Stochastic Time-Series Prediction Equation Using Wavelet Packets for Iran

MOHAMMADREZA NAJAFTOMARAEI,1 HABIB RAHIMI,1 G. TANIRCAN,2 and MOHAMMAD SHAHVAR3

Abstract—In this study, a stochastic simulation model proposed by Yamamoto and Baker (Bulletin of the Seismological Society of America 103:3044–3056, 2013) is applied to the Iranian strong motion database, which comprises more than 3828 recordings for the period between 1975 and 2018. Each ground motion is decomposed into wavelet packets. Amplitudes of wavelet packets are divided into two groups, and for each group, model parameters are estimated using the maximum likelihood method. Regression coefficients are then obtained relating model parameters to seismic characteristics such as earthquake magnitude, distance, and site condition. Inter-event residuals of coefficients and correlation of total residuals of those parameters are also calculated. An inverse wavelet packet transform is used to reconstruct the amplitudes in the time domain and perform the simulation. Finally, a validation test is performed. The comparison of ground motion intensity measures for recorded and simulated time series shows acceptable conformity in the application. The estimated parameters using the simulated data agree with the real data, indicating the acceptable validity of the estimated stochastic simulation model. The obtained regression equations can generate ground motions for future earthquake scenarios in Iran.

Keywords: Stochastic time-series prediction equation, wavelet packets, strong motion, Iran.

1. Introduction

The Iranian Plateau is located on the Alpine–Himalayan orogenic belt, and due to the northward motion of the Arabian plate toward Eurasia is experiencing continuous deformation. This deformation occurs mainly in the Alborz, Kopeh Dagh, and Zagros Mountains (Fig. 1a). Except for the Makran subduction zone to the southeast of the Iranian Plateau, where an oceanic lithosphere subducts northward under southeastern Iran (Nilforoushan et al., 2003), a continental form of the Arabian-Eurasian plate convergence causes the present tectonics of Iran (Jackson & McKenzie, 1984) that results in the tectonically complex region with structures of different density, velocity, etc. The Iranian Plateau is known as one of the most seismically active areas of the world. The 1909 Silakhor (Mw 7.4), 1962 Buin Zahra (Mw 7.2), 1972 Gir (Mw 6.7), 1978 Tabas (Mw 7.4), 1990 Manjil-Rudbar (Mw 7.4), 2003 Bam (Mw 6.6), 2012 Ahar-Varzaghan (Mw 6.4), and 2017 Ezgeleh (Mw 7.3) earthquakes are among the most destructive to date. Active seismic sources, young tectonics, and convergence rate due to regional and global tectonics have caused major earthquakes with extensive destruction and casualties over the past 50 years in Iran, where there are many populated industrial cities with critical infrastructures. This indicates a crucial need in seismic risk assessment and its mitigation tools.

Strong ground motion recordings play an indispensable role in earthquake engineering for investigating the individual characteristics of earthquakes and the response of structures under excitation. In almost all seismic regulations, structural design spectra are formed utilizing strong ground motion intensity measures (GMIM, e.g., peak ground acceleration, PGA), and spectral quantities (e.g., spectral acceleration, SA) calculated through hazard assessments. Assessments are essentially performed utilizing ground motion prediction equations (GMPE) compatible with a specific region. Hence, the precise calculation of GMPE is directly dependent on the number of strong motion recordings in an area. Strong motion simulation techniques are convenient ways to produce abundant ground motion...
as an input for structural analysis in earthquake engineering, especially for the regions that will be urbanized and that lack sufficient strong motion data. In addition, synthetic accelerograms provide demand parameters of earthquake engineers that can be extracted from them and are allowed in several design regulations for nonlinear time history analyses of structures. Generating synthetic ground motions whose features match the characteristics of the

Figure 1

(a) Five tectonic regions of Iran (Mirzaei et al., 1998). The distribution of the (b) 892 strong motion stations and (c) 1636 earthquakes studied in this study.
recorded ground motions can be a substitute approach.

There have been several works on ground motion simulation of earthquakes in Iran. Among them, Nicknam et al. (2009) present a probabilistic-based strong motion compatible with the source path and site soil conditions given the probability of exceedance for the citadel of Arg-e-Bam site bedrock. Soghrat et al. (2011), developed a GMPE for soil and rock sites in northern Iran, based on the calibrated specific barrier model (SBM) against the latest available strong motion data. Hassani et al. (2012), implemented stochastic finite-fault modeling based on dynamic corner frequency to simulate the 22 February 2005 (Mw 6.4) Zarrand earthquake. Zafarani and Soghrat (2012) developed a physically-based GMPE for soil and rock sites in the Zagros region based on the SBM. Zafarani et al. (2013) used a hybrid simulation method to calculate broadband ground motion time histories at the bedrock level for deterministic earthquake scenarios on the NTF, and used a hybrid technique to simulate the ground motion caused by an earthquake with Mw = 4, which occurred close to Tehran on 17 October 2009 and is the first local earthquake that has ever been recorded by the local strong ground motion network in Tehran. Gholami et al. (2013), applied a new analytical methodology for computing synthetic seismograms in three-dimensional (3-D) anelastic media to model the local records of the 2003 Bam Mw 6.6 earthquake, Southeastern Iran. Zafarani et al. (2015), employed a generalized inversion of S-wave amplitude spectra for deriving the site response and S-wave attenuation (quality factor) in the northwest region of Iran. Riahi et al. (2015), simulated the 2003 Bam earthquake acceleration recordings at three strong motion stations using the empirical Green’s function method. Vahidifard et al. (2017), presented a simulation of three components of near-field ground shaking recorded during the mainshock at three stations on 16 September 1978 in Tabas (Mw 7.4), Iran, using a hybrid method (composed of discrete wavenumber method) and stochastic finite-fault modeling based on a dynamic corner frequency.

Literature reviews of several models of earthquake ground motion construction performed by Douglas and Aochi (2008) and Ólafsson et al. (2001) show that all those techniques can be categorized into three groups: (1) The deterministic approach, where the model is based on physical principles, produces ground motions by modeling those parameters that affect their shape, duration, and frequency content (e.g., Atkinson & Boore, 2007; Atkinson & Somerville, 1994; Boore, 1983, 2003; Hanks, 1979; Hanks & McGuire, 1981; Pitarka et al., 1998; Silva et al., 1999). (2) The empirical or stochastic simulations are based on the assumption that future scenarios will be akin to those observed in earlier events and so try to match with target seismological and probabilistic features (e.g., Anderson, 1997; Abrahamson & Shedlock, 1997; Abrahamson et al., 2008; Baker & Cornell, 2006; Bommer & Alarcon, 2006; Campbell, 1986; Douglas, 2003; Joyner & Boore, 1988; Lee et al., 2000; Power et al., 2008; Scherbaum et al., 2004; Watson-Lamprey & Abrahamson, 2006). (3) Hybrid approaches combine factors of both (e.g., Campbell, 2003, 2007; Douglas et al., 2006; Hartzell et al., 2002; Hisada, 2008; Mai & Beroza, 2003; Pitarka et al., 2000; Scherbaum et al., 2006; Tavakoli & Pezeshk, 2005).

As mentioned earlier, some of the essential issues regarding the studies are as follows: (1) researchers performed ground motion simulations of a specific earthquake, not for a region or even for a whole country; (2) GMPEs predict a particular ground motion parameter (e.g., PGA, SA), not stories; (3) GMPEs predict ground motion at different frequencies with indeterminate time; (4) the research does not consider non-stationarity. As is known, temporal and spectral non-stationarity (changing amplitudes and changing frequency characteristics of time series, respectively) can affect nonlinear dynamic structural response.

Hence, methods that can reproduce nonstationary characteristics of ground motions are needed.

Amiri et al. (2011), predicted wavelet packet coefficient amplitudes using a neural network approach. Their simulations do not consider any seismological parameters such as source, path, and site effects, so ground motions containing all available variables of possible future earthquakes generated with their models are somewhat difficult.

Discrete wavelet transform (DWT) is another approach to generating synthetic ground motions.
Yamamoto and Baker (2013) (hereafter, YB2013) proposed a stochastic method that uses wavelet packet transform (WPT) to describe the waveform in the time and frequency domain and finally simulates time series using inverse wavelet transform. The reasons for selecting WPT among wavelet analysis approaches are as follows: (1) Because of the non-orthogonality of the wavelets at neighbor times and frequencies, reconstructing time series using the continuous wavelet transform (CWT) is complicated. (2) Time localization of low-frequency amplitudes is an issue in DWT, as mentioned above. (3) Controlling resolution in the time and frequency domains in WPT, a developed copy of the DWT, is easier because the basis functions are orthogonal, making it easy to reconstruct time series. (4) Simulated ground motions resulting from their proposed stochastic ground motion model are variable because they affect the variability of structural responses (e.g., Rezaeian, 2010; Rezaeian & Der Kiureghian, 2012). The YB2013 method has been successfully applied to strong motion data in various parts of the world. Among them, Huang and Wang (2015) used wavelet packets and cokriging analysis to simulate spatially distributed ground motions over a region stochastically.

Stochastic simulations, in general, require fewer parameters and are computationally less expensive than physics-based and hybrid simulations. Still, they do not consider physical phenomena because there are almost no parameters in the stochastic model to control these effects. The method proposed by YB2013 overcomes this problem. Unlike the previous works that simulated a specific earthquake, this method can be applied to a tectonic region. In addition, previous GMPEs can estimate only GMIMs, but YB2013 predicts the whole time series. Unlike the earlier studies, this approach locates the frequencies in time and considers temporal and spectral non-stationarity.

Considering the above-discussed superiority of the YB2013 over conventional stochastic methods, the YB2013 method is utilized to simulate the substantial, strong ground motion recordings in Iran. The motivation of this study comes from the fact that no work is available robustly predicting ground motion time series for Iran with an extensive strong ground motion database.

We begin this paper by introducing the general procedure of YB2013. Then the process is extended to Iranian strong ground motion data. Finally, a comparison of GMIMs of simulated versus recorded strong ground motion and SA at various periods is performed to validate the approach.

2. Methodology

YB2013 proposed a stochastic simulation method that uses wavelet packet transform to describe the waveform in the time and frequency domain and simulate time series using inverse wavelet transform. The significant advantage of this method is that it does not necessitate detailed information on the source rupture process, propagation path, or local site effects. The success of the physics-based and hybrid simulation techniques depends on the potential for characterizing those fundamental inputs. Fewer parameters are needed for the YB2013 stochastic simulation method. Compared to physics-based and hybrid simulations, they are computationally less expensive. Thirteen parameters are requisite in their proposed model that can be obtained by regression analysis, which comprises proper variability matching with the properties of the simulated ground motions. PGA, Arias intensity (Arias, 1970), and significant duration (Trifunac & Brady, 1976) of simulated ground motions obtained by the model and recorded ground motion were compared. The results showed a relatively appropriate agreement.

The following steps of the stochastic simulation model introduced by YB 2013 are summarized. The whole scheme of their stochastic modeling is also shown in Fig. 2.

2.1. Decomposing Time Series Using Wavelet Packet Transform

In the first step, the coefficients of the wavelet packets are obtained. To decompose a ground motion time series into wavelet packets in the time and
frequency domain, WPT was employed, and it is defined as follows:

\[ C_{j,k} = \int_{-\infty}^{\infty} x(t) \psi_{j,k}^* dt \]  

\[ n = FT \]  

where \( x(t) \) is the time series, \( C_{j,k} \) is the WPT coefficient, in which \( i \) is its location on the frequency axis, and \( j \) is the scale parameter number that shows the frequency resolution. \( K \) is its location in time and \( \psi_{j,k}^* \) is the wavelet packet function. \( n \) is the number of wavelet packet coefficients; \( T \) is the number of time steps and \( F \) is the number of frequency points. The mother wavelet function used here is the Meyer wavelet (Meyer, 1986). It has orthogonality characteristics and proper positioning in the time and frequency axis. It is computed from the finite-impulse response-based approximation. To minimize the residual of simulated data, wavelet packets in the lowest frequency (as low as 0.1 Hz) level had to be controlled; to this end, the \( dt \) and \( j \) were selected to be 0.02 s and 8, respectively.

2.2. Dividing Wavelet Packet Coefficients into Two Groups of Major and Minor

Amplitudes of WPT were divided into two groups, major (containing 70% of the total energy) and minor, because WPT has few coefficients with large amplitudes (less than 1% of the total number of wavelet packet coefficients). These two groups are united as follows to obtain the total wavelet packet coefficients:

\[ |C_{j,k}|^2 = |C_{j,k,\text{maj}}|^2 + |C_{j,k,\text{min}}|^2 \]  

where \( C_{j,k,\text{maj}} \) and \( C_{j,k,\text{min}} \) are the wavelet packet coefficients in the major and minor groups, respectively.

The energy fraction criteria maximize the difference in the characteristics of the wavelet packet coefficients of the major and minor groups.

2.3. Finding Parameters for Both Major and Minor Groups

The 13 parameters that explain the time and frequency characteristics of the acceleration time series are given in Table 1. Equation 4 is obtained for all amplitudes (major and minor groups), Eqs. 5–15 are obtained for two groups separately, and Eq. 15 is obtained for just the major group. Note that the 13th parameter is the standard deviation of independent and identically distributed lognormal random variables with median 1 and logarithmic standard deviation of the residual of the wavelet packets in the minor group computed from the bivariate lognormal function. In both groups, the correlations between time and frequency are transformed into a new parameter using the cumulative density function of the standard normal distribution, called a transformed coefficient because it is more appropriate for prediction using regression analysis.
Table 1
Thirteen parameters that explain the time and frequency characteristics of the acceleration time series

| Parameter name                  | Parameter denotation | Group name | Parameter name                  | Parameter denotation | Group name |
|---------------------------------|----------------------|------------|---------------------------------|----------------------|------------|
| 1. Temporal centroid            | $E(t)_{maj}$         | Major group| 7. Temporal centroid            | $E(t)_{min}$         | Minor      |
| 2. Temporal variance            | $S^2(t)_{maj}$       |            | 8. Temporal variance            | $S^2(t)_{min}$       |            |
| 3. Spectral centroid            | $E(f)_{maj}$         |            | 9. Spectral centroid            | $E(f)_{min}$         |            |
| 4. Spectral variance            | $S^2(f)_{maj}$       |            | 10. Spectral variance           | $S^2(f)_{min}$       |            |
| 5. Correlation between time and frequency of wavelet packet coefficients | $\rho(t,f)_{maj}$ |            | 11. Correlation between time and frequency of wavelet packet coefficients | $\rho(t,f)_{min}$ |            |
| 6. Mean of squared amplitudes of the WT packets | $E(a)_{maj}$ |            | 12. Standard deviation of the cumulative squared acceleration of the WT coefficients | $E_{acc}$ |            |
| 13. Cumulative squared acceleration | $E_{acc}$ | Total WT coefficients | |

\[
E_{acc} = \sum_i \sum_k |C_{j,k}^i|^2
\]  
(4)

\[
E(t)_{maj} = \frac{\sum_i \sum_k t_{i,k_{maj}} |C_{j,k_{maj}}^i|^2}{E_{acc}}
\]  
(5)

\[
E(t)_{min} = \frac{\sum_i \sum_k t_{i,k_{min}} |C_{j,k_{min}}^i|^2}{E_{acc}}
\]  
(6)

\[
S^2(t)_{maj} = \frac{\sum_i \sum_k \{t_{i,k_{maj}} - E(t)_{min}\}^2 |C_{j,k_{maj}}^i|^2}{E_{acc}}
\]  
(7)

\[
S^2(t)_{min} = \frac{\sum_i \sum_k \{t_{i,k_{min}} - E(t)_{min}\}^2 |C_{j,k_{min}}^i|^2}{E_{acc}}
\]  
(8)

\[
E(f)_{maj} = \frac{\sum_i \sum_k f_{i,k_{maj}} |C_{j,k_{maj}}^i|^2}{E_{acc}}
\]  
(9)

\[
E(f)_{min} = \frac{\sum_i \sum_k f_{i,k_{min}} |C_{j,k_{min}}^i|^2}{E_{acc}}
\]  
(10)

\[
S^2(f)_{maj} = \frac{\sum_i \sum_k \{f_{i,k_{maj}} - E(f)_{maj}\}^2 |C_{j,k_{maj}}^i|^2}{E_{acc}}
\]  
(11)

\[
S^2(f)_{min} = \frac{\sum_i \sum_k \{f_{i,k_{min}} - E(f)_{min}\}^2 |C_{j,k_{min}}^i|^2}{E_{acc}}
\]  
(12)

\[
\rho(t,f)_{maj} = \frac{\sum_i \sum_k \{t_{i,k_{maj}} - E(t)_{maj}\} \{f_{i,k_{maj}} - E(f)_{maj}\} |C_{j,k_{maj}}^i|^2}{S(t)_{maj}S(f)_{maj}E_{acc}}
\]  
(13)

\[
\rho(t,f)_{min} = \frac{\sum_i \sum_k \{t_{i,k_{min}} - E(t)_{min}\} \{f_{i,k_{min}} - E(f)_{min}\} |C_{j,k_{min}}^i|^2}{S(t)_{min}S(f)_{min}E_{acc}}
\]  
(14)

\[
E(a)_{maj} = \sum_i \sum_k |C_{j,k_{maj}}^i|^2
\]  
(15)

where $E_{acc}$ is the cumulative squared acceleration for the total wavelet packet coefficients, $C_{j,k}^i$ is the wavelet packet amplitudes, $E(t)$ and $E(f)$ are the center of the time (temporal centroid) and frequency (spectral centroid) location of the wavelet packet coefficients, respectively, $S^2(t)$ is the temporal variance, $S^2(f)$ is the spectral variance, $\rho(t,f)$ is the correlation between the time and frequency of the wavelet packet coefficients, and $E(a)_{maj}$ is the mean of the squared amplitudes of the wavelet packet coefficients of a major group. In all equations, maj denotes the major group, and min represents the minor group.

2.4. Finding Coefficients of Regression Analysis of Those 13 Parameters (Regression Analysis)

To simulate the ground motion time series, 13 parameters of the model must be predicted as a function of the source, path, and site of future scenarios through a two-stage regression analysis...
approach (Joyner & Boore, 1993, 1994) with a moment magnitude ($M_w$), hypocentral distance ($R_{hyp}$), rupture distance ($R_{rup}$), and average shear-wave velocity within 30 m depth ($V_{S30}$) as predictor variables.

The 13 model parameters are estimated using the maximum likelihood estimation to obtain the regression coefficient. The functional form used to conduct a regression analysis is as follows:

$$Y = \alpha + \beta_1 M + \beta_2 \ln(M) + \beta_3 \exp(M) + \beta_4 \ln(\sqrt{R_{hyp}^2 + h^2}) + \beta_5 \ln(V_{S30}) + \eta + \sigma$$

Except for the case of correlation, $Y$ is the natural logarithm of a model parameter (for correlation, $Y$ is the transformed coefficient), and $\eta$ and $\delta$ are intra and inter-event residuals, respectively. The parameters $\alpha$ and $\beta_i$ are the regression coefficients that must be predicted. Other details of the regression analysis can be found in the original paper. Through the source parameters, the logarithmic $\ln(M)$ term is used only for predicting $E(a)_{maj}$ and $E_{acc}$ because, at the large magnitude, they may confront saturation.

2.5. Major and Minor and Group of Wavelet Packets

The predicted parameters must lead to distributions in time and frequency to simulate future ground motions of a target event.

A statistical survey of wavelet packets shows that large wavelet packet coefficients differ in time and frequency distributions from those of small ones. Therefore, the WPT coefficients were divided into two groups, the major and minor groups. Observations from the major wavelet packet coefficients demonstrate that their locations depend on time and frequency but are independent of their amplitudes. To illustrate this statistically, the squared amplitudes have exponential distributions, and the locations are consistent with bivariate lognormal distributions. Parameters and characteristics of the functional form of WPT coefficients in both groups are listed in Table 2.

2.6. Inverse Wavelet Packet Transform

After obtaining wavelet packet coefficients for both groups through the above-mentioned distributions, which are acquired by estimated parameters calculated via regression equation and combining them, the inverse WPT is used to reconstruct the time series from wavelet packets as follows:

$$x(t) = \sum_{i=1}^{2^j} \sum_{k=1}^{2^{N-i}} C_{j,k} \psi_{j,k}^i$$

where $2^N$ is the number of time steps in the time series. Another required case for reconstructing the time series is the sign of the amplitudes of the wavelet packet obtained here randomly because it does not have a significant effect on the ground motion characteristics.
3. Strong Motion Data and Data Processing

The strong motion database of the Road, Housing, and Urban Development Research Center (BHRC) was used between 1975 and 2018. The database comprises 1636 events (Fig. 1c) with Mw between 3 and 7.5. A total of 3828 recordings of 892 stations (Fig. 1b) with hypocentral distances less than 300 km were selected. We selected those events that were recorded in at least five stations. Stations are equipped with SMA1, Guralp, or SSA2 accelerometers with a sampling rate of 200 samples per second. Stations and epicentral distribution of earthquakes are shown in Fig. 1b and c. Figure 3a and b show magnitude versus Repi and Vs30 versus Repi, respectively. From the 3828 three-component waveforms used in this study, 1469 records were not associated with measured Vs30 data (from 504 stations), so we used the estimated Vs30 obtained by these records in Najaftomaraei et al. (2020). Their approach estimates V\textsubscript{S30} stations, using a normal distribution and PDF and CDF of the natural logarithm of horizontal-to-vertical ratio (H/V) of PGA and PSAs at 24 periods (388 stations).

A suitable window for high-quality accelerograms must be selected for initial data processing. To choose the appropriate section of waveforms, arrival times of P and S waves were determined. The considered window starts 2 s before the P wave arrival and continues until the end of the waveform. A linear trend removal was applied to the data. Low-quality data with a signal-to-noise ratio [SNR, Eq. (17)] smaller than 3 (SNR < 3) were eliminated. Further data with no pre-event and even those with no P wave (truncated data) were not considered. Since having information about the style of faulting is restricted to just some of the events, the effect of this parameter on regression analysis was neglected, which needs to be studied and improved.

\[
\text{SNR} = \frac{\text{RMS of Signal}}{\text{RMS of Noise}} = \sqrt{\frac{\sum Y^2}{n_{\text{sig}}}} \sqrt{\frac{\sum N^2}{n_{\text{noise}}}} \tag{18}
\]

where \(Y\) is signal amplitudes, \(N\) is noise amplitudes, \(n_{\text{sig}}\) and \(n_{\text{noise}}\) define the number of samples of signal and noise, respectively, and RMS is the root mean square.

4. Stochastic Simulation

Essentially major steps given in the methodology section were followed for the simulation. WPT was applied to the processed ground motion recordings in the first phase of the analysis. For example, time, frequency, and wavelet packet coefficients at the corresponding time and frequency of an arbitrary signal are shown in Fig. 4, and the same plot for the 2000/2/7 Mw 4.2 event recorded at Ab-bar station are shown in Fig. 5. We can see the whole wavelet packet transform amplitudes of the record in this figure; large amplitudes are concentrated at about
12.5 s and between 3 and 4 Hz, and it is clear that the number of small amplitudes is much greater than that of large ones. This fact leads to the idea of dividing the amplitudes into two groups. The time and frequency nonstationary characteristics can also be seen in this figure.

Then, 13 model parameters were estimated using the observed time series. And regression coefficients were calculated utilizing two-stage regression analysis in MATLAB. For the two-stage regression analysis, we apart from the relation into two regression equations: The first one is related to intra-event effects considering distances and site conditions, and the second is associated with the inter-event impacts with the characteristics of the earthquakes (Joyner & Boore, 1993, 1994). To predict all parameters, regression analysis was carried out using the functional form of Eq. 4 for all 3828 records. Figure 6 shows the Major and Minor amplitudes of recordings given in Fig. 5. The major wavelet packet amplitudes are the largest, consisting of 70% of the total energy (less than one percent of the whole number of coefficients). The remaining packets with smaller amplitudes are placed in the minor group. The relationship between wavelet packet coefficients and some of these parameters is shown in Fig. 7. In this figure, $E(t)$ is the temporal centroid, which is associated with the time of the decisive part of the ground motion, $S^2(t)$ is the temporal variance related to significant duration containing 90% of the total energy, and $E(f)$ is the spectral centroid. It is associated with the dominant frequency, and $S^2(f)$ is the spectral variance related to significant bandwidth that contains 90% of the total energy.
In the final phase, stochastic simulations at 3828 acceleration time series from 1636 earthquakes utilizing the calculated stochastic time-series prediction model.

5. Results and Discussion

The resulting regression coefficients are shown in Table 3. In Table 4, the correlations of total residuals for each pair of model parameters have been tabulated. There can be some factors that do not exist in the regression equation, which may influence the model parameters computed from observed ground motions. Thus, a correlation must be performed through the intra- and inter-event residuals.

To evaluate and compare the simulated data with the observations, the parameters related to the two categories (simulated and observed data) are plotted relative to each other and are shown in Fig. 8. These parameters are the PGA and spectral accelerations (Sa) at \( T = 0.01; 0.06; 0.5; 1; 5; 10 \) and 20 s, Arias intensity (Arias, 1970), and significant duration. As shown in Fig. 8, the estimated parameters from the simulated data are in good agreement with the real data, indicating the acceptable validity of the estimated stochastic simulation model. Figure 8a is simulated versus observed PGA, Fig. 8b is the significant duration of simulated ground motion versus observed, Fig. 8c is Arias intensity of simulated ground motion versus observed, Fig. 8d–j show simulated versus observed PSA at \( T = 0.01; 0.06; 0.5; 1; 5; 10 \) and 20 s, respectively. Generally, the model can produce the observed PGA and PSA over...
Figure 7
Relationship between wavelet packet coefficients and some of the parameters used in regression analysis. $E(t)$ and $E(f)$ are the center of the time (temporal centroid) and frequency (spectral centroid) location of the wavelet packet coefficients, respectively. $S^2(t)$ is temporal variance, $S^2(f)$ is the spectral variance.

Table 3

|                  | $\alpha$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\beta_5$ | $\sigma$ | $\tau$ | $\sigma*$ |
|------------------|----------|-----------|-----------|-----------|-----------|-----------|---------|-------|---------|
| $E(t)_{Major}$   | 2.362    | 0         | 0         | -0.0002   | 0.57      | -0.21     | 0.4     | 0.29  | 0.49    |
| $E(f)_{Major}$   | 0.142    | -0.003    | 0         | 0         | -0.45     | 0.23      | 0.46    | 0.33  | 0.56    |
| $S(t)_{Major}$   | 1.828    | 0         | 0         | -0.0002   | 0.59      | -0.25     | 0.45    | 0.34  | 0.57    |
| $S(f)_{Major}$   | -0.062   | -0.022    | 0         | 0         | -0.45     | 0.19      | 0.48    | 0.36  | 0.6     |
| $\rho_{t,f}^{2}_{Major}$ | -10.936 | 0.165     | 0         | 0         | 2.38      | 0.42      | 2.6     | 5.03  | 5.66    |
| $E(t)_{Minor}$   | 1.873    | 0         | 0         | -0.0002   | 0.53      | -0.21     | 0.38    | 0.28  | 0.47    |
| $E(f)_{Minor}$   | -0.705   | -0.008    | 0         | 0         | -0.33     | 0.22      | 0.4     | 0.28  | 0.48    |
| $S(t)_{Minor}$   | 2.591    | 0         | 0         | -0.00015  | 0.48      | -0.21     | 0.37    | 0.28  | 0.47    |
| $S(f)_{Minor}$   | -0.133   | -0.007    | 0         | 0         | -0.18     | 0.17      | 0.35    | 0.26  | 0.44    |
| $\rho_{t,f}^{2}_{Minor}$ | 0.318   | 0.009     | 0         | 0         | -0.05     | 0.1       | 0.3     | 0.22  | 0.37    |
| $E(a)_{Major}$   | 6.613    | 0.275     | -0.76     | -0.76     | -0.15     | -0.46     | 1.75    | 1.45  | 2.27    |
| $E_{acc}$        | 10.555   | 0.122     | -0.21     | -0.207    | 0.02      | -0.57     | 1.63    | 1.35  | 2.11    |
| $S(\eta)$        | 1.4      | 0         | 0         | 0         | 0         | 0.06      | 0       | 0.06  | 0.06    |
Table 4  
Correlation of total residuals

| $E(t)_{\text{Major}}$ | $E(f)_{\text{Major}}$ | $S(t)_{\text{Major}}$ | $S(f)_{\text{Major}}$ | $\rho'_{f, \text{Major}}$ | $E(t)_{\text{Minor}}$ | $E(f)_{\text{Minor}}$ | $S(t)_{\text{Minor}}$ | $S(f)_{\text{Minor}}$ | $\rho'_{f, \text{Minor}}$ | $E(a)_{\text{Major}}$ | $E_{\text{acc}}$ |
|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| 1.00 | 0.43 | 0.89 | 0.3 | 0.05 | 0.88 | 0.36 | 0.8 | 0.24 | 0.02 | 0.11 | 0.13 |
| 0.43 | 1.00 | 0.49 | 0.9 | 0.13 | 0.39 | 0.73 | 0.42 | 0.5 | 0.13 | 0.2 | 0.18 |
| 0.89 | 0.49 | 1.00 | 0.35 | 0.08 | 0.83 | 0.37 | 0.8 | 0.25 | 0.08 | 0.12 | 0.13 |
| 0.3 | 0.9 | 0.35 | 1.00 | 0.18 | 0.29 | 0.71 | 0.32 | 0.5 | 0.05 | 0.21 | 0.18 |
| 0.05 | 0.13 | 0.08 | 0.18 | 1.00 | 0.16 | 0.18 | 0.21 | 0.21 | -0.02 | 0.17 | 0.15 |
| 0.88 | 0.39 | 0.83 | 0.29 | 0.16 | 1.00 | 0.38 | 0.97 | 0.26 | 0.07 | 0.13 | 0.14 |
| 0.36 | 0.73 | 0.37 | 0.71 | 0.18 | 0.38 | 1.00 | 0.39 | 0.82 | 0.04 | 0.3 | 0.26 |
| 0.8 | 0.42 | 0.8 | 0.32 | 0.21 | 0.97 | 0.39 | 1.00 | 0.28 | 0.1 | 0.18 | 0.18 |
| 0.24 | 0.5 | 0.25 | 0.5 | 0.21 | 0.26 | 0.82 | 0.28 | 1.00 | 0.02 | 0.39 | 0.36 |
| 0.02 | 0.13 | 0.08 | 0.05 | -0.02 | 0.07 | 0.04 | 0.1 | 0.02 | 0.02 | 0.12 | 0.09 |
| 0.11 | 0.2 | 0.12 | 0.21 | 0.17 | 0.13 | 0.3 | 0.18 | 0.39 | 0.12 | 1.00 | 0.96 |
| 0.13 | 0.18 | 0.13 | 0.18 | 0.15 | 0.14 | 0.26 | 0.18 | 0.36 | 0.09 | 0.96 | 1.00 |

Figure 8  
Comparison of GMIMs of simulated versus recorded strong ground motion and spectral accelerations at various periods: a PGA, b significant duration, c Arias Intensity, d SA at $T = 0.01$ s, e SA at $T = 0.06$ s, f SA at $T = 0.5$ s, g SA at $T = 1$ s, h SA at $T = 5$ s, i SA at $T = 10$ s, j SA at $T = 20$ s

Figure 9  
Recorded and simulated ground motion of the 2015/08/25 Firoozkooh earthquake (Mw 4.2) at Tehran82. Station: a and b acceleration (cm/s²), c and d velocity (cm/s), e and f displacement (cm), g Fourier amplitude spectrum (cm/s), h SA versus period for recorded and simulated ground motion, respectively
periods from 0.01 to 20 s, but for $T = 0.01$, $T = 5$, and $T = 10$ s, data distribution around the equality line ($y = x$) is more acceptable.

For all records, wavelet packet coefficients and resulting time series were simulated. Finally, PGA, significant duration, Arias Intensity, and SA at 7
periods were extracted for simulated and recorded ground motion. The left panels of Figs. 9, 10, 11, and 12 simulated and observed acceleration, velocity, and displacement time series are shown for the tabulated earthquake in Table 5. In addition, the Fourier and response spectra of simulated and recorded ground motions are displayed in the right panel of Figs. 9, 10, 11, and 12.

These figures show the contrast between the recorded and simulated acceleration time series. A match between the observed and simulated acceleration waveforms, velocity, and displacement is acceptable. The durations of the simulated and observed time series are similar and show qualified matching. The estimated PSA is more significant at long periods than observed, and the model overestimates it. Fourier amplitude spectra matching at frequencies lower than about 8 Hz seems to be the best.

As shown in Table 5, the magnitude of these earthquakes was considered from 4.2 to 7.3 to evaluate the validity of the estimated simulation model in a broader range of magnitudes. Also, the effect of shear wave velocity in 30 m (V_{30}), which indicates the site effect below the stations, was from 296 to 700 km/s for the selected stations to evaluate the performance of the derived simulation model according to different soil effects. In addition, the epicentral distance for the mentioned stations varied from 42 to 111 km (Table 5).

Due to the limitations of the stochastic simulation method and the lack of consideration parameters such as slip distribution on the fault, 3-D velocity, and attenuation mode, the acceptable agreement between the simulated and observed acceleration, velocity and displacement can be seen in the time domain. On the other hand, as it is clear in Figs. 9, 10, 11, and 12, there is an almost good agreement between the Fourier and response spectra of simulated and observed time series.

### 6. Conclusion

To simulate the ground motion time series, 13 statistical model parameters extracted from wavelet packet analysis were obtained and estimated using the maximum likelihood method using the recorded ground motion from the BHRC database for Iran. These parameters were assumed to function with seismic properties such as earthquake magnitude, distance, and Vs30. So through a two-stage regression analysis, the regression coefficient was obtained. Intra- and inter-event residuals of coefficients and correlation of total residuals of those 13 parameters were also calculated. Because temporal and spectral non-stationarity can be governed by this computationally inexpensive method, and since there are some regions in Iran where mega constructions will be established in the future but which also lack sufficient recorded ground motions for engineering purposes, the time-series prediction equation coefficient was obtained to simulate ground motions, followed by a validation test for all records, comparing simulated and recorded parameters such as PGA, significant duration, Arias Intensity, and Sa at seven periods, as well as for four earthquakes. This comparison of engineering demand parameters for recorded and simulated time series showed excellent consistency. Indeed, because of our constraints on data, such as not having the style of faulting, this study can be improved by inputting some other seismological parameters in the model to make it more realistic.

| Record # | Earthquake name | Date year/month/day | Mw | Record ID | PGA (g%) | Epicentral distance (km) | Station name | Vs30 (m/s) |
|----------|-----------------|---------------------|----|-----------|----------|------------------------|--------------|------------|
| 1        | Firoozkooh      | 2015/08/25          | 4.2| 6587      | 0.99     | 111                    | Tehran82     | 296        |
| 2        | Ahar            | 2012/08/11          | 6.1| 5602      | 57       | 72                     | Meshkinshahr  | 700        |
| 3        | Sarpolzahab     | 2017/11/12          | 7.3| 7377-01   | 270      | 46                     | Ghoorsefid   | 403        |
| 4        | Babamonir       | 2011/03/05          | 5.2| 5210-01   | 94       | 42                     | Ghaemiyeh    | 617        |
Acknowledgements

We offer our sincere appreciation to BHRC for providing us with a strong ground motion database. We especially appreciate the Editor and anonymous reviewers for their constructive comments. We acknowledge the Road, Housing, and Urban Development Research Center (BHRC) (BHRC) for providing the required waveforms for this study.

Funding

The article was supported by the Iranian National Science Foundation (grant number 9902274).

Declarations

Conflict of interests The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

Abrahamson, N., Atkinson, G., Boore, D., Bozorgnia, Y., Campbell, K., Chiou, B., Idriss, I. M., Silva, W., & Youngs, R. (2008). Comparisons of the NGA ground-motion relations. *Earthquake Spectra, 24*(1), 45–66. https://doi.org/10.1193/1.2924363

Abrahamson, N. A., & Shedlock, K. M. (1997). Overview. *Seismological Research Letters, 68*(1), 9–23.

Amiri, G. G., Shahjouei, A., Saadat, S., & Ajallooeian, M. (2011). Hybrid evolutionary-neural network approach in generation of artificial accelerograms using principal component analysis and wavelet-packet transform. *Journal of Earthquake Engineering, 15*(1), 50–76.

Anderson, J. G. (1997). Nonparametric description of peak acceleration above a subduction thrust. *Seismological Research Letters, 68*(1), 86–93.

Arias, A. (1970). A measure of earthquake intensity. In R. J. Hansen (Ed.), *Seismic design for nuclear power plants* (pp. 438–483). MIT Press.

Atkinson, G. M., & Boore, D. M. (2007). Earthquake ground-motion prediction equations for Eastern North America. *Bulletin of the Seismological Society of America, 97*(3), 1032. https://doi.org/10.1785/0120070023

Atkinson, G. M., & Somerville, P. G. (1994). Calibration of time history simulation methods. *Bulletin of the Seismological Society of America, 84*(2), 400–414.

Bai, Q., Fu, L. H., Bao, W. B., Jin, S. J., & Zhang, D. S. (2012). An ensemble of ground motions simulation based on discrete stationary wavelet transform. *Applied Mechanics and Materials, 166–169, 2408–2411. https://doi.org/10.4028/www.scientific.net/amm.166-169.2408

Baker, J. W., & Cornell, C. A. (2006). Spectral shape, epsilon and record selection. *Earthquake Engineering and Structural Dynamics, 35*(9), 1077–1095. https://doi.org/10.1002/eqe.571

Bommer, J. J., & Alarcon, J. E. (2006). The prediction and use of peak ground velocity. *Journal of Earthquake Engineering, 10*(1), 1–31.

Boore, D. M. (1983). Stochastic simulation of high-frequency ground motions based on seismological models of the radiated spectra. *Bulletin of the Seismological Society of America, 73*(6), 1865–1894.

Boore, D. M. (2003). Simulation of ground motion using the stochastic method. *Pure and Applied Geophysics, 160*(3–4), 635–676. https://doi.org/10.1007/PL00012553

Campbell, K. W. (1986). An empirical estimate of near-source ground motion for a major, Mb = 6.8, earthquake in the eastern United States. *Bulletin of the Seismological Society of America, 76*(1), 1–17.

Campbell, K. W. (2003). Prediction of strong ground motion using the hybrid empirical method and its use in the development of ground-motion (attenuation) relations in eastern North America. *Bulletin of the Seismological Society of America, 93*(3), 1012–1033.

Campbell, K. W. (2007). Validation and update of hybrid empirical ground motion (attenuation) relations for the CEUS. Tech. rep., ABS Consulting, Inc. (EQECAT), Beaverton, USA, Award number: 05HQGR0032.

Douglas, J. (2003). Earthquake ground motion estimation using strong-motion records: A review of equations for the estimation of peak ground acceleration and response spectral ordinates. *Earth-Science Reviews, 61*(1–2), 43–104.

Douglas, J., & Aochi, H. (2008). A survey of techniques for predicting earthquake ground motions for engineering purposes. *Surveys in Geophysics, 29*, 187–220.

Douglas, J., Bungum, H., & Scherbaum, F. (2006). Ground-motion prediction equations for southern Spain and southern Norway obtained using the composite model perspective. *Journal of Earthquake Engineering, 10*(1), 33–72.

Gholami, V., Hamzehloo, H., Ghayamghamian, M. R., Vaccari, F., & Panza, G. F. (2013). Simulation of the 2009, M w = 4 Tehran earthquake using a hybrid method of modal summation and finite difference. *Journal of Geophysics and Engineering, 10*(2), 025007. https://doi.org/10.1088/1742-2132/10/2/025007

Hanks, T. C. (1979). b values and x-c seismic source models: Implications for tectonic stress variations along active crustal fault zones and the estimation of high-frequency strong ground motion. *Journal of Geophysical Research, 84*(B5), 2235–2242.

Hanks, T. C., & McGuire, R. K. (1981). The character of high-frequency strong ground motion. *Bulletin of the Seismological Society of America, 71*(6), 2071–2095.

Hartzell, S., Leeds, A., Frankel, A., Williams, R. A., Odum, J., Stephenson, W., & Silva, S. (2002). Simulation of broadband ground motion including nonlinear soil effects for a magnitude 6.5 earthquake on the Seattle fault, Seattle, Washington. *Bulletin of the Seismological Society of America, 92*(2), 831–853.
Hassani, B., Zafarani, H., & Farjoodi, J. (2012). Stochastic Finite-Fault Simulation of 22 February 2005 (Mw 6.4) Zarand Earthquake (Iran), based on Dynamic Corner Frequency.

Hisada, Y. (2008). Broadband strong motion simulation in layered half-space using stochastic Green’s function technique. *Journal of Seismology, 12*(2), 265–279. https://doi.org/10.1007/s10950-008-9090-6

Huang, D., & Wang, G. (2015). Stochastic simulation of regionalized ground motions using wavelet packets and cokriging analysis. *Earthquake Engineering & Structural Dynamics, 2015*(44), 775–794. https://doi.org/10.1002/eqe.2487

Jackson, J., & McKenzie, D. (1984). Active tectonics of the Alpine–Himalayan belt between western Turkey and Pakistan. *Geophysical Journal of the Royal Astronomical Society, 77*, 185–264.

Joyner, W. B., & Boore, D. M. (1988). Measurement, characterization, and prediction of strong ground motion. In: Proceedings of earthquake engineering & soil dynamics II, Geotechnical division, ASCE, pp 43–102.

Joyner, W. B., & Boore, D. M. (1993). Methods for regression analysis of strong-motion data. *Bulletin of the Seismological Society of America, 83*(2), 469–487.

Joyner, W. B., & Boore, D. M. (1994). Errata, methods for regression analysis of strong-motion data. *Bulletin of the Seismological Society of America, 84*(3), 955–956.

Lee, Y., Anderson, J. G., & Zeng, Y. (2000). Evaluation of empirical ground-motion relations in southern California. *Bulletin of the Seismological Society of America, 90*(6B), S136–S148.

Mai, P. M., & Beroza, G. C. (2003). A hybrid method for calculating near-source, broadband seisograms: Application to strong motion prediction. *Physics of the Earth and Planetary Interiors, 137*(1–4), 183–199. https://doi.org/10.1016/S0031-9201(03)00014-1

Meyer, Y. (1986). Principe d’incertitude, bases hilbertiennes et algèbres d’operateurs, Séminaire N. Bourbaki, 662, 209–223.

Mirzaei, N., Mengtan, G., & Yuntai, C. (1998). Seismic source regionalization for seismic zoning of Iran: Major seismotectonic provinces. *J. Earthq. Pred. Res., 7*, 465–495.

Najaftomaraei, M. R., Rahimi, H., Tanircan, G., Mahani, A. B., & Shahrvar, M. (2020). Site classification and probabilistic estimation of Vs30 for the Iranian strong-motion network. *Physics of the Earth and Planetary Interiors, 308*, 106583.

Nicknam, A., Yazdani, A., & Yaghmaei Sabegh, S. (2009). Predicting probabilistic-based strong motion time series for citadel of Arg-E-Bam (South-East of Iran). *Journal of Earthquake Engineering, 13*, 482–499.

Nilforoushan, F., Masson, F., & Vernant, P. (2003). GPS network monitors the Arabia-Eurasia collision deformation in Iran. *Journal of Geodesy, 77*, 411. https://doi.org/10.1007/s00190-003-0526-5

Ölafsson, S., Remseth, S., & Sighjörnsson, R. (2001). Stochastic models for simulation of strong ground motion in Iceland. *Earthquake Engineering and Structural Dynamics, 30*(9), 1305–1313.

Pitarak, A., Irikura, K., Iwata, T., & Sekiguchi, H. (1998). Three-dimensional simulation of the near-fault ground motion for the 1995 Hyogo-ken Nanbu (Kobe), Japan, Earthquake. *Bulletin of the Seismological Society of America, 88*(2), 428–440.

Pitarak, A., Somerville, P., Fukushima, Y., Uetake, T., & Irikura, K. (2000). Simulation of near-fault strong-ground motion using hybrid Green’s functions. *Bulletin of the Seismological Society of America, 90*(3), 566–586.

Power, M., Chiou, B., Abrahamson, N., Bozorgnia, Y., Shantz, T., & Roblee, C. (2008). An overview of the NGA project. *Earthquake Spectra, 24*(1), 3–21. https://doi.org/10.1193/1.2894833

Rezaeian, S. (2010). Stochastic modeling and simulation of ground motions for performance-based earthquake engineering. University of California.

Rezaeian, S., & Der Kiureghian, A. (2012). Simulation of orthogonal horizontal ground motion components for specified earthquake and site characteristics. *Earthq. Eng. Struct. Dynam., 41*(2), 335–353.

Riahi, A., Sadeghi, H., & Hosseini, K. (2015). Simulation of 2003 Bam (Iran) earthquake using empirical Green’s function method via very small and near-fault events. *Geophysical Journal International, 201*(3), 1264–1286. https://doi.org/10.1093/gji/gsv074

Sasaki, F., Maeda, T. & Yamamoto, Y. (2003). Artificial ground motion with non-stationarity generated using the wavelet analysis. In *Proc. of the 17th Structural Mechanics in Reactor Technology (SMiRT17)*, Prague, Czech Republic, 17–22 August 2003.

Scherbaum, F., Cotton, F., & Smit, P. (2004). On the use of response spectral-reference data for the selection and ranking of ground-motion models for seismic-hazard analysis in regions of moderate seismicity: The case of rock motion. *Bulletin of the Seismological Society of America, 94*(6), 2164–2185. https://doi.org/10.1785/0120030147

Scherbaum, F., Cotton, F., & Staedtke, H. (2006). The estimation of minimum-misfit stochastic models from empirical ground-motion prediction equations. *Bulletin of the Seismological Society of America, 96*(2), 427–445. https://doi.org/10.1785/0120050015

Silva, W., Gregor, N., & Darragh, B. (1999). Near fault ground motions. Tech. rep., Pacific Engineering and Analysis, El Cerrito, USA, PG & E PEER—Task 5.A

Soghrat, M. R., Khajii, N., & Zafarani, H. (2011). Simulation of strong ground motion in northern Iran using the specific barrier model. *Geophysical Journal International, 188*(2), 645–679. https://doi.org/10.1111/j.1365-246X.2011.05287.x

Tavakoli, B., & Pezeshk, S. (2005). Empirical-stochastic ground-motion prediction for eastern North America. *Bulletin of the Seismological Society of America, 95*(6), 2283–2296. https://doi.org/10.1785/0120050030

Trifunac, M. D., & Brady, A. D. (1976). A study on the duration of strong earthquake ground motion. *Bulletin of the Seismological Society of America, 65*(3), 581–626.

Vahidifard, H., Zafarani, H., & Sabbagh-Yazdi, S. R. (2017). Hybrid broadband simulation of strong-motion records from the 16 September, 1978, Tabas, Iran, earthquake (M w 7.4). *Natural Hazards, 87*, 57–81. https://doi.org/10.1007/s11069-017-2753-2

Watson-Lamprey, J., & Abrahamson, N. (2006). Selection of empirical ground-motion relations in southern California. *Bulletin of the Seismological Society of America, 103*(3), 57–81. https://doi.org/10.1785/10.1785/0120120312

Yamamoto, Y., & Baker, J. W. (2013). Stochastic model for earthquake ground motion using wavelet packets. *Bulletin of the Seismological Society of America, 103*(6), 3044–3056. https://doi.org/10.1785/10.1785/0120120312

Zafarani, H., Rahimi, M., Noorazad, A., Hassani, B., & Khazaei, B. (2015). Stochastic simulation of strong-motion records from the 2012 Ahar-Varzaghan Dual Earthquakes, Northwest of Iran.
Zafarani, H., & Soghrat, M. (2012). Simulation of ground motion in the Zagros Region of Iran using the specific barrier model and the stochastic method. *Bulletin of the Seismological Society of America, 102*(5), 2031–2045. https://doi.org/10.1785/0120110315

Zafarani, H., Vahidifard, H., & Ansari, A. (2013). Prediction of broadband ground-motion time histories: the case of Tehran, Iran. *Earthquake Spectra, 29*(2), 633–660.

(Received January 24, 2022, revised June 19, 2022, accepted June 21, 2022, Published online July 13, 2022)