

Estimation of peak floor acceleration based on support vector regression and p-wave features

Tzu-Kang Lin^a and Chang-Han Wu

Department of Civil Engineering, National Chiao Tung University, Hsinchu 300, Taiwan

Abstract. This study proposes a support vector regression (SVR)-based method to evaluate the peak floor acceleration (PFA) in a building subject to a major earthquake. Support vector machine, which constructs a hyperplane in high-dimensional space to solve multivariate problems, is well known for the outstanding performance in classification and regression. Six P-wave parameters obtained from the vertical component of the first three seconds of the acceleration time history of the ground motion, as well as the floor height and the fundamental period of the structure, are adopted as the eight input variables. With these and the support of SVR, the PFA of a specific building can be rapidly estimated to avoid the necessity of installing accelerometers on the building. The SVR model is trained on 2274 representative earthquake records from the Structure Strong Earthquake Monitoring System (SSEMS) of the Central Weather Bureau (CWB) in Taiwan and tested on 757 independent test earthquake records. As demonstrated, the accuracy of the predicted PFA, located within a one-level difference on the seismic intensity scale of Taiwan from the monitored PFA, is 95.51%. The developed algorithm can be integrated into an existing earthquake early warning system (EEWS) to provide comprehensive protection during major earthquakes.

1 Introduction

Two types of EEWS, the regional early warning system and the on-site early warning system, have been adopted in Taiwan. Although an early warning of an earthquake can be announced by the regional early warning system a few seconds before a major earthquake arrives, it requires the P-wave information of several stations and, therefore, it may take approximately 20 to 30 seconds to announce the warning signal. Under this circumstance, regions that are 60 km from the epicenter then belong to the blind space of the early warning system (for the velocity of the S-wave estimated at 3 km/s). On the other hand, as the P-wave information of only one station is required by the on-site early warning to predict the peak ground acceleration (PGA) of an earthquake, the blind space in this instance can be much smaller than the previously described one.

In order to enable earlier earthquake alerts to reduce possible casualties when an earthquake occurs, the National Center for Research on Earthquake Engineering (NCREE) developed an on-site EEWS, which can predict the PGA of the whole time history of the earthquake before the major vibration arrives [1]. The entire operation process is shown in Figure. 1.

However, all current EEWSs focus only on the prediction of the PGA, but the PFA is more

^a Corresponding author : tklin@nctu.edu.tw

important than the PGA from a structural perspective. As a result, this study proposes an approach to predict the PFA so that the earthquake’s influence on a building can be easily assessed. It is expected that an early warning for each floor potentially under threat can be sent out automatically so that efficient evacuation can be conducted to reduce possible casualties.

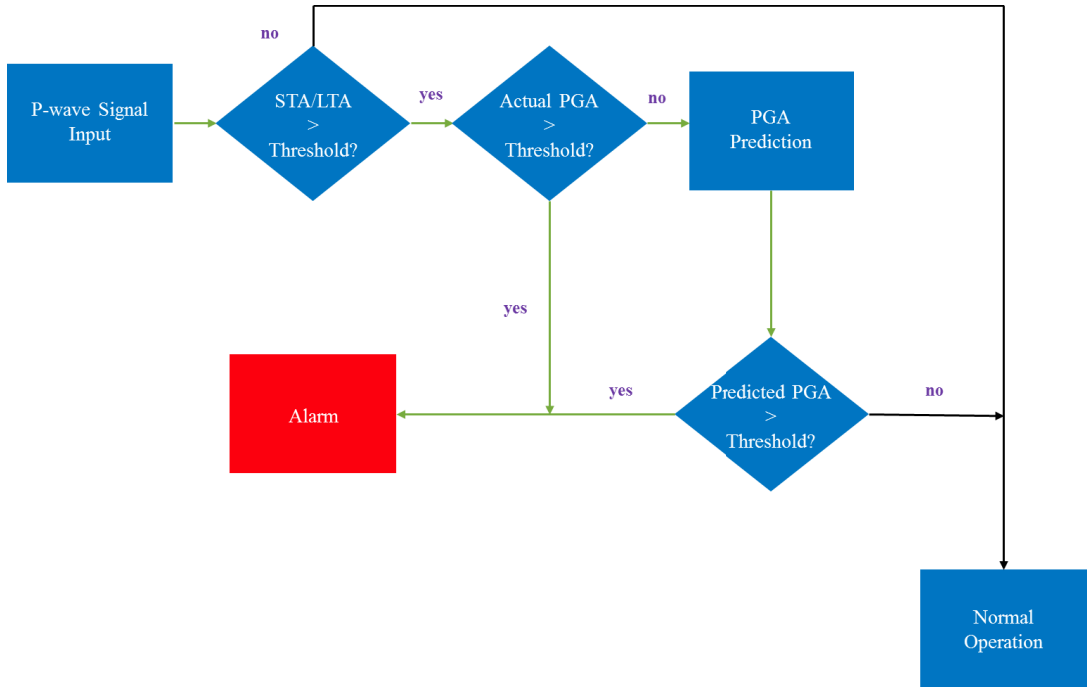


Figure 1. The operation process of an on-site EEWs.

2 The proposed PFA prediction method

2.1 P-wave features

According to Satriano [2], the P-wave features including peak measurements, integral quantities [3], the predominant period, and the average peak can be applied to an EEWs. Since the average peak is similar to the peak measurements, and as it is necessary to set some thresholds in computation, the average peak is not adopted as a parameter for analysis in this study.

Let t_p seconds be the time of observation after a P-wave arrives at the station from the epicenter, and the time that the P-wave arrives at the destination is called P-wave Arrival, which can be determined by a short-term average/long-term average (LTA/STA) method. Three seconds after the P-wave arrives, that is t_p equals 3 seconds, the P-wave features, which are the peak measurement of acceleration (P_a), peak measurement of velocity (P_v), peak measurement of displacement (P_d), effective predominant period (T_c), integral of squared velocity (IV2), and cumulative absolute velocity (CAV), can be quickly estimated from the time history of vertical earthquake acceleration. Of these parameters, peak measurements (P_a , P_v , and P_d) are associated with the PGA, peak ground velocity, and peak ground displacement, respectively, and Kanamori [4] suggested that T_c and IV2 are correlated to the magnitude of an earthquake. Additionally, CAV can be used to detect whether there is a destructive earthquake coming. The T_c , IV2 [5], and CAV [6] can be calculated as:

$$T_c = \frac{2\pi}{\sqrt{r}}, \text{ where } r = \frac{\int_0^{t_p} \dot{u}(t)^2 dt}{\int_0^{t_p} u(t)^2 dt} \quad (1)$$

$$IV2 = \int_0^{t_p} \dot{u}(t)^2 dt \quad (2)$$

$$CAV = \int_0^{t_p} |\ddot{u}(t)| dt \quad (3)$$

where $u(t)$, $\dot{u}(t)$, and $\ddot{u}(t)$ are the vertical components of the displacement, velocity, and acceleration time histories of ground motion after P-wave Arrival, respectively.

Since the primary earthquake database from the CWB consists of acceleration time-history information, it is necessary to integrate the acceleration to obtain the velocity and displacement for P-wave features. Moreover, a high-pass filter is used to correct for bias shifting at low frequency, where the order and the cutoff frequency of the filter are 2 and 0.075 Hz, respectively.

2.2 Support vector regression

Support vector machine (SVM), a data recognition technique, can be used to generalize and classify the database through image recognition. SVM can be classified as either support vector classification (SVC) or support vector regression (SVR), in accordance with its application purpose. Compared with other analysis methods, the major advantage is that SVR seldom has problems with local minima, overfitting, or underfitting. However, when establishing the SVM model, the amount of training data must be of sufficient representativeness, so that the trained model can be used to classify or regress accurately.

The main concept of SVR is that the original data space can be projected to the hyperspace by using a nonlinear projection function, and the precise outcome can be predicted easily. Consequently, the prediction of PFA for a building is attempted by using SVR [7].

Take a training data set $(x, y) \in \mathbb{R}$, where x represents input data, and each set of data can be one or multiple parameters, and y represents the actual value related to the input training data. If the values of input data x have a direct or indirect correlation to the corresponding output value y , then SVR can be employed to find the mathematical relationship between them. The regression function can be expressed as:

$$f(w, b) = w \cdot \phi(x) + b \quad (4)$$

where w is the regression coefficient vector representing the vector of the hyperspace and b is the offset, which represents the distance between the hyperspace and the origin.

If the error between y_i and $f(x_i)$ is less than the error tolerance ε for each x_i , then $f(x_i)$ is approximately equal to y_i . However, as the error of the data is often larger than the error tolerance, which is out of the ε -insensitive area, it is necessary to add ξ_i and ξ_i^* to allow the condition.

The ν -support vector regression problem can be expressed as the following convex optimization problem [8]:

$$\min_{w, b, \xi, \xi^*} \left[\frac{1}{2} \|w\|^2 + C \left(\nu \varepsilon + \frac{1}{m} \sum_{i=1}^m (\xi_i + \xi_i^*) \right) \right] \quad (5)$$

subjected to:

$$w \cdot \phi(x_i) + b - y_i \leq \varepsilon + \xi_i$$

$$y_i - w \cdot \phi(x_i) - b \leq \varepsilon + \xi_i^*$$

$$\xi_i, \xi_i^* \geq 0 \text{ for } i = 1, \dots, m$$

where m is the number of training data, C is a positive constant that determines the degree of penalized loss when a training error occurs, and $v\varepsilon$ is a slack variable controlled by the coefficient v .

By introducing the method of Lagrange multipliers and setting the Lagrange multipliers α_i and β_i , Eq. 5 can be solved by applying the standard quadratic programming algorithm:

$$\max_{\alpha, \beta} \left[-\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (\alpha_i - \beta_i)(\alpha_j - \beta_j) k(x_i, x_j) + \sum_{i=1}^m y_i (\alpha_i - \beta_i) \right] \quad (6)$$

subjected to:

$$\sum_{i=1}^m (\alpha_i - \beta_i) = 0 \leq \alpha_i$$

$$\beta_i \leq \frac{C}{m} \text{ for } i = 1, 2, \dots, m$$

$$\sum_{i=1}^m (\alpha_i + \beta_i) = 0 \leq Cv$$

where m is the number of training data and $k(x_i, x_j)$ is the kernel function to yield the inner products in feature space, which is adopted as a radial basis kernel function as follows:

$$k(x_i, x_j) = \exp(-|x_i - x_j| / 2\sigma^2) \quad (7)$$

After the Lagrange multipliers, α_i and β_i , are determined, the parameters w and b can be estimated under the Karush–Kuhn–Tucker complementarity conditions [9]. Therefore, the prediction function can be expressed as:

$$f(x) = \left[\sum_{i=1}^j (\alpha_i - \beta_i) k(x, x_i) \right] + b \quad (8)$$

where j is the number of nonzero terms $(\alpha_i - \beta_i)$, i.e., the support vectors.

3 Earthquake time-history database

Earthquake data has been measured by the Structure Strong Earthquake Monitoring System (SSEMS) of the CWB since 1992. The system collects original earthquake records from 39 structures in Taiwan, where accelerometers are installed at different heights. The sample frequency of each data set is 200 Hz or 250 Hz, and the resolution is typically 16 bits.

Initially, earthquakes whose magnitudes are larger than $M_L = 6.0$ are chosen. In addition to the six P-wave features used to predict the PFA of a building, the floor height and the fundamental period of a structure are also considered. Consequently, the height of each accelerometer in the structure is

checked one by one and the fundamental period of a structure is estimated as one tenth of the structure. The database is carefully filtered, and the data with a time duration of less than three seconds are removed. As a result, 20 stations are finally adopted, and the locations of each structure are shown in Figure. 2. Some 788 sets of earthquake time-history data, namely 3031 points of sample data, are used after the converting for the height of each structure.

Based on the definition of PFA, the peak acceleration values of each accelerometer channel in the whole seismic event are first selected, and then compared with the peak acceleration value of another accelerometer channel at the same height to determine the real PFA. Consequently, for a seismic event, a set of six P-wave features, the fundamental period of a structure, several floor heights, and several real PFAs can be established. The details of the 20 structure stations are given in Table 1.

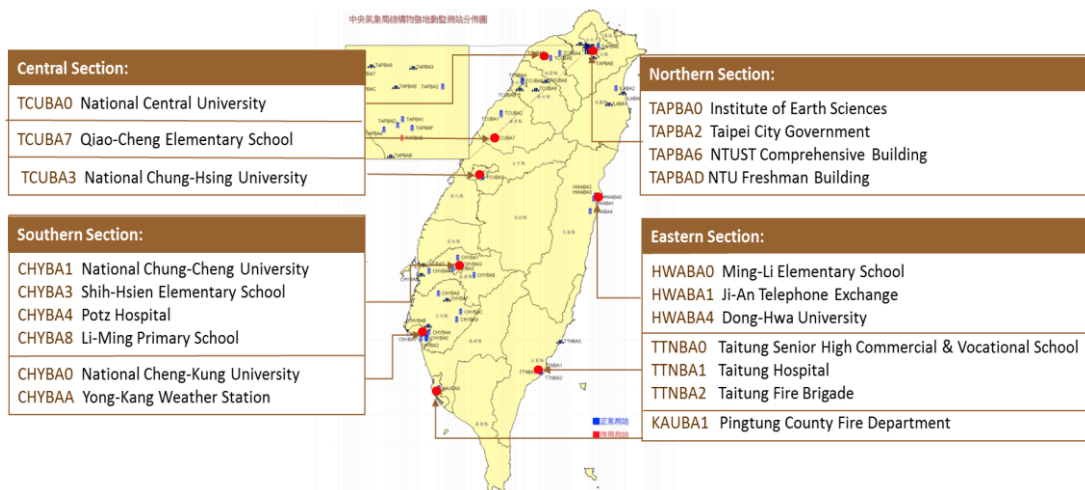


Figure 2. The locations of each station structure.

Table 1. Distribution of sample data.

Station code	Number of floor types	Number of earthquake events	Number of sample data
CHYBA0	3	17	51
CHYBA1	4	84	336
CHYBA3	5	55	275
CHYBA4	5	42	210
CHYBA8	2	35	70
CHYBAA	4	57	228
HWABA0	4	50	200
HWABA1	4	46	184
HWABA4	4	72	288
KAUBA1	4	4	16
TAPBA0	5	16	80
TAPBA2	4	22	88
TAPBA6	4	61	244
TAPBAD	4	32	128
TCUBA0	4	48	192
TCUBA3	3	39	117
TCUBA7	3	12	36
TTNBA0	3	34	102
TTNBA1	3	38	114
TTNBA2	3	24	72
Total		788	3031

4 Analysis result

A total of 2274 points of sample data from the SSEMS in the CWB are used as the representative earthquake data to train the SVR model by means of the eight relevant parameters, namely the six P-wave features, the floor height, and the fundamental period of a structure. On the basis of the correlation between the eight relevant parameters and the real PFA, the predicted PFA can be quickly evaluated. Furthermore, as the floor height and the fundamental period of a structure are fixed, they can be preset before an earthquake occurs.

In this study, a grid search method is used to determine the optimal values of C and σ , where the ranges of C and σ are 2^5 to 2^{15} and 2^{-1} to 2^{-12} , respectively. After searching within the confined range, the optimal values of C and σ can be determined with the smallest mean squared error.

As mentioned above, the data samples are used to train the SVR model, and the reserved testing samples are inputted into the trained SVR model to obtain the predicted PFA. For the purpose of checking the performance of the SVR model, the result is defined as “accurate” if the seismic intensity of the predicted PFA is located within a one-level difference of the monitored PFA, according to the classification criterion of the CWB.

Figure. 3 shows the regression result for the 2274 representative earthquake data. The best values of C and σ are 4096 and 0.0156, respectively, and the squared correlation coefficient is 0.89813. Figure. 4 shows the regression results for the 757 testing earthquake data, and the squared correlation coefficient is 0.319626. The accuracy of the predicted PFA located within a one-level difference of the seismic intensity scale from the real PFA is 95.51%.

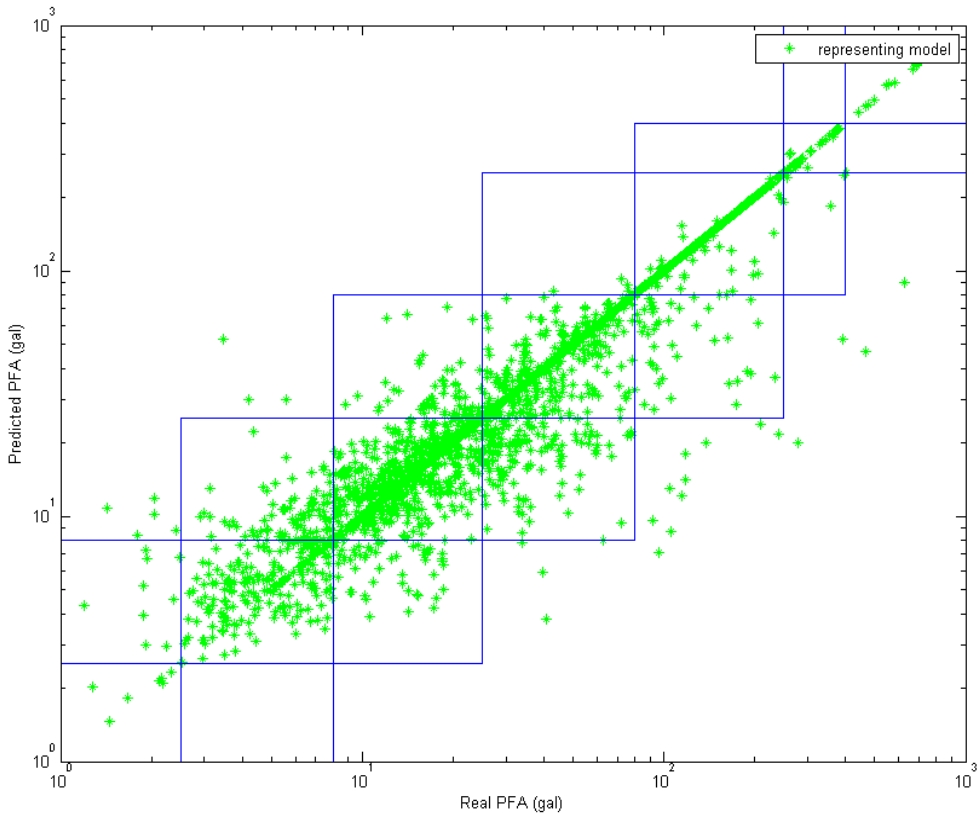


Figure 3. The training results of the representative earthquake records.

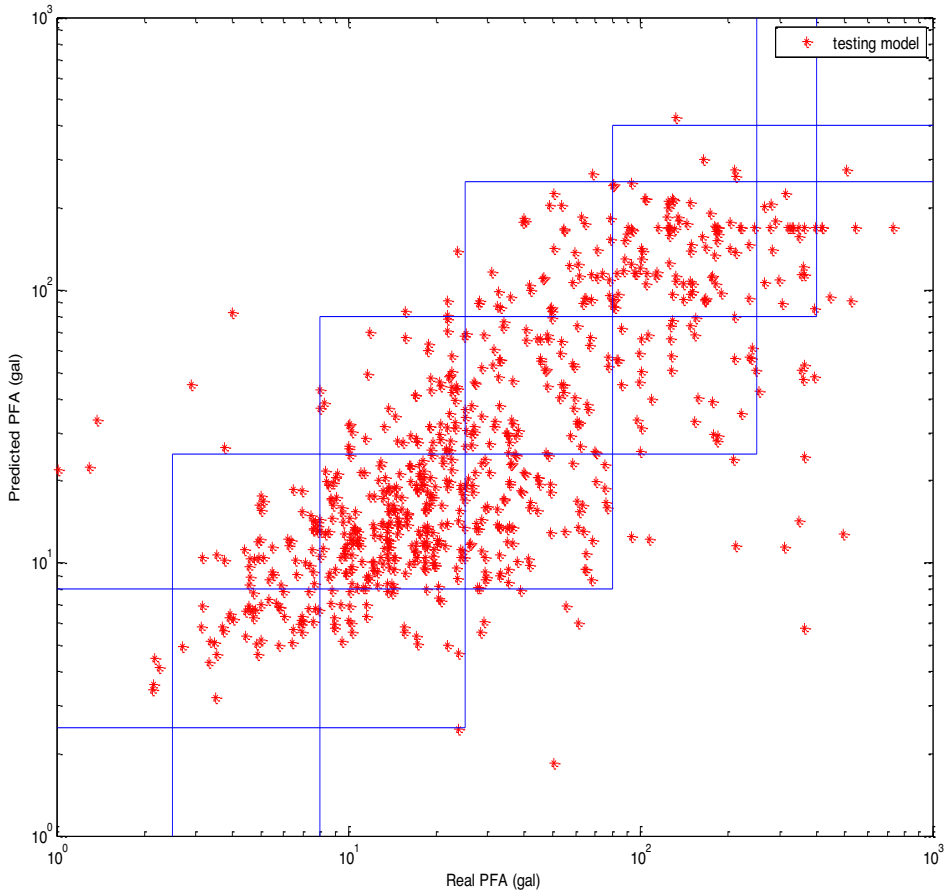


Figure 4. The testing results of the testing earthquake records.

In order to evaluate the practical applicability of this approach more rigorously, three earthquake events are randomly selected from the testing earthquake records, and the corresponding information is shown in Table 2.

Table 2. Details of the three earthquake events.

Earthquake events	Station code	Earthquake occurrence time
1	CHYBA0	2000.06.10
2	TAPBA6	2009.10.03
3	HWABA4	2007.09.06

For these three earthquake events, the seismograms are first examined. A comparison of the PFAs and seismic intensities is also made, and the results are shown in Figures. 5 and 6, Figures. 7 and 8, and Figures. 9 and 10, respectively. As indicated by the bar charts, it is obvious that most of the predictions are certainly larger than their corresponding real PFAs, except in the case of earthquake event 1 at H = 20.95 m, and all the predicted intensities are equivalent to or one-level larger than the real value. Therefore, a conservative estimation of the seismic intensity can be provided by these three earthquake events, and an earthquake early warning can be effectively provided to the residents.

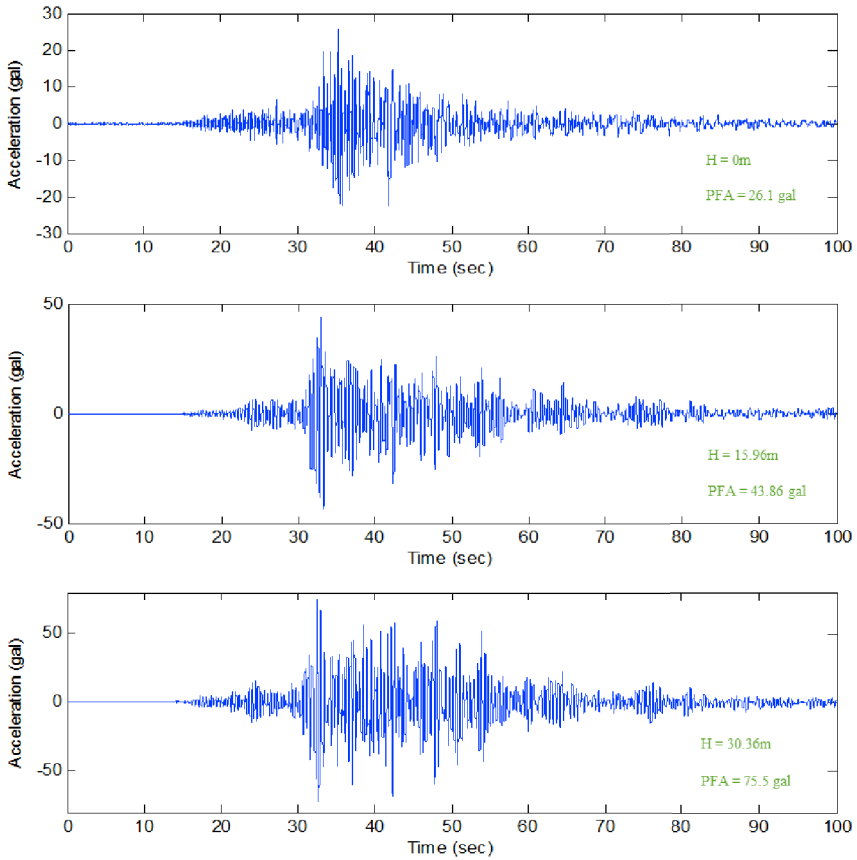


Figure 5. Earthquake event 1: seismograms.

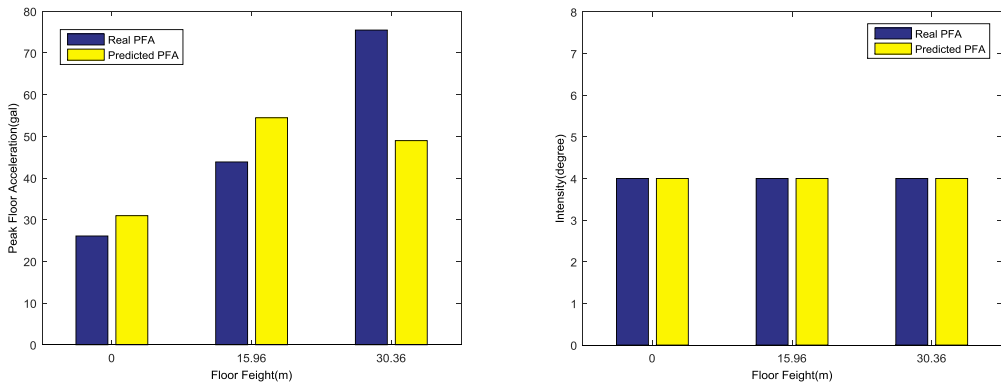


Figure 6. Earthquake event 1: comparison of PFAs and intensities.

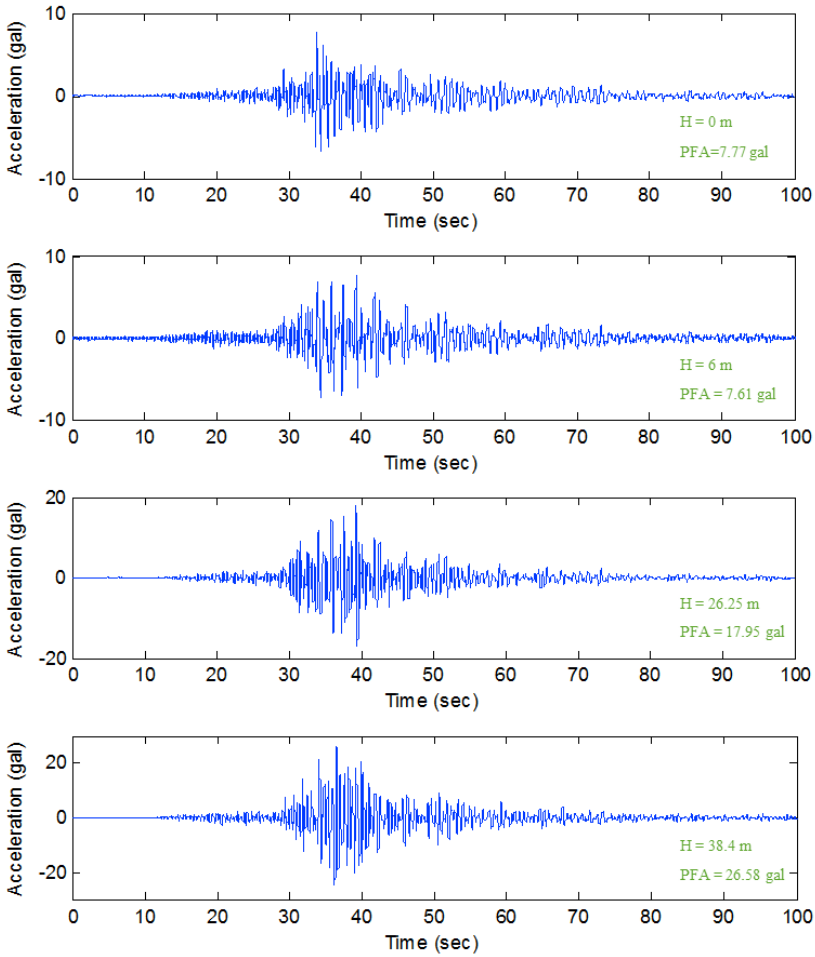


Figure 7. Earthquake event 2: seismograms.

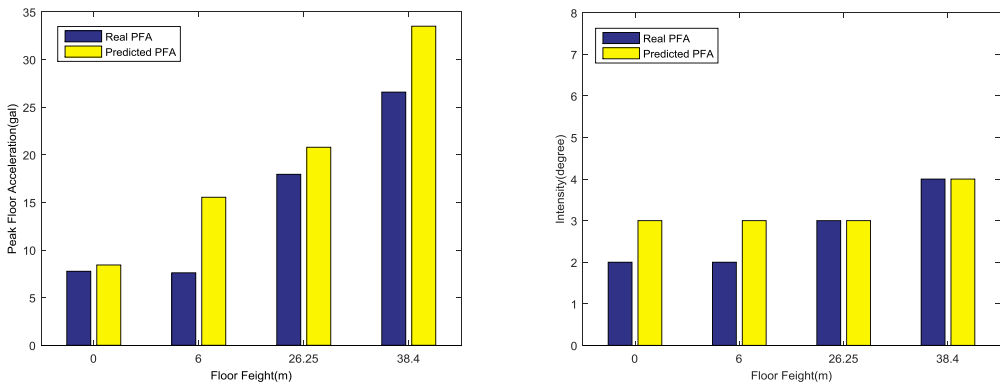


Figure 8. Earthquake event 2: comparison of PFAs and intensities.

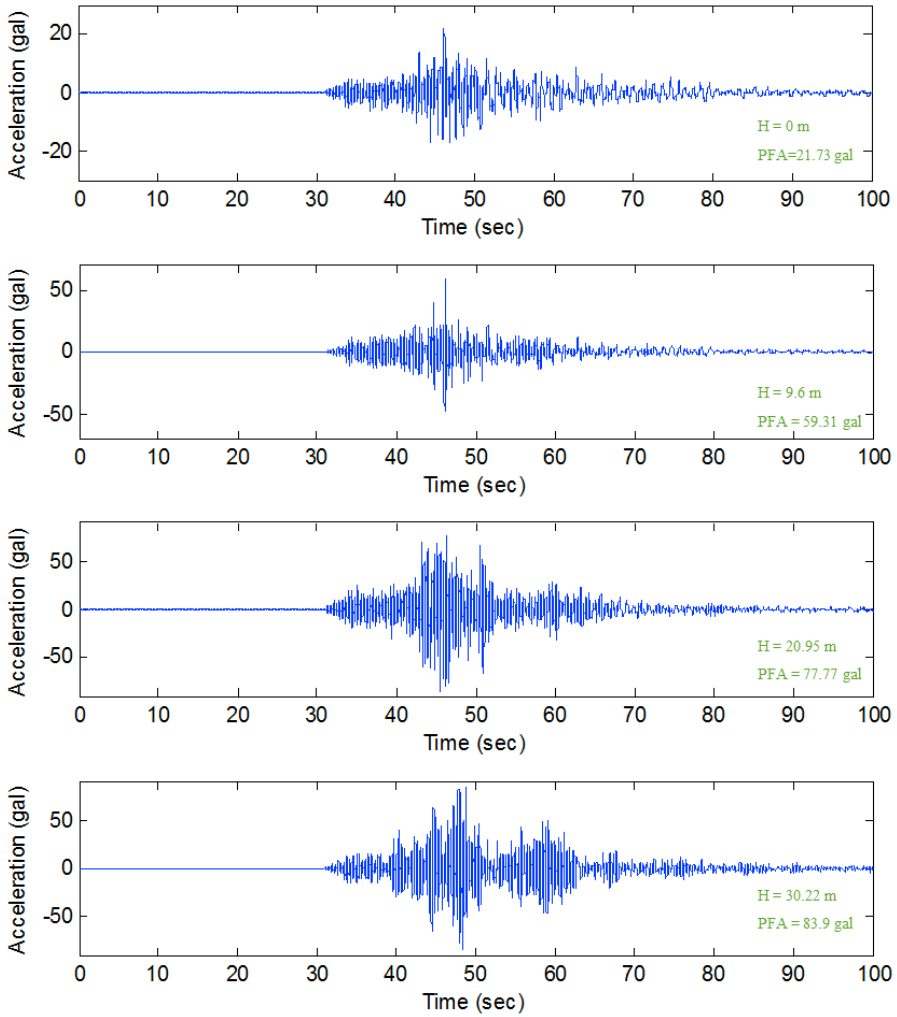


Figure 9. Earthquake event 3: seismograms.

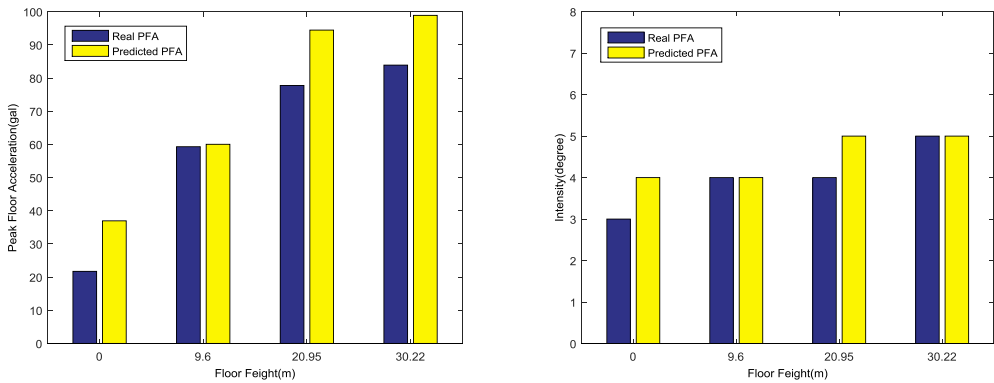


Figure 10. Earthquake event 3: comparison of PFAs and intensities.

5 Summary and conclusion

EEWS have been developed to quickly predict the peak ground acceleration of an impending and to provide early warnings instantly. To enhance the functionality of the existing EEWS system, making precise and rapid predictions of the PFA of a building to allow an evacuation has also become an important issue. In this study, an approach for predicting the PFA from an on-site EEWS was proposed. In addition to using the six P-wave features as characteristic parameters, the floor height and the fundamental period of a structure were also included to estimate the PFA of a building through SVR when an earthquake occurs. For the earthquake time-history database adopted in this study, after testing the earthquake records, which are independent of the representative earthquake records, the accuracy of the predicted PFA located within a one-level difference of the seismic intensity scale from the real PFA is 95.51%. Thus, the accuracy demonstrates that the PFA evaluated by this approach is relatively reliable.

To enhance the performance of the proposed method, more records will be included in the SVR-based method in the near future, especially earthquake records with smaller earthquake magnitudes or records collected from other on-site stations.

References

1. T.Y. Hsu, S.K. Huang, Y.W. Chang, C.H. Kuo, C.M. Lin, T.M. Chang, K.L. Wen, and C.H. Loh, Rapid on-site peak ground acceleration estimation based on support vector regression and P-wave features in Taiwan, *SDEE*, **49**, 210-217 (2013)
2. C. Satriano, Y.M. Wu, A. Zollo, and H. Kanamori, Earthquake early warning: concepts, methods and physical grounds, *SDEE*, **31**, 106-118 (2011)
3. M. Böse, T. Heaton, and E. Hauksson, Rapid estimation of earthquake source and ground-motion parameters for earthquake early warning using data from a single three-component broadband or strong-motion sensor, *BSSA*, **102**, 738-750 (2012)
4. H. Kanamori, Real-time seismology and earthquake damage mitigation, *Annu Rev Earth Planet Sci*, **33**, 195-214 (2005)
5. G. Festa, A. Zollo, and M. Lancieri, Earthquake magnitude estimation from early radiated energy, *Geophys Res Lett*, **35** (22) (2008)
6. M. Erdik, Y. Fahjan, O. Ozel, H. Alcik, A. Mert, and M. Gul, Istanbul earthquake rapid response and the early warning system, *B EARTHQ ENG*, **1** (1), 157-63 (2003)
7. V.N. Vapnik, *Springer-Verlag*, The nature of statistical learning theory (1995)
8. B. Schölkopf, A. Smola, R. Williamson, and P.L. Bartlett, New support vector algorithms, *Neural Comput*, **12**, 1207-45 (2000)
9. R. Fletcher, *Practical Methods of Optimization*, Second Edition, John Wiley (1987)