Stochastic nature of earthquake ground motion

Srđan Kostić a,∗, Nebojša Vasović b, Matjaž Perc c, Marinko Toljić a, Dobrica Nikolić d

Department of Geology, University of Belgrade Faculty of Mining and Geology, Belgrade, Serbia
Department of Applied Mathematics, University of Belgrade Faculty of Mining and Geology, Belgrade, Serbia
Department of Physics, Faculty of Natural Science and Mathematics, University of Maribor, Maribor, Slovenia
Department of Physics, University of Belgrade Faculty of Mining and Geology, Belgrade, Serbia

HIGHLIGHTS

● Deterministic nature of earthquake ground motion is challenged by surrogate data.
● Earthquake ground motion is likely to be the product of a random process.
● Ground accelerations belong to a class of linear stationary stochastic processes.
● Determinism test indicates a high degree of stochasticity in the examined data.

ARTICLE INFO

Article history:
Received 2 April 2013
Received in revised form 7 April 2013
Available online 3 May 2013

Keywords:
Earthquake
Surrogate data
Deterministic chaos
Stochasticity

ABSTRACT

In this paper, we analyze the irregular behavior of earthquake ground motion as recorded during the Kraljevo M5.4 earthquake, which occurred on November 3rd, 2010 in Serbia. We perform the analysis for the ground accelerations recorded at 6 seismological stations: Grua, Ruda, Rada, Bara, Zaga and Bdv. The latter were carefully chosen based on their corresponding tectonic zone and the local geological setting. For each station, we analyze the horizontal component of the ground acceleration in the north–south direction, which is the one of primary interest for engineering design. We employ surrogate data testing and methods of nonlinear time series analysis. The obtained results indicate that strong ground accelerations are stochastic, in particular belonging to a class of linear stationary stochastic processes with Gaussian inputs or distorted by a monotonic, instantaneous, time-independent nonlinear function. This type of motion is detected regardless of the corresponding tectonic setting and the local geological conditions. The revealed stochastic nature is in disagreement with the frequently assumed deterministically chaotic nature of earthquake ground motion.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Modeling of strong ground motion is one of the most challenging problems in earthquake engineering, since it represents the key input in earthquake resistant design [1,2]. Namely, a dynamic aseismic design of high rise buildings and large structures commonly uses strong earthquake accelerograms as input excitations [3]. Nevertheless, the site at which a major construction is required seldom contains past strong motion records. For such sites the simulation of strong motion is a required task. However, in order to be able to generate a synthetic accelerogram, the type of seismic source, attenuation characteristics and wave propagation for each specific site have to be determined [4], which is not always a straightforward
process. The main reason for this is that each recording represents the result of the physical interaction of many complex processes. The final signal is the convolution of each operation that transfers energy from the particle to the station [5,6]. Apparently, the seismic energy, radiated during the motion along the rough fault surface, is partially scattered by random heterogeneities, and further distorted due to interaction with the surface of the Earth. The two most important attenuating factors are certainly tectonic setting and local geological composition, with significant influence on the final shape of strong ground motion. Previous research showed that the magnitude of the acceleration is controlled by the tectonic stress in the region around the fault and by the frictional stress along the fault surface [7]. Hence, the recorded accelerograms have their own different characteristics depending on the location where earthquake occurred. Concerning this, every earthquake that produces good strong motion records gives a new insight into the physical phenomena of ground shaking.

During the past decades, much effort has been given to obtaining reliable simulations of strong ground motion, using theoretical or semi-empirical models of the parameters affecting the shape, the duration and frequency content of strong motion records. Generally, there are two main approaches to ground motion simulation: deterministic and stochastic. One of the simplest techniques for predicting strong ground motion from the largest earthquakes is the deterministic empirical Green’s function method. Hartzell [8] first utilized observed records from small events as Green’s functions to simulate the mainshock records. Since then, his original idea has been widely used and developed [9,4,10–12]. On the other hand, a stochastic method was introduced by Hanks and McGuire [13], who indicated that the observed high frequency ground motion can be represented by the windowed and filtered white noise. The band limited white-noise method, firstly described by Boore [14], has also been widely applied to simulate ground motion from point sources [15,16].

However, despite these numerous works on engineering models of earthquake motion, there is still a lack of direct experimental evidence confirming its stochastic or deterministic nature. The underlying complexity of earthquake dynamics is still being widely investigated, indicating the possibility that the existing models may not capture all the relevant information relating to the seismic events [17,18]. The results of previous analyses suggest that strong ground motion could be a deterministic feature. Pavlos et al. [19] applied a chaotic analysis approach to an earthquake time series recorded in the area of Japan in order to test the assumption that the earthquake process could be the manifestation of a chaotic low dimensional process. The results of their study show that the underlying mechanism, as expressed by the time series, can be described by low dimensional chaotic dynamics. Yang et al. [20] used qualitative and quantitative methods to verify the chaotic behavior of earthquake ground motion recorded at 24 stations during the great $M_w$ 7.6 Chi–Chi earthquake. Their results indicate that the strong earthquake ground motion exhibits chaotic and fractal characteristics rather than pure random behavior. We have to emphasize that irregular behavior may often be mistakenly advertised as chaos, where in fact unpredictable disturbances render the behavior stochastic rather than deterministically chaotic [21]. Indeed, this fact was a strong motivation for us to conduct thorough research on recorded strong ground motion.

In this paper, our aim is to examine the possible deterministic/stochastic behavior of earthquake ground motion by applying methods of nonlinear time series analysis [22]. These methods were previously not used in the field of seismology, even though they were successfully applied in many other fields of research [23], including Earth and geophysical sciences [24–26]. These studies have proved that nonlinear time series analysis methods have great potential in studying the various types of experimentally recorded time series.

The presented analysis was done for the recorded acceleration, since the peak ground acceleration represents one of the main parameters that enters the calculation for civil engineering design [27]. For every recording, we analyze only the horizontal component of the recorded acceleration in the north–south direction, because, from the engineering point of view, the horizontal accelerations are commonly used to describe the ground motion due to their natural relationship to inertial forces [28,29]. Six different accelerograms are studied, which are recorded at stations located in different tectonic zones, and with various local geological conditions.

Hence, the aim of our research is twofold: we want to confirm or reject the hypothesis that the strong ground motion is essentially stochastic and we want to exclude the plausible influence of tectonic setting and local geological conditions on the stochastic/deterministic character of the recorded ground accelerations.

The scheme of this paper is as follows. In Section 2, we give brief description of Kraljevo M5.4 earthquake, together with the description of acceleration recordings, where we indicate the choice of accelerometer stations. In Section 3, we perform surrogate data analysis, by testing three null hypotheses on the stochastic nature of recorded ground acceleration. Next, we conduct a determinism test, on the basis of the calculated optimal embedding delay and minimum embedding dimension. In the final section we discuss the obtained results, with suggestions for further research.

2. Kraljevo M5.4 earthquake

The wider area of Kraljevo is prone to frequent seismic activity, which is initiated by movement along the fault system in the west/northwest–south/southeast direction, along which transcurrent left lateral movements occur. These structures are developed along the southern margin of the Cacak–Kraljevo basin, as a part of the Trans-Serbian dislocation. Subsequent seismic activity propagates also along faults in the northwest–southeast direction, which are developed at the northeastern margin of the Cacak–Kraljevo basin. These structures are the part of an older tectonic build, which was reactivated during the Oligocene–Miocene period.

In this paper, we study the earthquake ground motion caused by the last major earthquake M5.4, that occurred near Kraljevo on 3rd November 2010. According to the Seismological Survey of Serbia, the earthquake epicenter was located
121 km south from Belgrade, in the village Sirca, along the 8 km long fault. The hypocentral depth was about 13 km. After the main quake, a series of aftershocks was registered in the wider zone of Kraljevo, with magnitudes in the range from 1.0 to 4.4. The epicentral map of the Kraljevo earthquake is shown in Fig. 1. For the nonlinear analysis of the recorded ground motion, 6 different accelerographic stations were chosen: Grua, Ruda, Rada, Bara, Zaga and Bdva, according to the following two criteria:

- different tectonic zone. Although different approaches of tectonic zonation of Serbia exist due to complex geological composition [30,31], we accept a suggestion by Marovic et al. [30], who proposed the following division of tectonic units in Serbia: Dinarides, Hellenides, Serbo-Macedonian massif, Carpatho-Balkanides and large basins: Pannonian and Dacian basin. The distribution of chosen stations is shown in the tectonic map of Serbia (Fig. 2):
  - Grua is located in the Cacak–Kraljevo basin, closest to the earthquake epicenter,
  - Ruda is located in the Vardar zone of Dinarides,
  - Rada belongs to the Ophiolite zone of Dinarides,
  - Bara belongs to the edge of the Neogene–Quaternary Leskovac basin, at the boundary with the Serbian–Macedonian massif,
  - Zaga is located in the pile of Gethian overthrusts of Carpatho-Balkanides,
  - Bdva belongs to the Pannonian Basin.

- various local geological conditions, according to Eurocode 8\(^1\) ground types [32]:
  - stations Grua, Rada, Bara and Zaga are settled in ground type A (rock or other rock-like geological formation, including at most 5 m of weaker material at the surface),
  - Ruda is settled in ground type B–C (deposits of very dense to medium dense sand, gravel, or very stiff clay, with thickness from several tens to many hundreds of meters, characterized by a gradual increase of mechanical properties with depth),
  - Bdva is settled in ground type E (soil profile consisting of a surface alluvium layer with thickness varying between about 5 and 20 m, underlain by stiffer material with shear wave velocity \(V_s > 800 \text{ m/s}\)).

\(^1\) Eurocode 8 represents a set of design requirements in civil engineering for earthquake resistant structures.
The ground acceleration was recorded by the Seismological Survey of Serbia, with equally spaced intervals of 0.0048 s (Grua), 0.0049 s (Bara and Rada) and 0.005 s (Ruda, Bdva and Zaga), which gave the recording length between 40 s and 89 s (and corresponding data size, as well, in the range 9600–17 800), depending on duration of ground motion at the specific site. All the recordings were preprocessed in order to reduce noise and side effects of the original signal (bandpass filtering, instrument-and baseline-correction). The recorded acceleration time series which are examined in this paper are shown in Fig. 3. It is evident that the amplitudes of ground motion depend on the distance from the epicenter—the highest values of the acceleration are recorded at stations in the epicentral area.

3. Surrogate data analysis

Surrogate data analysis is one of the most important steps in preprocessing the observed data, since it enables us to compare the value of possible nonlinear statistics for the observed data and the approximate distribution for various classes of linear systems. In other words, by applying the method of surrogate data, we can test whether the data have some characteristics which are distinct from stochastic linear systems [33].

The testing is performed by assuming that the observed data belong to some class of linear systems (null hypothesis). In this case, we test three null hypotheses: (1) data are independent random numbers drawn from some fixed but unknown distribution; (2) data originate from a stationary linear stochastic process with Gaussian inputs and (3) data originate from a stationary Gaussian linear process that has been distorted by a monotonic, instantaneous, time-independent nonlinear function [34]. The process of surrogate data testing is the following. Firstly, in order to ensure that the testing results are valid, we generate 20 surrogates for each of the three null hypotheses, as already proposed in Ref. [22]. Then, in order to compare the original data and generated surrogates, we calculate the zeroth-order prediction error $\varepsilon$ [34]. If the zeroth-order prediction error for the original recordings ($\varepsilon_0$) is smaller in comparison to the calculated error for the surrogate data ($\varepsilon$), then a null hypothesis can be rejected. On the other hand, if $\varepsilon_0 > \varepsilon$ at any instance of the test, the null hypothesis is confirmed. Usually, more than one wrong result out of 20 is not considered acceptable [35]. Our aim is to achieve a significance level of $\alpha = 0.95$ when confirming or rejecting a null hypothesis.

In this paper, the surrogates are generated by using Matlab toolkit MATS, developed by Kugiumtzis and Tsimpiris [36], while the zeroth-order prediction error $\varepsilon$ is calculated according to the algorithm in C suggested by Kantz and Schreiber [35].
As was already emphasized in the introductory section, the surrogate data analysis is done for the acceleration recordings at 6 seismological stations, namely Grua, Ruda, Rada, Bara, Zaga and Bdva, which belong to different tectonic zones with various local geological conditions.

We start with a very simple null hypothesis that the data are independent random numbers drawn from some fixed but unknown distribution. In order to test this null hypothesis we generate 20 surrogates by randomly shuffling the data (without repetition), thus yielding surrogates with exactly the same distribution yet independent construction. Then we calculate the zeroth-order prediction error for the original recording ($\varepsilon_0$) and for each of the 20 generated surrogates ($\varepsilon$). In all cases, except for the acceleration recording at Grua station, $\varepsilon_0$ is smaller than $\varepsilon$, which allows us to reject the null hypothesis with significance level 95%.

As a next step, in order to test the second null hypothesis that data originate from a stationary linear stochastic process with Gaussian inputs, we employ the phase randomization analysis. In this case, adequate surrogate data sets must contain correlated random numbers which have the same power spectrum as the data. This is the case if the Fourier transforms of the original data and the surrogates differ only in their phases but have the same amplitudes which go into the power spectrum. This could be achieved by randomizing the Fourier surrogates of the original data, and then by computing the inverse transform to obtain randomized time series. The detailed description of this method is given in Ref. [27]. In this case,
we also constructed 20 surrogates for each of the recorded accelerograms. The results are shown in Fig. 4. Obviously, we could not reject the null hypothesis (at 95% significance level) since $\varepsilon_0$ is within $\varepsilon$ for all the tested surrogates.

In order to test the third null hypothesis that the data originate from a stationary Gaussian linear process that has been distorted by a monotonic, instantaneous, time-independent nonlinear function, we calculate the amplitude adjusted Fourier-transformed (AAFT) surrogates [37]. This AAFT algorithm guarantees that the distribution and power spectrum will be approximately preserved. This could be achieved by the following procedure. After rescaling the original data to a normal distribution, a Fourier-transformed surrogate of the rescaled data is constructed. The final surrogate is then scaled to the
distribution of the original data. As in the previous case, we generated 20 surrogates for each of the observed cases and calculated the prediction error $\varepsilon$ (Fig. 5).

As apparent from Fig. 5, rather interesting behavior appears when testing the third null hypothesis. The null hypothesis could not be rejected for the ground acceleration recorded at Grua station (Fig. 5(a)) and Bdva station (Fig. 5(f)), since $\varepsilon_0 > \varepsilon$ for at least one surrogate and several prediction time steps $n$. In all the other cases, the null hypothesis could be rejected, with significance level 95%. This kind of prediction behavior could result from the very nature of the applied method itself.
since the generation of amplitude adjusted surrogates results in changes to the power spectrum of the final surrogate, which further causes power spectrum whitening of the original data [38]. Hence, we further use a method of iterated AAFT surrogates, by performing a series of iterations in which the power spectrum of the AAFT surrogate is adjusted back to that of the original data before the distribution is rescaled back to the original distribution. This is obtained by adjusting back the amplitudes of the Fourier transformed AAFT surrogates to the Fourier transformed surrogates of the rescaled original data. The obtained surrogates are then inverse Fourier transformed and rescaled back to the original data distribution by sorting the original data according to the ranking of the Fourier-transformed surrogate [29]. These two steps are iterated a number of times (in our case 500), until the whitening of the power spectrum becomes sufficiently small. As in the previous cases, we generated 20 such surrogates and calculated the zeroth-order prediction error $\varepsilon$ (Fig. 6). It is clear that, in all cases, $\varepsilon_0$ is well within $\varepsilon$, allowing us to reject the null hypothesis.

4. Determinism test

The results of the surrogate data analysis could be further confirmed by applying the determinism test, which enables us to verify if the studied time series originates from a deterministic process. In this paper, in order to examine the possible deterministic nature of the underlying process, we use the method developed by Kaplan and Glass [39], which assumes that if a time series originates from a deterministic process, it can be described by a set of more or less complex first-order ordinary differential equations. This further implies that its vector field can be drawn easily. If the time series originated from a deterministic system, the obtained vector field should consist solely of vectors that have unit length, indicating the average length of all directional vectors $\kappa$ to be equal to 1. If the solutions in the phase space are to be unique, then the unit vectors inside each box may not cross, since that would violate the uniqueness condition at each crossing. In other words, if the system is deterministic, the average length of all directional vectors $\kappa$ will be 1, while for a completely random system $\kappa \approx 0$ [40]. However, it should be emphasized that Kaplan and Glass [39] indicated that an average length of the vector of the deterministic system equals 1 only for the infinitely long data set.

In order to apply this test, we have to embed the observed scalar series into the appropriate phase space via the embedding procedure, originally proposed by Takens [41]. The embedding is performed by using the open-source software, already applied in Refs. [22–24]. As a first step of embedding, we need to determine the optimal embedding delay. Here we use a technique that measures the general dependence of two variables, such as the average mutual information method. According to Ref. [42], the value of $\tau$ that produces the first local minimum of mutual information should be used for phase portraits. The values of optimal embedding delay calculated by using the average mutual information method for all the observed time series are: $\tau = 9$, $\tau = 6$ and $\tau = 14$, for Grua, Ruda and Rada, respectively, and $\tau = 14$, $\tau = 13$ and $\tau = 29$, for Bara, Zaga and Bdva, respectively.

The next step in our analysis is to determine the minimal required embedding dimension $m$ in order to fully resolve the complex structure of the attractor. In this paper, we use the procedure suggested in Ref. [43], which identifies the number of “false nearest neighbors”, points that appear to be nearest neighbors because the embedding space is too small. We use the criterion which utilizes the fact that the normalized distance between the embedding coordinates of two presumably neighboring points is larger than a given threshold ($R_{tr}$), if these two points are false neighbors. According to Kennel et al. [35], the value of $R_{tr} = 10$ proves to be a good choice for most data sets. This suggestion also agrees with the proposition made by Abarbanel [44], who stated that in practice, for values of $R_{tr}$ in the range $10 \leq R_{tr} \leq 50$, the number of false neighbors identified by this criterion is constant. The performed analysis showed that the fraction of false nearest neighbors drops to zero only for the recorded acceleration at stations from the same tectonic unit as the epicentral area ($m = 6$, $m = 12$ and $m = 13$ for Grua, Ruda and Rada). In all other cases, the fraction of false nearest neighbors rises with the increase of embedding dimension, which could indicate a high level of stochasticity in the input data. Since all the recordings are preprocessed, which included bandpass filtering, instrument-and baseline-correction, this increase could be explained by the high level of stochastic components in the original time series.

Now we can conduct the determinism test. For this purpose, we coarse grained the previously calculated embedding space into maximum $42^6$, $42^12$ and $42^{13}$ grids, for Grua, Ruda and Rada, respectively. To calculate the determinism factor, we included only those boxes visited at least one time by the trajectory. The obtained values of determinism factor $\kappa$ are 0.76, 0.89 and 0.85 for the recordings at Grua, Ruda and Rada, respectively, which indicates the possible stochastic signature. At the same time, although the embedding dimension is needed as an input parameter for the determinism test, we also examined acceleration recordings at other three stations (Bdva, Zaga and Bara) for different values of embedding dimension (Fig. 7). As apparent from Fig. 7, the value of the determinism factor $\kappa$ is in the range 0.5–0.9, indicating the possible stochastic nature of the underlying process.

5. Discussion

Investigating the time dynamics of complex natural systems is one of the most important scientific challenges, and, in particular, recognizing the dynamical properties in seismicity data could give a better understanding of the geophysical mechanisms underlying tectonic processes. In this paper, we demonstrate that the earthquake ground motion evolves as linear stochastic process, regardless of the corresponding tectonic setting and local geological conditions.
The analysis of the recorded acceleration time series was performed in several steps:

- The basic method used in this paper, in the search for possible stochastic behavior, was surrogate data analysis \([22,35,37,38]\). The results of testing the three null hypotheses indicate that stochasticity is an important factor in the dynamics of earthquake ground motion,

- The results of surrogate data analysis were further confirmed by the application of a determinism test \([39]\). Rather low values of the determinism factor \((\kappa < 0.9)\) indicate the possible absence of determinism in the recorded ground motion.
It should be noted that results of certain analyses suggest that earthquakes represent deterministic features for some range of parameter values. The main focus of these works was the frequency and distribution of the registered earthquakes, in the search for deterministically chaotic behavior [45–47]. Also, nonlinear dynamical analysis of the Burridge–Knopoff model, as a phenomenological model of earthquake nucleation, indicates the possibility of deterministically chaotic behavior [48–50]. However, in all of the previously examined cases, strong ground motion was not examined explicitly, so the results of our analyses are not directly opposed to these previous research, but they do suggest that more than determinism, stochastic behavior, as is traditionally associated with economic data [51–54], governs earthquake ground motion. We propose our results to be a complement to the study of earthquakes as complex phenomena.

It has to be emphasized that similar data exist for other earthquakes recorded at different locations globally. Our aim was firstly to examine the character of ground motion for one specific earthquake, and then to expand our study to various accelerograms, recorded for different registered earthquakes.

Taking all of this into consideration, there are two main conclusions from our analysis. Firstly, our investigation strongly suggests that earthquake ground motion represents a linear stochastic process, which is in contrast to the works of Pavlos et al. [19] and Yang et al. [20], who stated that strong ground motion is chaotic. Secondly, we showed that the tectonic setting and local geological conditions at the location where the seismological station is settled has no influence on the character of strong ground motion.

Although other more detailed studies and further work will be needed, we believe that our results are important for several reasons. Methods which are used here are numerical on the whole, which is a rather convenient tool for engineers to conduct the analysis without engaging with complicated and lengthy analytical tools. On the other hand, by confirming the stochastic behavior of the earthquake ground motion, we lend support to the validity of the stochastic approach in examining strong ground motion. In other words, our results justify the use of the methods of stochastic ground motion simulation in engineering design.

In order to improve our work, it could be interesting to study the other two components of the ground motion and to examine the possible difference in the dynamics at the same or different recording stations. Moreover, the effect of different types of seismic waves on the ground motion could be thoroughly investigated. It is known that longitudinal (P) waves have a primary effect on ground motion in the vertical direction, while transversal (S) waves mainly influence the horizontal ground motion. Surface waves are detected in all three channels at the seismograph. By separating the initial signal into these three different parts, and by examining separately each part, we could evaluate the influence of each of the seismic waves on ground motion. On the other hand, it could be useful to apply mean field method from statistical physics, in order to investigate the complex interactions and dynamics among different tectonic zones.

Acknowledgments

The authors would like to thank the Seismological Survey of Serbia on providing the very reliable and extensive data. Special thanks also go to Uroš Đurić from University of Belgrade regarding the quality of Fig. 1. This research was supported...
by the Serbian Ministry of Education, Science and Technological Development (Grants 176016, 171017, 176015 and 33003) and the Slovenian Research Agency (Grant J1-4055).

References