



Tenth U.S. National Conference on Earthquake Engineering
Frontiers of Earthquake Engineering
July 21-25, 2014
Anchorage, Alaska

STOCHASTIC SIMULATION OF SPATIALLY DISTRIBUTED GROUND MOTIONS USING WAVELET PACKETS AND KRIGING ANALYSIS

D. Huang¹ and G. Wang²

ABSTRACT

Considering spatial distribution of ground-motion is important in seismic hazard analysis of spatially distributed infrastructure systems such as long-span bridges, lifelines, railways. It is often the case that distributed ground-motion time histories are required to perform analyses at multiple locations where recorded time histories are not available. In this paper, we propose a geostatistics-based method to simulate spatially-distributed synthetic ground motions using wavelet packets and kriging analysis. In this model, thirteen wavelet parameters are used for time-frequency characterization of earthquake ground motions. The spatial correlation of these parameters is determined through semivariogram analysis using densely populated recordings from the Northridge and Chi-Chi earthquakes. It is observed that the spatial correlations of most wavelet parameters are closely related to regional site conditions. Kriging technique is then used to estimate the wavelet parameters at unmeasured locations using the spatial correlations. It is demonstrated that the simulated ground motions are in good agreement with the actual recordings. This method can be used for time history analysis of spatially distributed infrastructure systems and circumvent difficulties in traditional ground-motion selection and modification processes. It is also useful to derive time-histories at an unobserved location using recorded ground-motion data in the neighborhood.

¹ Graduate Student Researcher, Dept. of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong

² Assistant Professor, Dept. of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong



Tenth U.S. National Conference on Earthquake Engineering
Frontiers of Earthquake Engineering
July 21-25, 2014
Anchorage, Alaska

Stochastic Simulation of Spatially Distributed Ground Motions Using Wavelet Packets and Kriging Analysis

D.R. Huang¹ and G. Wang²

ABSTRACT

Considering spatial distribution of ground-motion is important in seismic hazard analysis of spatially distributed infrastructure systems such as long-span bridges, lifelines, railways. It is often the case that distributed ground-motion time histories are required to perform analyses at multiple locations where recorded time histories are not available. In this paper, we propose a geostatistics-based method to simulate spatially-distributed synthetic ground motions using wavelet packets and kriging analysis. In this model, thirteen wavelet parameters are used for time-frequency characterization of earthquake ground motions. The spatial correlation of these parameters is determined through semivariogram analysis using densely populated recordings from the Northridge and Chi-Chi earthquakes. It is observed that the spatial correlations of most wavelet parameters are closely related to regional site conditions. Kriging technique is then used to estimate the wavelet parameters at unmeasured locations using the spatial correlations. It is demonstrated that the simulated ground motions are in good agreement with the actual recordings. This method can be used for time history analysis of spatially distributed infrastructure systems and circumvent difficulties in traditional ground-motion selection and modification processes. It is also useful to derive time-histories at an unobserved location using recorded ground-motion data in the neighborhood.

Introduction

In current seismic design practice, earthquake ground-motion time histories are needed for nonlinear structure analysis. However, it is often the case that recorded ground motions are not available at the location being analyzed. In addition, ground motions at design levels usually have low probability of occurrence and are rare in strong-motion databases. To date, many methods have been proposed for engineering design purposes, by selecting and modifying existing ground motions, or by generating artificial time histories using stochastic approaches. Among a variety of ground-motion simulation techniques, wavelet packet transform is becoming more widely used [1-2]. The method decomposes an acceleration time history into wavelet packets localized in time and frequency domain. Reversely, synthetic motions can be simulated if the time-frequency distribution of wavelet packets can be quantified.

¹ Graduate Student Researcher, Dept. of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong.

² Assistant Professor, Dept. of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong. Corresponding Author. Email: gwang@ust.hk

Most recently, a stochastic model was proposed by Yamamoto and Baker [3] using wavelet packets. The model requires 13 wavelet parameters that can be predicted using seismological variables, such as earthquake moment magnitude, distances and site conditions, through regression analysis of strong motion data. The Yamamoto and Baker model is successful in simulating ground motions whose time-frequency characteristics (e.g. non-stationarity) are consistent with the actual recorded motions. Another important feature is the model introduces variability in the simulated ground motions, as the predictive equations for the model parameters provide not only the median, but also the variability of these parameters.

However, the Yamamoto and Baker model can only be applied to simulate ground-motion time histories at individual site locations. Yet, considering spatial distribution of ground motions is important in seismic hazard analysis of spatially-distributed infrastructure, such as long-span bridges, lifelines and railways. It is desirable to simulate simultaneous occurrence of ground-motion time histories at multiple locations when performing time-history analysis of the spatially-distributed systems. It is also very useful to derive time-histories at an unobserved location using recorded ground-motion data in the neighborhood. For all these purposes, the spatial correlation of wavelet parameters should be further studied.

This study uses well-recorded and densely-populated ground-motion data from the 1994 Northridge and 1999 Chi-Chi earthquakes to develop empirical spatial correlations for thirteen wavelet parameters used in the Yamamoto and Baker model. The estimated spatial correlations can be used to simulate spatially-distributed ground motions based on earthquake scenarios, or they can be used to interpolate wavelet parameters, and thus the time histories, at unobserved locations using the kriging technique and recorded data in the neighborhood. This paper will demonstrate that the simulated ground motions considering spatial correlations are consistent with the actual recorded data in a specific region. These synthetic ground motions can be well used in time history analysis and loss assessments of spatially-distributed infrastructure.

Wavelet Characterization of Nonstationary Ground Motions

This study adopts the stochastic model proposed by Yamamoto and Baker [3] to decompose ground-motion time histories using the wavelet packet transform. The decomposed wavelet packets are further separated into a major group containing 70% of total energy and a minor group containing the remaining 30% energy. A suite of thirteen wavelet parameters are defined to account for statistical attributes of the amplitude coefficients of wavelet packets in time-frequency domain. Of the thirteen parameters, the sum of all wavelet coefficients represents the total energy contained in the ground motion, defined as E_{acc} [3],

$$E_{acc} = \sum_i \sum_k |c_{j,k}^i|^2 = \int_{-\infty}^{\infty} |x(t)|^2 dt \quad (1)$$

For coefficients in the major group, six wavelet parameters are defined by Eqs. 2-7,

$$E(a)_{major} = \sum_i \sum_k |c_{j,k,major}^i|^2 / n_{major} \quad (2)$$

$$E(t)_{major} = \sum_i \sum_k t_{i,k,major} / n_{major} \quad (3)$$

$$S^2(t)_{major} = \sum_i \sum_k \{t_{i,k,major} - E(t)_{major}\}^2 / (n_{major} - 1) \quad (4)$$

$$E(f)_{major} = \sum_i \sum_k f_{i,k,major} / n_{major} \quad (5)$$

$$S^2(f)_{major} = \sum_i \sum_k \{f_{i,k,major} - E(f)_{major}\}^2 / (n_{major} - 1) \quad (6)$$

$$\rho(t, f)_{major} = \frac{\sum_i \sum_k [t_{i,k,major} - E(t)_{major}] [f_{i,k,major} - E(f)_{major}]}{S(t)_{major} S(f)_{major} (n_{major} - 1)} \quad (7)$$

where $c_{j,k,major}^i$ denotes wavelet coefficients in the major group, n_{major} denotes the number of coefficients in the major group, and $t_{i,k,major}$ and $f_{i,k,major}$ represent centers of wavelet packet coefficients in time and frequency domain, respectively. Similarly, another six wavelet parameters are defined for characterizing statistical properties in minor group,

$$E(t)_{minor} = \sum_i \sum_k t_{i,k,minor} |c_{j,k,minor}^i|^2 / (0.3 \cdot E_{acc}) \quad (8)$$

$$S^2(t)_{minor} = \sum_i \sum_k \{t_{i,k,minor} - E(t)_{minor}\}^2 |c_{j,k,minor}^i|^2 / (0.3 \cdot E_{acc}) \quad (9)$$

$$E(f)_{minor} = \sum_i \sum_k f_{i,k,minor} |c_{j,k,minor}^i|^2 / (0.3 \cdot E_{acc}) \quad (10)$$

$$S^2(f)_{minor} = \sum_i \sum_k \{f_{i,k,minor} - E(f)_{minor}\}^2 |c_{j,k,minor}^i|^2 / (0.3 \cdot E_{acc}) \quad (11)$$

$$\rho(t, f)_{minor} = \frac{\sum_i \sum_k [t_{i,k,minor} - E(t)_{minor}] [f_{i,k,minor} - E(f)_{minor}]}{0.3 \cdot E_{acc} \cdot S(t)_{minor} S(f)_{minor}} \quad (12)$$

where $t_{i,k,minor}$ and $f_{i,k,minor}$ represent centers of minor-group wavelet packet coefficients in time and frequency domain, respectively. Additionally, the randomness of wavelet coefficients in the minor group is denoted as $S(\xi)$

Spatial Correlation of Wavelet Parameters

The spatial distributions of wavelet parameters are investigated using semivariogram. The semivariogram is a widely used geostatistical tool for modeling regionalized variables, such as spatially-distributed ground-motion intensity measures (e.g. [4-6]). It characterizes the dissimilarity or decorrelation of spatial data, which can be thought of as a stationary regionalized variable $\{Z(\mathbf{u}) : \mathbf{u} \in D\}$, in which the spatial index \mathbf{u} varies continuously over the region D . For a data pair separated by a vector \mathbf{h} , the semivariogram is defined as [7],

$$\gamma(\mathbf{h}) = \frac{1}{2} \text{E} \left[(Z(\mathbf{u} + \mathbf{h}) - Z(\mathbf{u}))^2 \right] \quad (13)$$

Previous studies suggested that the vector lag distance \mathbf{h} in Eq. 13 can be replaced by a scalar variable $h = \|\mathbf{h}\|$ based on the assumption that the variable Z is spatially isotropic and second-order stationary. However, it is not always the case that two sites are separated by an exact lag distance h . Therefore, this study employs a separation distance bins $[h - \Delta h, h + \Delta h]$ with a bin size of Δh to group all data pairs when computing semivariograms.

Recent advances in geostatistical studies of earthquake ground motions have led to various estimators being applied to estimate semivariograms, such as the method-of-moments estimator and the robust estimator [7]. It was demonstrated that both estimators yield similar semivariograms [5]. In order to provide consistent results, the method-of-moments estimator is used for computing semivariograms throughout the study. Its formulation is provided in Eq. 14,

$$\tilde{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{i=1}^{N(h)} [z(u_i + h) - z(u_i)]^2 \quad (14)$$

where $N(h)$ represents the number of distinct data pairs in the separation-distance bin $[h - \Delta h, h + \Delta h]$. Four basic continuous models are commonly used to fit empirical semivariograms, namely, the exponential model, the spherical model, the Gaussian model and the nugget effect model. Among all these models, the exponential model is found to have the best fitting capability, and it will be adopted in this study.

$$\tilde{\gamma}(h) = a[1 - \exp(-3h/b)] \quad (15)$$

where a and b are defined as sill and range of the semivariogram, respectively. The exponential model specifies that 95% of the spatial correlation vanishes beyond the range b .

Spatial Correlations of Wavelet Parameters

In this section, spatial correlations of thirteen wavelet parameters are developed using ground-motion data from two well-recorded past earthquakes, the 1994 Northridge earthquake and 1999 Chi-Chi earthquakes. These two earthquakes are selected to represent different regional geological conditions. The Northridge earthquake represents a heterogeneous region, and the Chi-Chi earthquake represents a homogeneous region, based on their estimated ranges of V_{s30} . Previous research works reported that the range of V_{s30} for the Northridge earthquake is 0 km, indicating an independent V_{s30} distribution over the region. On the other hand, the range of V_{s30} for the Chi-Chi earthquake was estimated as 27 km representing a homogeneous geological condition [5].

1994 Northridge earthquake and 1999 Chi-Chi earthquake

The earthquake data used in developing empirical spatial correlations are selected from a subset

of NGA database used by Boore and Atkinson [8]. Ground motions with lowest useable frequency higher than 1Hz are excluded, and only fault-normal components are adopted in the analyses. These criteria result in 148 ground-motion recordings available for the Northridge earthquake and 381 recordings for the Chi-Chi earthquake. Residuals of the wavelet parameters are computed using the prediction model proposed by Yamamoto and Baker [3]. To construct semivariograms, the bin size is set as 4 km, such that the lack of data within short separation distances can be compensated.

Fig. 1 presents semivariograms of E_{acc} residuals for the 1994 Northridge and the 1999 Chi-Chi earthquake using the weighted-least-square (WLS) method that fits an exponential model. Ranges of semivariograms for E_{acc} are estimated as 9.7 km and 41.6 km for the Northridge and the Chi-Chi earthquake, respectively. These ranges are consistent with those obtained from Arias Intensity (I_a) residuals in the same region [5], since E_{acc} and I_a all represent the integration of acceleration time histories (they only differ by a constant multiplication factor). Fig. 2 shows semivariograms for remaining wavelet parameters, whose range values varied from 7.7 km to 19.7 km for the Northridge earthquake, and from 13.2 km to 58.8 km for the Chi-Chi earthquake.

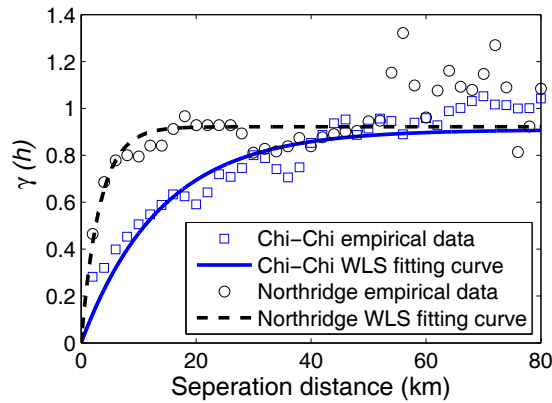


Figure 1. Semivariograms of E_{acc} residuals for 1994 Northridge and 1999 Chi-Chi earthquake using weighted-least-square (WLS) method that fits the exponential model

It is observed that the spatial correlations of most wavelet parameters are closely related to regional site conditions. The correlation ranges for the Northridge earthquake are generally smaller than those obtained from the Chi-Chi earthquake. Therefore, it is important to preserve the regional-specific spatially correlations of these wavelet parameters when they are utilized to generate the spatially-distributed synthetic motions.

The Influence of Regional Site Conditions on the Spatial Correlation

The influence of regional site conditions can be further summarized and compared using the ratios of the correlation ranges from the Chi-Chi and Northridge earthquake, as shown in Table 1, where a larger ratio means that the spatial correlation is more significantly affected by the regional site conditions. Accordingly, all wavelet parameters are divided into four groups.

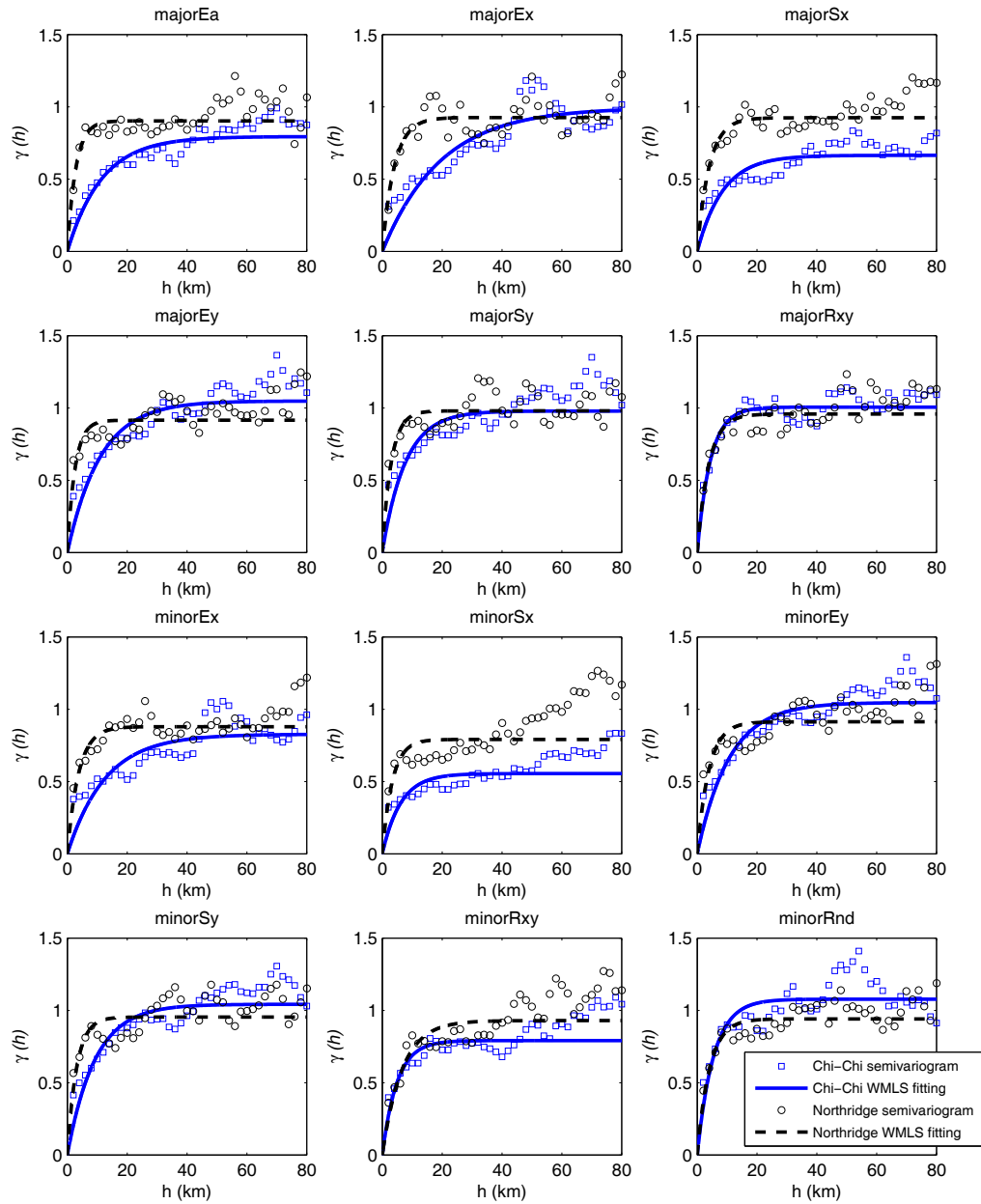


Figure 2. Semivariograms of the twelve wavelet parameter residuals for 1994 Northridge and 1999 Chi-Chi earthquake using weighted-least-square (WLS) method that fits the exponential model

Group I includes two wavelet parameters that define the centroids of the major-group coefficients in the time-frequency domain ($E(t)_{major}$ and $E(f)_{major}$), as well as two wavelet parameters for the total and major-group energy (E_{acc} and $E(a)_{major}$). They are found to be most strongly affected by regional geological conditions.

In Group II and III, the influence of regional site conditions becomes slightly less pronounced. It may be attributed to the intrinsic meaning of those parameters. Group II represents the centroid of the minor-group-coefficient distribution; Group III is time and frequency variance of both major and minor group coefficients. The ratios of ranges are approximately 3 and 2 for Group II and Group III, respectively. On the other hand, the spatial correlations of wavelet parameters in Group IV are not significantly affected by regional conditions. Group IV describes the time-frequency correlation (i.e. $\rho(t, f)_{major}$ and $\rho(t, f)_{minor}$) and randomness in the minor group (i.e. $S(\xi)$).

Table 1. Estimated correlation ranges of wavelet parameters and the ratios for the Chi-Chi and Northridge earthquakes.

Groups	Wavelet parameters	Northridge	Chi-Chi	Ratio of ranges	Averaged ratio
I	E_{acc}	9.7	41.6	4.29	4.39
	$E(a)_{major}$	8.4	34.5	4.11	
	$E(t)_{major}$	12.1	58.8	4.86	
	$E(f)_{major}$	7.7	33.1	4.30	
II	$E(t)_{minor}$	11.3	38.1	3.37	3.07
	$E(f)_{minor}$	11.6	32.0	2.76	
III	$S(t)_{major}$	11.9	25.7	2.20	2.39
	$S(f)_{major}$	8.9	22.9	2.57	
	$S(t)_{minor}$	10.0	18.4	1.84	
	$S(f)_{minor}$	9.3	27.7	2.98	
IV	$\rho(t, f)_{major}$	11.5	13.2	1.15	1.08
	$\rho(t, f)_{minor}$	19.7	14.8	0.75	
	$S(\xi)$	12.3	16.4	1.33	

Kriging Estimate of Wavelet Parameters at Unmeasured Locations

Using spatial correlations developed in previous sections, wavelet parameters at unmeasured locations in the study region can be predicted. A spatial interpolation technique named ordinary kriging is employed in this study. The method provides least-square linear estimates of variables at the unsampled point, by minimizing the estimation variance using a predefined semivariogram model [9]. For a regionalized random field $Z(\mathbf{u})$, the ordinary kriging estimator $Z^*(\mathbf{u})$ is defined as [9]

$$\begin{aligned}
Z^*(\mathbf{u}) &= m(\mathbf{u}) + \sum_{i=1}^{N(\mathbf{u})} \lambda_i(\mathbf{u}) [Z(\mathbf{u}_i) - m(\mathbf{u})] \\
&= \sum_{i=1}^{N(\mathbf{u})} \lambda_i(\mathbf{u}) \cdot Z(\mathbf{u}_i) + \left[1 - \sum_{i=1}^{N(\mathbf{u})} \lambda_i(\mathbf{u}) \right] \cdot m(\mathbf{u}) \\
&= \sum_{i=1}^{N(\mathbf{u})} \lambda_i(\mathbf{u}) \cdot Z(\mathbf{u}_i)
\end{aligned} \tag{16}$$

where N is the number of measured data, $m(\mathbf{u})$ is the expected value of $Z(\mathbf{u})$, $\lambda_i(\mathbf{u})$ is the kriging weight assigned to $Z(\mathbf{u}_i)$, which will be determined from the emivariogram model. The term of $m(\mathbf{u})$ can be removed by reinforcing the summation of kriging weights to be 1. Subject to the statistical constraint Eq. 16, the Lagrange multiplier technique is applied to minimize the kriging variance. The Lagrangian $L(\mathbf{u})$ is a function of a newly introduced Lagrange multiplier μ and the kriging weight vector $\lambda(\mathbf{u})$ as follows [9],

$$L(\lambda_1(\mathbf{u}), \dots, \lambda_i(\mathbf{u}), \mu(\mathbf{u})) = \sigma_E^2(\mathbf{u}) + 2\mu(\mathbf{u}) \left[\sum_{i=1}^{N(\mathbf{u})} \lambda_i(\mathbf{u}) - 1 \right] \tag{17}$$

The optimal estimate of kriging weight vector $\lambda(\mathbf{u})$ can be obtained by setting zero to partial derivatives of L , resulting in an ordinary kriging system expressed in terms of semivariograms as follows,

$$\begin{cases} \sum_{i=1}^{N(\mathbf{u})} \lambda_i(\mathbf{u}) \gamma(\mathbf{u}_i - \mathbf{u}_j) - \mu(\mathbf{u}) = \gamma(\mathbf{u}_j - \mathbf{u}), & j = 1, \dots, N(\mathbf{u}) \\ \sum_{i=1}^{N(\mathbf{u})} \lambda_i(\mathbf{u}) = 1 \end{cases} \tag{18}$$

where $\lambda(\mathbf{u})$ and μ can be solved using the above $N+1$ equations.

Example for generating synthetic motions using Northridge earthquake data

This section provides an example for generating synthetic ground motions using spatial correlations developed for the Northridge earthquake. For validation purpose, the ground-motion recording at Los Angeles Baldwin Hills (i.e. NGA sequence No. 985, termed as LA-BH site) was removed from the database throughout the analysis, so that a synthetic ground motion can be produced at this site. The simulated motion will be compared with the real recording. Fig. 3 shows a map illustrating locations of the epicenter of 1994 Northridge earthquake (red star), 148 recording stations (blue circles) and the LA - Baldwin Hills site (yellow triangle).

The thirteen wavelet parameters at the “unobserved” LA-BH site were computed using the ordinary kriging technique based on empirical spatial correlations. Table 2 summarized the wavelet parameters estimated in this study, which are in a good agreement with those directly

obtained from the actual recording using wavelet packet analyses. Further, simulated wavelet packet coefficients in time-frequency domain are presented in Fig. 4, whose distribution is found to be generally similar to that of the actual recording, except that the latter one exhibits a more irregular distribution. The simulated ground motion time history and its response spectrum are presented in Fig. 5, which are compared with those obtained from actual data. This demonstrated that the spatially correlated parameterization developed in this study is a good representation of non-stationary ground motions.

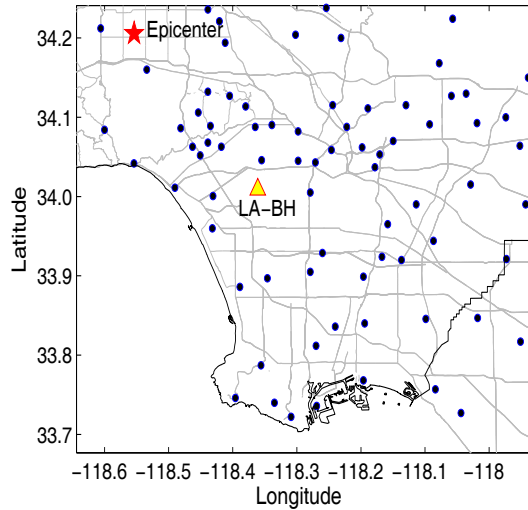


Figure 3. A map showing the epicenter of 1994 Northridge earthquake (red star), 148 recording stations (blue circles) and Los Angeles - Baldwin Hills site (yellow triangle).

Table 2. Summary of simulated and recorded wavelet parameters for generating synthetic ground motions at the LA - Baldwin Hills site for the 1994 Northridge earthquake

Wavelet parameters	E_{acc}	$E(a)_{maj}$	$E(t)_{maj}$	$S(t)_{maj}$	$E(f)_{maj}$	$S(f)_{maj}$	$\rho(t, f)_{maj}$
Simulated	4.583	0.026	12.471	6.039	3.684	2.532	-0.357
Recorded	3.775	0.019	11.971	6.570	3.791	2.758	-0.479
Wavelet parameters	$E(t)_{min}$	$S(t)_{min}$	$E(f)_{min}$	$S(f)_{min}$	$\rho(t, f)_{min}$	$S(\xi)$	
Simulated	12.263	8.387	4.864	5.474	-0.165	1.191	
Recorded	12.905	9.492	5.217	6.738	-0.232	1.220	

Conclusions

Spatial correlations of decomposed wavelet packet parameter are developed empirically based on ground-motion records from 1994 Northridge and 1999 Chi-Chi earthquake. It is observed that most of these parameters are strongly influenced by local geological conditions. Kriging

technique is adopted to populate wavelet parameters at unmeasured locations using spatial correlations. The simulated ground motions agree well with the actual recorded data. The spatially correlated wavelet parameterization proposed in this study is a good representation of the spatially distribution of ground motions. Using this method, regional distributed synthetic ground motions can be generated. The method can be applied in time history analyses of distributed infrastructure systems.

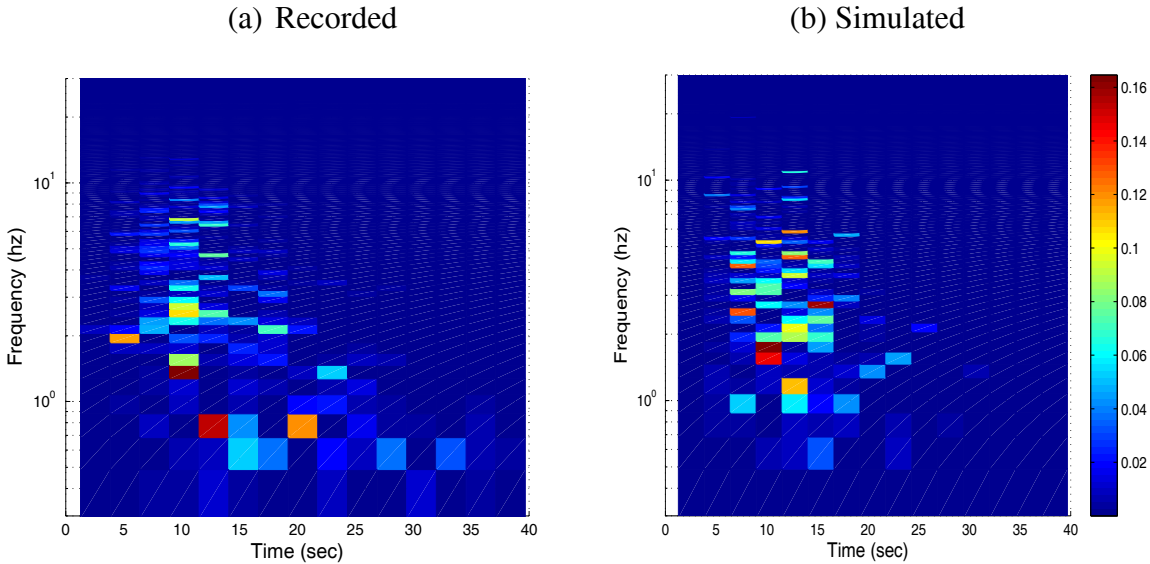


Figure 4. Wavelet packet coefficients in time-frequency domain for ground motions (a) recorded at the LA-BH site during 1999 Northridge earthquake, and (b) simulated using kriged wavelet parameters based on spatial correlations

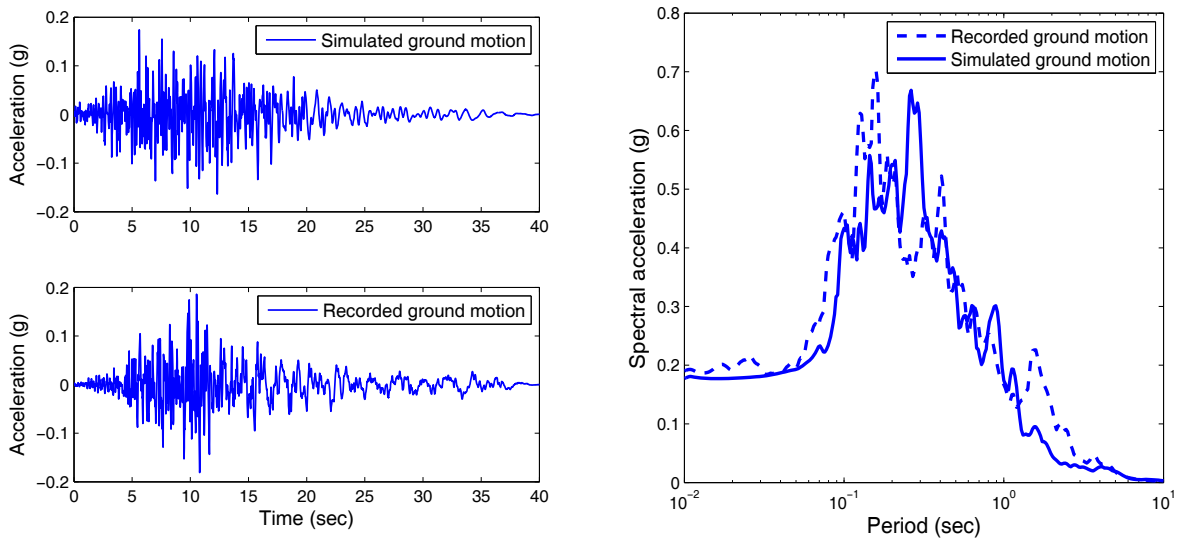


Figure 5. Simulated and recorded ground motions at the LA-BH site during 1999 Northridge earthquake (left), and a comparison of their response spectra.

Acknowledgements

The study was financially supported by Hong Kong RGC/HKUST grant DAG12EG07-3 and FSGRF13EG09.

References

1. Amiri GG, Shahjouei A, Saadat S, Ajallooeian M. Hybrid evolutionary-neural network approach in generation of artificial accelerograms using principal component analysis and wavelet-packet transform. *Journal of Earthquake Engineering* 2011; **15** (1): 50-76.
2. Percival DB, Walden AT. *Wavelet Methods for Time Series Analysis*. Cambridge University Press: Cambridge, 2000.
3. Yamamoto Y, Baker JW. Stochastic model for earthquake ground motion using wavelet packets. *Bulletin of the Seismological Society of America* (in press).
4. Jayaram N, Baker JW. Correlation model for spatially distributed ground-motion intensities. *Earthquake Engineering and Structural Dynamics* 2009; **38**: 1687-1708.
5. Du W, Wang G. Intra-Event spatial correlations for cumulative absolute velocity, Arias Intensity, and Spectral Accelerations based on regional site conditions. *Bulletin of the Seismological Society of America* 2013; **103** (2A): 1117-1129.
6. Wang G, Du W. Spatial cross-correlation models for vector intensity measures (PGA, Ia, PGV and SAs) considering regional site conditions. *Bulletin of the Seismological Society of America* 2013; **103** (6):DOI: 10.1785/0120130061.
7. Goovaerts P. *Geostatistics for Natural Resources Evaluation*. Oxford University Press: Oxford, New York, 1997.
8. Boore DM, Atkinson GM. Ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods between 0.01 s and 10.0 s. *Earthquake Spectra* 2008; **24**(1): 99-138.
9. Hohn ME. *Geostatistics and petroleum geology*. Kluwer Academic Publisher, 1999.