from the simulation data.

For confidence level ν, in Student's t table, it is possible to find number $t_{n-1,1-\nu}$, for which

$$P(|t_{n-1}| \leq t_{n-1,1-\nu}) = \nu \qquad (3.5.2)$$

Using and transforming expressions (3.2.1) and (3.2.2) it can be obtained following:

$$P(\mathbf{RP}\text{-}t_{n-1,1-v} * \frac{S}{\sqrt{n}} \leq \mathbf{RRP} \leq \mathbf{RP} + t_{n-1,1-v} * \frac{S}{\sqrt{n}}) = \nu$$

In other words, return period for the any level of acceleration, with probability ν, is in interval whose limits are:

$$G_L = \mathbf{RP} - t_{n-1,1-v} * \frac{S}{\sqrt{n}}$$

$$G_D = \mathbf{RP} + t_{n-1,1-v} * \frac{S}{\sqrt{n}}$$

Final report of the analysis described in this paper would refer to the locations in area in which are collected data during simulation and would represent data about limits of confidence intervals for return periods of different ground movement amplitude at observed locations. By the analysis of gathered data for different locations in the region, it would be possible to sketch adequate maps which would express the results in aspect which is needed and used in circumstances of defining seismic hazard. Such carts, more or less precise, depending on the purpose, would represent good base for projecting and building of objects or systems of objects in areas with high seismic activity.

4. CONCLUSION

In this paper we have described one of the possible aspects of seismic hazard definition in areas with high seismic activity. Seismic and geological data and analysis are used for determination of parameters which are used in 'future seismic history' simulation for the area. The data about the implications of simulated earthquakes on the locations in observed region are gathered during the simulation and processed in the end of it, in order to define seismic hazard for the region.

Advantage of such approach is that during the ‘seismic history’ simulation it is possible to gather data about different consequences that can occur at the observed locations. For the objects that are planed to be build there it can be observed what would be the return periods of different levels of damage on the objects. If the task of seismology engineers is to plan and build some infrastructure system in the area, by this approach it is possible to predict effects of future seismic activity on the system and to choose the optimal solution [2].

BIBLIOGRAPHY

[1] http://www.seismo.cg.yu;

[2] Bulajić M., ″*Analiza seizmičkog rizika za sistem objekata infrastrukture* ″, Magistarski rad, Institut za zemljotresno inženjerstvo i inženjersku seizmologiju, Univerzitet ″Kiril i Metodij″, Skoplje, 1985;

[3] Ogata, Y., “*Space-time Point Process Models for Earthquake Occurrences*”, Annals of the Institute of Statistical Mathematics, 50,1998;

[4] Shah H.C. and Benjamin J.R., *″Earthquake Engineering and Seismic Risk Analysis ″*, The John A. Blume Earthquake Engineering Center, Stanford University, Stanford, 1983;
[5] Mihailov V., *″Prilog stohastičkom modeliranju seizmičnosti ″* , Doktorska disertacija, Zagreb, 1978;
[6] Vukovic, N., "*PC statistika i verovatnoca*", Fakultet organizacionih nauka, Beograd, 2004;
[7] Minster, J.B., and Williams, N., "*Systematic global testing of intermediate-term earthquake prediction algorithms*", Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, Sand Diego, 2004;
[8] Seher, T.and Main, I. G., "*A statistical evaluation of a 'stress-forecast' earthquake*", The University of Edinburgh, 2003.