Research Article

Dynamic Time Warping Distance Method for Similarity Test of Multipoint Ground Motion Field

Yingmin Li,[1] Huiguo Chen,[1, 2] and Zheqian Wu[1]

[1] *College of Civil Engineering, Chongqing University, Chongqing 400045, China*
[2] *Department of Architecture & Civil Engineering, Logistics Engineering College, Chongqing 401311, China*

Correspondence should be addressed to Huiguo Chen, chenhg_mail@163.com

The reasonability of artificial multi-point ground motions and the identification of abnormal records in seismic array observations, are two important issues in application and analysis of multipoint ground motion fields. Based on the dynamic time warping (DTW) distance method, this paper discusses the application of similarity measurement in the similarity analysis of simulated multi-point ground motions and the actual seismic array records. Analysis results show that the DTW distance method not only can quantitatively reflect the similarity of simulated ground motion field, but also offers advantages in clustering analysis and singularity recognition of actual multi-point ground motion field.

1. Introduction

Considerable research studies have shown that nonuniform ground motion input has a significant influence on dynamic response of large scale structures [1, 2]. The influence of multi-point ground motion must be taken into account when analyzing the seismic response of large scale structures. Ground motion records from common stations only reflect one- or multidimension vibration of sites, while multi-point ground motion fields are often obtained from seismic arrays. For lack of actual records of seismic arrays, attenuation rules of actual multi-point ground motion field are usually used to synthesize multi-point ground motions for specific site conditions and distances. In this process, the traveling wave effect and coherence of multi-point ground motions recorded in seismic arrays with various distances are first analyzed, and then the coherence corresponding to various points are obtained through statistical analysis. Finally, the method for synthesizing spatial ground motions with the coherence effect and traveling effect of multi-point earthquake ground motions considered can be established [3–5]. In this method, the variations of ground motions with
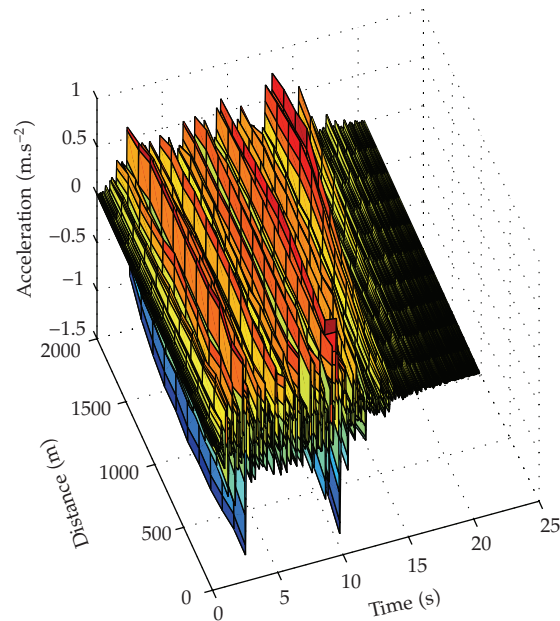
Figure 1: Multi-point ground motion field simulated using actual initial phase.

site conditions and motion propagation can be controlled in a certain range in the extent of engineering scales for both actual array records and artificially synthesized multi-point earthquake ground motions. Multi-point ground motion field in the engineering scale may have a certain degree of similarity. Moreover, if the site condition is homogeneous, the degree of similarity will gradually decrease with the increase in distance [6].

Based on the knowledge that multi-point ground motion fields have similarity in engineering scale, there are at least two issues that can be addressed by determining the similarity of ground motions. First, the reliability of actual array records can be ensured. Since all actual array records are obtained from ground motion recording instruments, the reliability of records from all stations cannot be guaranteed in a particular triggering earthquake. Due to equipment failure or equipment differences some station records may be seriously distorted. When to use seismic array records for analyzing the propagation of multi-point ground motions, the distorted data should be removed to ensure the reasonability of array records and such distorted data can be excluded by using similarity test of actual multi-point ground motions. Second, the reasonability of artificially synthesized multi-point ground motions can also be tested. By judging whether the artificial multi-point ground motion field is similar to actual ground motion field, the efficiency and rationality of the synthesis method for multi-point ground motions can be tested.

For this reason, it is necessary to establish a standardized and efficient similarity test method for multi-point ground motions. For similarity judgment of ground motions, correlation coefficients are frequently used. However, correlation coefficients are mainly defined as a measurement of linear feature of two variables, and they are not suitable for similarity judgment between two sequences with nonlinear correlations. Spatial multi-point earthquake ground motions are often influenced by site conditions, coherence effect and traveling effect, therefore the use of the correlation coefficient is insufficient to reflect the similarity of ground motions. Figures 1 and 2 show two sets of multi-point ground motions
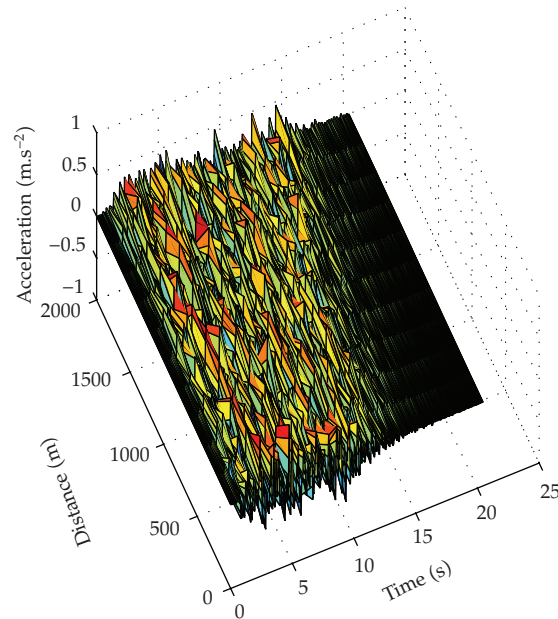
Figure 2: Multi-point ground motion field simulated using random initial phase.

synthesized using the trigonometric series method [7] with two different initial phase values. For condition 1, the initial phase is the actual ground motion phase, and for condition 2, the initial input phase is random. In both cases, the distance interval of synthesized ground motions is 200 m. Figure 3 shows the correlation coefficients between the artificial ground motion and the actual ground motions. For condition 2, due to the impact of the random initial phase value, the nonlinear relationships among ground motions of various points are more prominent, which indicates less correlations among various points. It is clear that in condition 2 the similarity among multi-point ground motions is not correctly presented.

Similarity measurement of time series is an important problem in data mining. By using the basic theory of similarity measurement, it not only can reflect the intrinsic similarity of time series or eigenvectors but also can quantitatively evaluate their similarity features. Similarity measurement has been an important tool in series clustering, pattern matching, classification, rule identification and anomaly detection [8, 9]. Among the methods commonly used for similarity measurement, the dynamic time warping distance (DTW) method proposed by Berndt and Clifford [10] is the most widely used, which can effectively determine the similarity between two series under the circumstances of shift and stretching of amplitude, and the bending of time axis [11]. In this study, the DTW method is applied to the similarity determination and singularity recognition of multi-point ground motions.

2. The Basic Principle of Similarity Measurement Method

In general, the similarity measurement between series means that a function $\text{Sim}(X, Y)$ can be defined where X and Y are two different time series of the same type of data set. The value of the function Sim lies in $(0, 1]$, that is, the bigger the value is, the higher the similarity degree between two series is, and vice versa. In particular, two series are completely similar
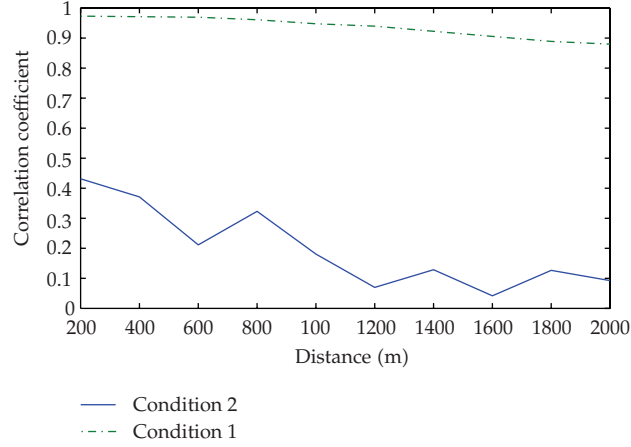
Figure 3: Correlation coefficient of multi-point ground motion field simulated under different conditions.

when Sim = 1. For simple time series, the similarity measurement function can be represented using correlation coefficients or cosine value between two series, while for the complex data, it is difficult to accurately represent the similarity degree using above function. Therefore, the similarity degree between the series is generally represented by defining a specific distance between two series called similarity distance. There are several methods available to define the similarity distance and the Minkowski distance is most frequently used, which is defined as follows:

$$d(X, Y) = \left(\sum_{i=1}^{n} |x_i - y_i|^P \right)^{1/P}. \tag{2.1}$$

When $P = 2$, the distance between two series is called Euclidean Distance.

Referring to formula (2.1), if two series are equal, their distance is zero, which means two series are completely similar. The larger difference between two series yields larger distance and less similarity. When to calculate the Minkowski distance, it is required that two series have the same length, the values of two series have point-to-point correspondence, and the weight of each pair of difference is equal. Due to such a correspondence, the Minkowski distance cannot be applied to the similarity measurement of complex series with shift and stretching of amplitude. To solve this problem, the dynamic time warping distance method is often used. In this method, the distance is designated to depict the greatest similarity between series by calculating the minimum distance between them, which is defined as follows.

Let $X(x_1, x_2, \ldots, x_n)$ and $Y(y_1, y_2, \ldots, y_m)$ be two series with the length of n and m, respectively, and an $n \times m$ matrix M can be defined to represent the point-to-point correspondence relationship between X and Y, where the element M_{ij} indicates the distance $d(x_i, y_j)$ between x_i and y_j. Then the point-to-point alignment and matching relationship between X and Y can be represented by a time warping path $W = \langle w_1, w_2, \ldots, w_K \rangle, \max(m, n) \leq K < m + n - 1$, where the element $w_k = (i, j)$ indicates

the alignment and matching relationship between x_i and y_j. If a path is the lowest cost path between two series, the corresponding dynamic time warping distance is required to meet

$$\text{DTW}(X, Y) = \min_{W}\left\{\sum_{k=1}^{K} d_k, W = \langle w_1, w_2, \ldots, w_K \rangle\right\}, \tag{2.2}$$

where $d_k = d(x_i, y_j)$ indicates the distance represented as $w_k = (i, j)$ on the path W.

Then the formal definition of dynamic time warping distance between two series is described as

$$\text{DTW}(\langle\ \rangle, \langle\ \rangle) = 0,$$

$$\text{DTW}(X, \langle\ \rangle) = \text{DTW}(\langle\ \rangle, Y) = \infty,$$

$$\text{DTW}(X, Y) = d(x_i, y_j) + \min\begin{cases} \text{DTW}(X, Y[2:-]), \\ \text{DTW}(X[2:-], Y), \\ \text{DTW}(X[2:-], Y[2:-]), \end{cases} \tag{2.3}$$

where $\langle\ \rangle$ indicates empty series, $[2:-]$ indicates a subarray whose elements include the second element to the final element in an one-dimension array, $d(x_i, y_j)$ indicates the distance between points x_i and y_j which can be represented by the different distance measurements, for example, Euclidean Distance. The DTW distance of two-time series can be calculated by the dynamic programming method based on accumulated distance matrix [10], whose algorithm mainly is to construct an accumulated distance matrix:

$$r(i, j) = d(x_i, y_j) + \min\{r(i-1, j), r(i, j-1), r(i-1, j-1)\}. \tag{2.4}$$

Any element $r(i, j)$ in the accumulated matrix indicates the dynamic time warping distance between series $X_{1:i}$ and $Y_{1:j}$. Series with high similar complexity can be effectively identified because the best alignment and matching relationship between two series is defined by the dynamic time distance.

3. Similarity Analysis of the Simulated Multi-Point Ground Motions

For structural dynamic analysis, artificial multi-point ground motion field plays a significant role. Current methods of artificial ground motion mainly cover triangular series method and Green function method [3]. According to the general rule of multi-point ground motion field in engineering scale, the multi-point ground motion fields simulated using above methods are of high similarity. Due to site effect, wave propagation and other factors, multi-point ground motions often show relatively more complex nonlinear relationships such as time shift and stretching. The correlation coefficient method shown in Figure 3 cannot accurately represent the similarity of simulated ground motions. For ground motions at the points with different distances, the corresponding dynamic warping time distances between them and the input ground motion can be calculated via the method
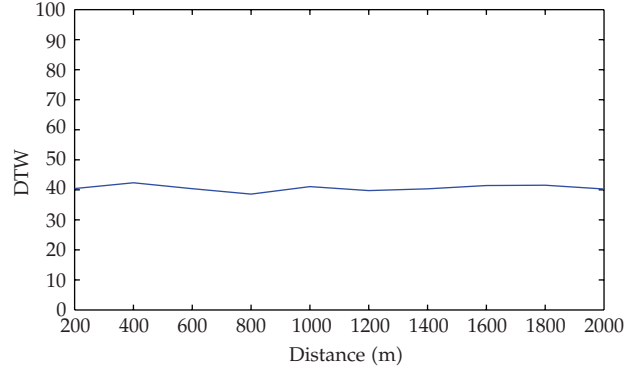
Figure 4: The dynamic warping time distance between the points of the synthesized ground motion in Figure 2 and the input ground motion.

discussed above. Figure 4 shows the dynamic time warping distance between the points and the input ground motion under condition 2 (Figure 2). It can be seen that the dynamic time warping distances between the points and input ground motion is basically around 40, indicating that ground motions of the points have an identical degree of similarity.

However, because different multi-point ground motion fields have different dynamic time warping distances, the degree of similarity between ground motions of various points cannot be intuitively presented. If a maximum distance (or lowest similarity) can be approximately defined between the input ground motion and a assume time series, the similarity degree of ground motions can be calculated numerically. That is to say, if regard the dynamic time warping distance between a particular time series and the input ground motion as a reference standard of maximum distance, the similarity degree of the particular time series and input ground motion can be consider as zero. In this paper, according to the basic understanding that the values of ground motion are not always maintained at the peak value as time increasing, a particular time series can be approximately assumed whose values are constantly the peak value of input ground motion as time increasing. If the dynamic time warping distance between the particular series and the input ground motion is define as D (the maximum distance), then according to the dynamic time warping distance $d(i)$ between the points and input ground motions, the approximate degree of similarity between the points and input ground motions can be calculated as follows:

$$\text{Sim}(a_i, a_1) = 1 - \frac{d(i)}{D}. \tag{3.1}$$

Figure 5 shows the degree of similarity between the points of simulated multi-point ground motion filed in Figure 2 and the input ground motion. By comparing with the results of similarity degree of multi-point ground motion using correlation coefficient as shown in Figure 3, it can be found the degree of similarity via dynamic warping time distance is more reasonable.
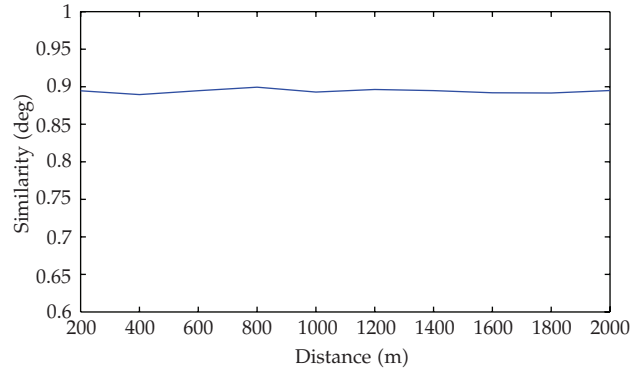
Figure 5: The similar degrees of similarity between the points of the synthesized ground motion filed in Figure 2 and the input ground motion.
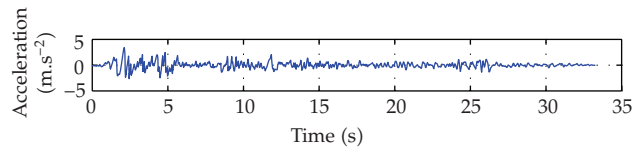

Figure 6: Time history curve of El Centro NS.

4. Clustering Analysis and Singularity Recognition of Multi-Point Ground Motions

More generally, for ground motions in an unknown multi-point ground motion field, if a certain ground motion is seriously distorted, it can also be effectively identified using the dynamic time warping distance. The DTW method provides a basis for clustering analysis and singularity recognition of multi-point ground motion data. To validate the efficiency of this method, the real ground motion (El Centro 1940) recorded at Imperial Valley, California, USA is used whose time history of acceleration is shown in Figure 6.

Firstly, an experiment is performed using the artificially simulated ground motion field with multi-point similarity. Figures 7 and 8 show the time-history curves of artificial ground motions at some points as shown in Figures 1 and 2. By comparing the propagation rule and waveform of El Centro NS and the artificial motions, we can see there are obvious differences. If substitute El Centro wave for the original input seismic wave, a multi-point ground motion field including artificial singular ground motion can be established. Obviously, in this artificially synthesized ground motion field, El Centro wave should be the singular seismic wave. Using the dynamic warping distance method, the similarities of multi-point ground motion field in Figures 1 and 2 are calculated, respectively, and the results are shown in Tables 1 and 2. It can be seen in both Tables the dynamic warping distances between El Centro wave and all other points (indicated in the table as bold characters) show consistent anomalies. It can be concluded that El Centro is the singular seismic wave while the rest are in the same type of ground motion field.
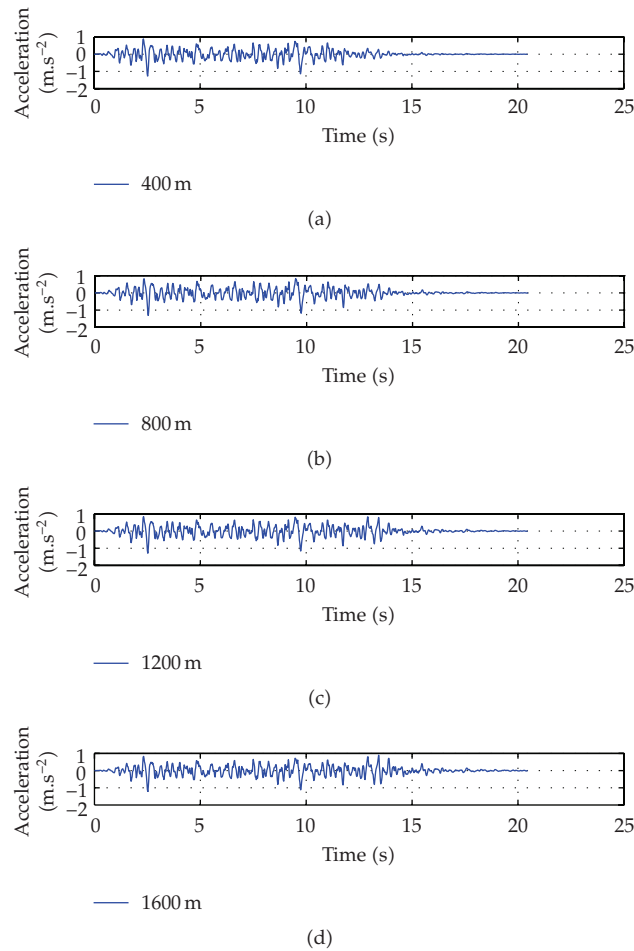
Figure 7: Time history curve of artificial ground motion at some points as shown in Figure 1.

Secondly, the data recorded at the SMART1 Array located in Lotung, Taiwan during Event 40 are used (Figure 9). Similarly, the El Centro wave was introduced in the array records to calculate the dynamic time warping distance between various points. The values are shown in Table 3. It is seen that the singular earthquake ground motion can be accurately determined via dynamic time warping distance calculation.

5. Conclusion

There exists similarity among the ground motions of multi-point ground motion filed and the degree of similarity can be accurately evaluated using the dynamic time warping distance method.

The dynamic time warping distance method is an efficient method for singularity recognition of actual array data, and it can be used in the preprocessing and clustering analysis of actual array data of multi-point ground motion field.
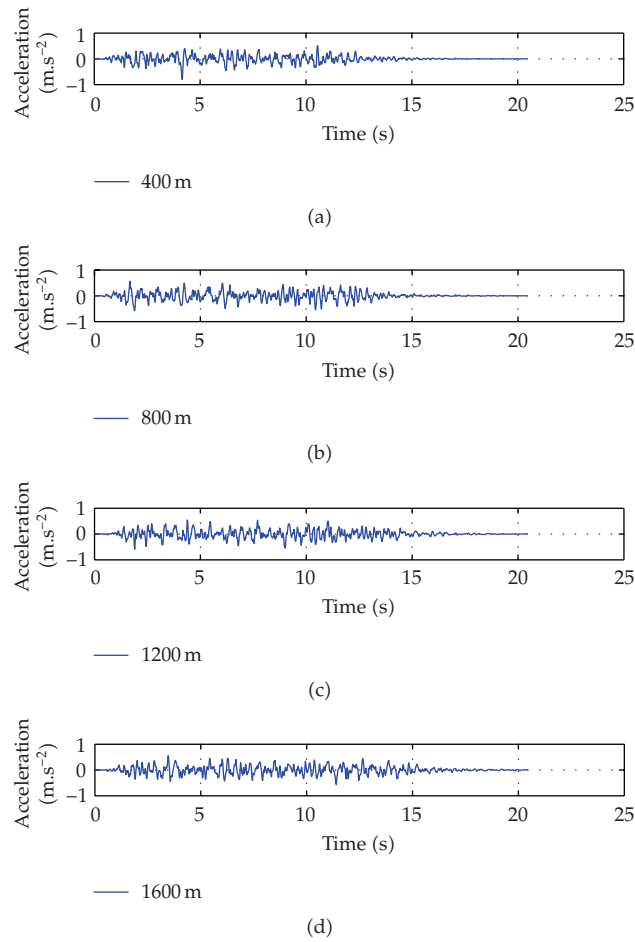
Figure 8: Time history curve of artificial ground motion at some points as shown in Figure 2.

Table 1: The dynamic time warping distance between each point after mixing singular ground motion into the artificial multi-point ground motion field in Figure 1.

DTW distance	El Centro	200 m	400 m	600 m	800 m	1000 m	1200 m	1400 m	1600 m	1800 m	2000 m
El Centro	0	**104**	**106**	**109**	**111**	**114**	**114**	**115**	**115**	**114**	**108**
200 m	**104**	0	28	33	37	45	49	54	56	52	67
400 m	**106**	28	0	21	28	35	38	42	43	57	65
600 m	**109**	33	21	0	19	24	29	31	45	55	64
800 m	**111**	37	28	19	0	18	18	31	40	50	60
1000 m	**114**	45	35	24	18	0	19	25	35	45	58
1200 m	**114**	49	38	29	18	19	0	17	27	38	53
1400 m	**115**	54	42	31	31	25	17	0	19	31	48
1600 m	**115**	56	43	45	40	35	27	19	0	23	41
1800 m	**114**	52	57	55	50	45	38	31	23	0	33
2000 m	**108**	67	65	64	60	58	53	48	41	33	0
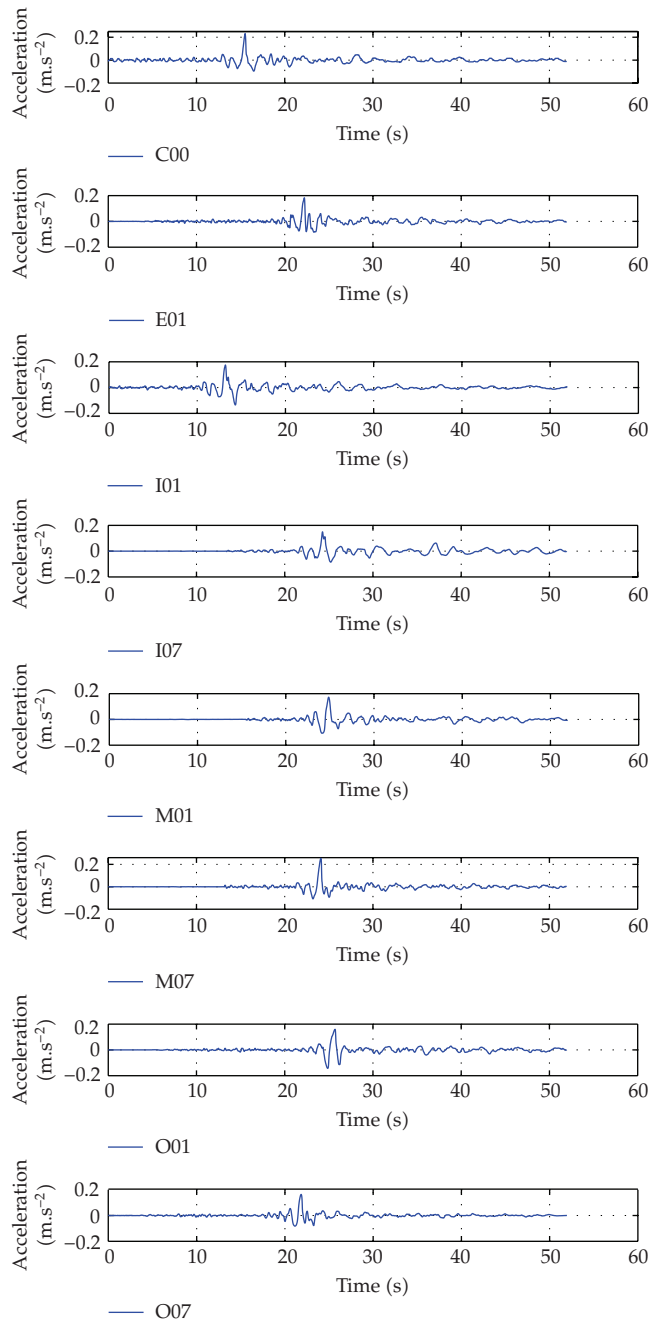
Figure 9: Time history curve of SMART1-40.

Acknowledgment

Table 2: The dynamic time warping distance between each point after mixing singular ground motion into the artificial multi-point ground motion field in Figure 2.

DTW distance	El Centro	200 m	400 m	600 m	800 m	1000 m	1200 m	1400 m	1600 m	1800 m	2000 m
El Centro	0	**75**	**77**	**79**	**78**	**85**	**85**	**84**	**85**	**88**	**83**
200 m	**75**	0	41	41	42	41	41	43	41	41	40
400 m	**77**	41	0	42	43	44	44	45	42	44	43
600 m	**79**	41	42	0	45	45	46	46	45	44	45
800 m	**78**	42	43	45	0	47	50	49	49	49	49
1000 m	**85**	41	44	45	47	0	50	50	52	49	52
1200 m	**85**	41	44	46	50	50	0	52	52	52	51
1400 m	**84**	43	45	46	49	50	52	0	51	52	49
1600 m	**85**	41	42	45	49	52	52	51	0	53	50
1800 m	**88**	41	44	44	49	49	52	52	53	0	51
2000 m	**83**	40	43	45	49	52	51	49	50	51	0

Table 3: The dynamic time warping distance between each point after mixing singular ground motion into SMART1-40 multi-point ground motion field.

DTW distance	C00	E01	I01	I07	M01	M07	O01	O07	El Centro
C00	0	12	9	9	13	10	14	12	**28**
E01	12	0	10	11	9	9	12	8	**22**
I01	9	10	0	9	11	11	12	12	**26**
I07	9	11	9	0	10	11	11	12	**22**
M01	13	9	11	10	0	9	8	7	**19**
M07	10	9	11	11	9	0	10	10	**23**
O01	14	12	12	11	8	10	0	11	**25**
O07	12	8	12	12	7	10	11	0	**18**
El Centro	**28**	**22**	**26**	**22**	**19**	**23**	**25**	**18**	0

References

[1] P. Danguang, L. Menglin, and F. Lichu, "Status of seismic response analysis of long-span structures under multiple support excitations," *Journal of Tongji University*, vol. 29, no. 10, pp. 1213–1219, 2001.

[2] M. Berrah and E. Kausel, "Response spectrum analysis of structures subjected to spatially varying motions," *Earthquake Engineering and Structural Dynamics*, vol. 21, no. 6, pp. 461–470, 1992.

[3] H. Yuxian, *Earthquake Engineering*, Seismological Press, Beijing, China, 2006.

[4] H. Hao, C. S. Oliveira, and J. Penzien, "Multiple-station ground motion processing and simulation based on SMART-1 array data," *Nuclear Engineering and Design*, vol. 111, no. 3, pp. 293–310, 1989.

[5] M. K. Berrah and E. Kausel, "A modal combination rule for spatially varying seismic motions," *Earthquake Engineering & Structural Dynamics*, vol. 22, no. 9, pp. 791–800, 1993.

[6] D. Lubo, *Nonlinear analysis of free spanning submarine pipelines subjected to spatially varying earthquake ground motions*, doctoral dissertation, Dalian University of Technology, Dalian, China, 2008.

[7] Q. Tiejun and W. Qianxin, "Simulation of spatial correlative time histories of multi-point ground motion, part I: fundamental formulas," *Earthquake Engineering and Engineering Vibration*, vol. 18, no. 1, pp. 8–15, 1998.

[8] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "On clustering validation techniques," *Journal of Intelligent Information Systems*, vol. 17, no. 2-3, pp. 107–145, 2001.

[9] G. K. Das, H. Mannila, P. Renganatham, et al., "Rule discovery from time series," in *Proceedings of the 4th International Conference of Knowledge Discovery and Data Mining*, R. Agrawal, S. E. Paul, and P. S. Gregory, Eds., pp. 16–22, AAAI Press, New York, NY, USA, 1998.

[10] D. J. Berndt and J. Clifford, "Finding patterns in time series: a dynamic programming approach," in *Advances in Knowledge Discovery and Data Mining*, pp. 229–248, AAAI/MIT Press, Cambridge, Mass, USA, 1996.

[11] X. Hui, *Similarity search and outlier detection in time series*, doctoral dissertation, Fudan University, Shanghai, China, 2005.