SECOND EUROPEAN CONFERENCE ON EARTHQUAKE ENGINEERING AND SEISMOLOGY, ISTANBUL AUG. 25-29, 2014

Turkish Earthquake Foundation - Earthquake Engineering Committee
Prime Ministry, Disaster And Emergency Management Presidency

GROUND MOTION ESTIMATION USING FRONT SITE WAVE FROM DATA BASED ON RVM FOR EARTHQUAKE EARLY WARNING

Yincheng YANG[1] and Masato MOTOSAKA [2]

ABSTRACT

When a large earthquake occurs, an earthquake early warning system provides the warning several seconds before the impending disastrous ground motions occur, allowing for immediate mitigation actions to be taken. This paper presents the application of Sparse Bayesian Regression Method, which also can be called Relevant Vector Machine (RVM) for the estimation of magnitude, peak ground acceleration (PGA) and peak ground velocity (PGV) of magnitudes more than 5.0 and hypocentral distances less than 301 km. Earthquake PGA, PGA, PGD, pulse rise times, average period and the ratio of $V_{\mathrm{pmax}} / A_{p\max}$ of first 3 to 6 sconds of P-wave have been used as six input variables to train the regression model. This study shows that RVM is a useful tool for the prediction of peak ground shaking at a site.

Keywords: ground motion prediction, earthquake early warning, Relevant Vector Machine

INTRODUCTION

The Earthquake Early Warning System (EEWS) is considered to be one of the most useful tools for emergency response in reducing seismic hazards, as it requires the seismic data to be recorded and analyzed shortly after an earthquake occurs. As the development of seismic instrumentation, communication technology, computer capability, and data storage facility is enhanced, the “near” real-time seismology for rapid seismic analysis and notification is becoming reality. The EEWS can be divided into two approaches: (a) onsite warning and (b) regional warning. In (a), the observed beginning of P-wave arrivals at a single station are used to predict the ensuing ground motion at the same site. In (b), the observed beginning of P-wave arrivals at a site are used to locate an earthquake, identify the magnitude and estimate the ground motion at other sites.

For the EEWS, determing earthquake magnitude and ground shaking (PGA, PGV) are the most integral parts, as they provide information with a few seconds warning for impending ground motions due to a large earthquake. Many researchers have developed a large number of methods to estimate the real-time magnitude and ground shaking. There are two different approaches to the EEW magnitude estimation, one being the predominant period method (Nakamura 1988, Allen and Kanamori 2003, Kanamori 2005, M.Allen 2005), and another being the amplitude method (Toshikazu Odaka 2003, Wu and Kanamori 2005b, Wu 2006, Zollo 2006). The predominant method shows that the predominant

[1] Ph.D. student, Tohoku University, Sendai, Japan, yinchengyang@gmail.com

[2] Professor, International Research Institute of Disaster Science, Tohoku University, Sendai, Japan, Motosaka@saigai.str.archi.tohoku.ac.jp

period reflects the size of earthquakes and has a linear relationship with magnitude (Kanamori 2005). Wu and Kanamori (2005b), and Wu (2006) found a linear relationship among the peak displacement amplitude Pd , the hypocentral distance R and the magnitude. In recent years, some researchers have tried to find a direct relationship between the first three to six seconds of the P-wave PGA and the PGA of the whole time history (Hagiwara 2011). This methodology is aimed to give a regional warning in some certain areas.

In order to improve the precision and stability for the prediction of magnitude and ground shaking, we use another method by including various period parameters and amplitude parameters: peak acceleration, peak velocity, peak displacement, predominant period, the ratio of peak velocity and peak acceleration, and effective cumulative ascending pulse width of first width and of first displacement pulse for every second after P-wave arrival. All of the parameters of different frequency bands were synthesized using the sparse Bayesian Regression Model (Tipping and Faul 2003).

GROUND MOTION DATA

The ground motion records used in training have been obtained from the K-net database. A total of 69 records have been selected using the following criteria:

- The epicenter of earthquakes are located in a window bounded by $37^{\circ}30'00''\mathrm{N}$ to $39^{\circ}00'00''\mathrm{N}$ latitude and $141^{\circ}30'00''\mathrm{E}$ to $145^{\circ}00'00''\mathrm{E}$ longitude;
- The earthquakes used in this study were recorded by the following K-net stations: Ojika (MYG011) from 2005 to 2013;
- Magnitude larger than 5.0;
- The range of distance between the station (Ojika) and rupture surface is 21-301 km;
- The range of depth is 0-86 km;

Detailed information regarding these records can be found by NIED (Hsiao, Wu et al. 2011)

A baseline correction of all acceleration time histories has been performed by subtracting the mean value of time history. All accelerograms were lowcut filtered by removing the frequencies below 0.1 Hz. A fourth-order Butterworth lowcut filter was used for this filtering operation. For the K-NET, as the natural frequencies of all accelerographs were very high (about 200 Hz), there was no need of the instrument response correction.

PREDICTIVE AND TARGET VARIABLES

In this study, we chose six variables as predictive parameters which will be used as training samples to predict the target variables. The six parameters include three period parameters: P-wave pulse rise times $\tau_{1/2}$, average period τ_c, the ratio of $V_{\mathrm{pmax}}/A_{p\max}$ and three amplitude paremeters: PGA_p, PGV_p, PGD_p of the different time period of P-wave. In the regression model, the input vector represent i^{th} record as follows:

$$x_i = \left[\log_{10}\left(PGA_p\right)_i \quad \log_{10}\left(PGV_p\right)_i \quad \log_{10}\left(PGD_p\right)_i \quad (\tau_{1/2})_i \quad (\tau_c)_i \quad (V_{p\max}/A_{p\max})_i \right] \tag{1}$$

In the equation(1), all the parameter values were scaled linearly to [-1 1], in order to prevent different problems because of large variations between the ranges of variables.

The output (target) for the i^{th} record for magnitude, PGA and PGV seperately has the following form:

$$t_i = M_i + \varepsilon \tag{2}$$

$$t_i' = \log_{10}\left(PGA\right)_i + \varepsilon \tag{3}$$

$$t_i'' = \log_{10}\left(PGV\right)_i + \varepsilon \tag{4}$$

Once the earthquake occurs, the propagation of seismic wave will be received one after another by the different observed front sites, thus making the records continuous. As time goes on, the prediction can be updated by the increase of record data. In this research, the initial 1 to 6 seconds of the P-wave data had been used to obtain the six predictive parameters.

METHODOLOGY

The sparse Bayesian regression model (M. E. Tipping and A. C.Faul 2003) would initially take the form:

$$y(x,w)=\sum_{m=1}^{\mathrm{M}}\omega_m\phi_m(x) \tag{5}$$

Given a data set of input vectors $\{x_i\}_{i=1}^{\mathrm{N}}$, which is the feature value of training data, and corresponding scalar-valued targets $t=[t_1,t_2,\dots,t_{\mathrm{N}}]^{\mathrm{T}}$, we can then get the targets to follow the standard probabilistic formulation which are samples from the model with addictive noise:

$$\begin{aligned} t_n &= y(x_n;w)+\varepsilon_n \\ &=\Phi w+\varepsilon \end{aligned} \tag{6}$$

where the w is the parameter vector of $\mathrm{w_i}$, and Φ is the $\mathrm{N}\times\mathrm{M}$ design matrix compose of kernel function K, ε is an independent zero-mean Gaussian distribution noise samples, with variance σ^2, so $p(\varepsilon)=\prod_{n=1}^{\mathrm{N}}\mathrm{N}(\varepsilon_n\,|0,\sigma^2)$.

This research uses the Gaussian radial basis function (KBF) kernel:

$$K(x,x')=\exp(\gamma\left\|x-x'\right\|_2^2) \tag{7}$$

where γ is a free width parameter and the model accuracy and complexity is can be controlled by it.

After determining the weights and the noise variance, the predictive mean y_* for new input x_* can be obtain as:

$$y_*=\mu^{\mathrm{T}}\phi(x_*) \tag{8}$$

The final predictive error bars can be found by adding the maximizing value σ_{MP}^2 to the uncertainty as follow:

$$\sigma_*^2=\sigma_{MP}^2+\phi(x_*)^{\mathrm{T}}\Delta\phi(x_*) \tag{9}$$

RESULTS

In the former part of this paper, we had introduced the sparse Bayesian model which also can be called as relevant vector machine, the width paremeter γ was said to be used to control the regression's accuracy by changing its value. In this research, the γ was assaigned the range of [0 3]. For the purpose of finding the optimal value of γ for each magnitude, PGA and PGV, the 10-fold cross validation method (Andrew R. Webb 2011) was used. In this method, the training data is randomly assigened to 10 groups with consistant size, and nine groups are used to train the model while the remaining group is for assessing the validation error. This process is repeated 10 times until every group is used, and each time the validation group changes, the mean value of validation error for each γ is calculated. Finally, the optimal γ with the smallest validation error is determined by calculating the mean validation error for the whole range of γ.

After obtaining the onset time of the P-wave, the front site of the P-wave arrival is obtained every second until the S-wave arrives. In this research, we use 3 to 6 seconds of the P-wave arrival data to train the regression model and to obtain the peak value of imminent S-wave ground shaking. The standard deviation of regression results, which have been categorized as equation (2) (3) and(4), is tabulated in Table 1. The comparison between the observed data and regression output with six inputs is shown in Fig. 1-3.

Table 1 Standard deviation of regression results by using 3 to 6 seconds P-wave segments

Regression targets	Standard Deviation			
	3S	4S	5S	6S
Mag.(JMA)	0.39	0.28	0.30	0.29
Log(PGA_ew)	0.31	0.24	0.22	0.20
Log(PGA_ns)	0.29	0.24	0.21	0.20
Log(PGV_ew)	0.31	0.27	0.24	0.23
Log(PGV_ns)	0.18	0.25	0.23	0.23

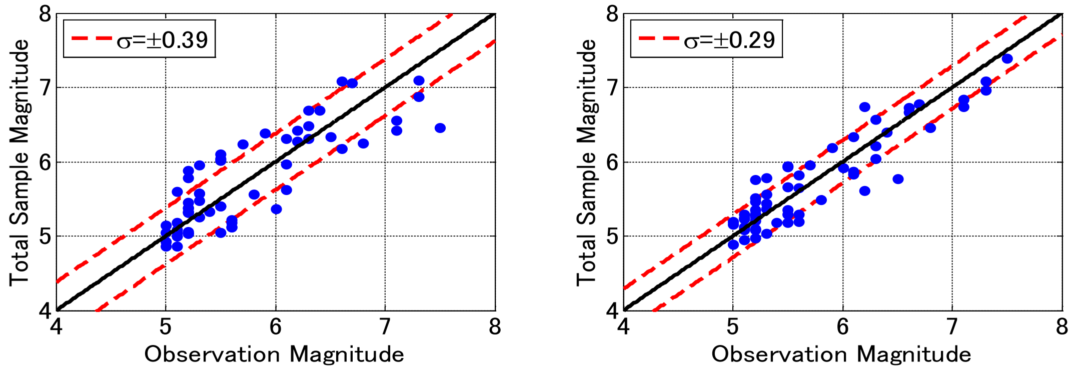

Figure 1 Magnitude regression results within 3 and 6 seconds of P-wave arrival

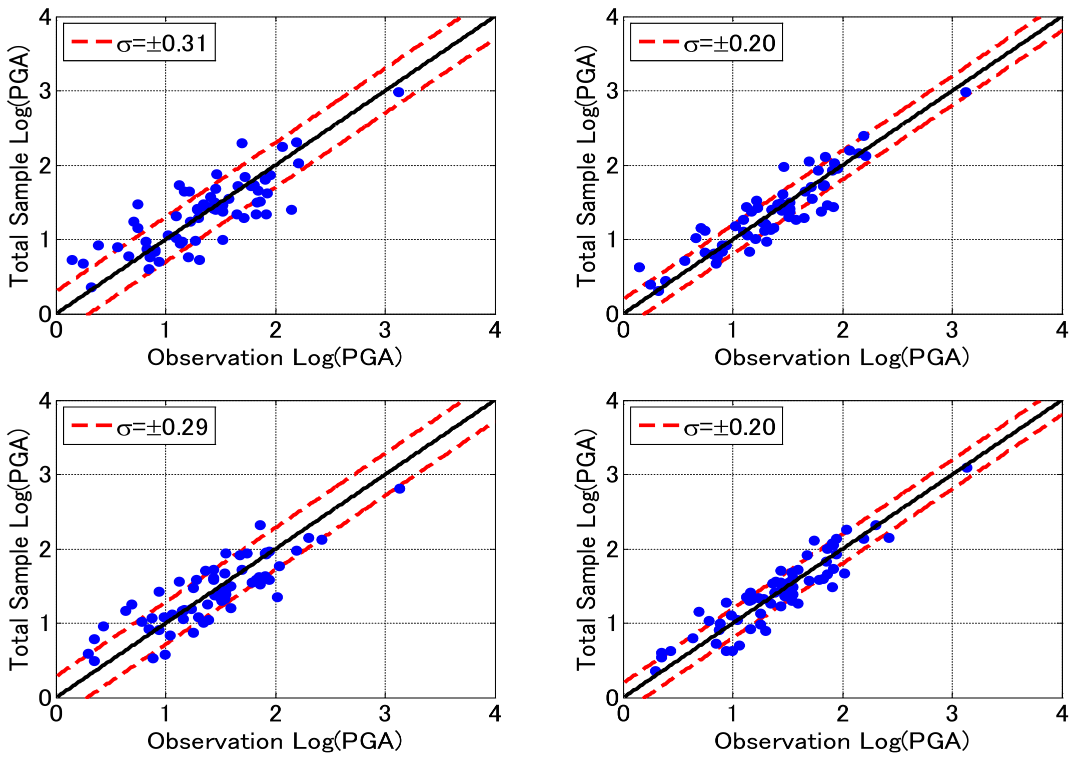

Figure 2 Horizontal Log (PGA) regression results within 3 and 6 seconds of P-wave arrival. The top two are in the EW direction and the bottom two are in the NS direction.

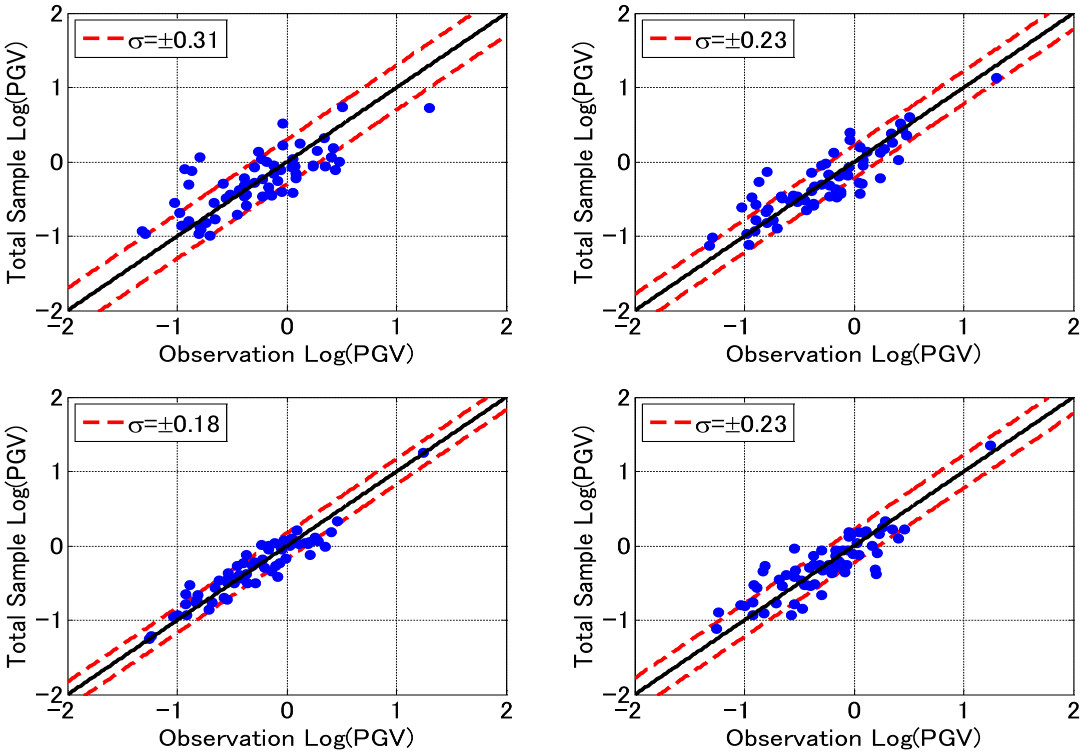

Figure 3 Horizontal Log (PGV) regression results within 3 and 6 seconds of P-wave arrival. The top two are in the EW direction and the bottom two are in the NS direction.

From the results, it is seen that the Sparse Bayesian Regression Model with six inputs has predicted accurate results on magnitude, PGA and PGV along with some inaccurate results. The standard deviation of predicted results has meet earthquake early warning needs.

APPLICATION

On 19:15 August 24, 2005, an earthquake (Mj=6.3) occurred in Miyagi-oki, Japan. The depth of this earthquake was 14 km approximately, and the distance of epicenter was 139 km. This earthquake's data is not included in the training part of regression model.

Table 2 Continual ground motion prediction using the initial P-wave segments

prediction targets	real value	ground shaking predicted results			
		3S	4S	5S	6S
Mag.(JMA)	6.30	6.55	6.84	6.40	6.46
Log(PGA_ew)	0.40	0.46	0.12	0.43	0.39
Log(PGA_ns)	0.50	0.54	0.51	0.48	0.45
Log(PGV_ew)	-0.85	-0.96	-0.83	-0.95	-1.02
Log(PGV_ns)	-0.69	-0.89	-0.81	-0.93	-1.10

From the results, it is seen that we can use the 3 to 6 seconds of P-wave arrival to predict the later peak ground motion of S-wave as well. The prediction for magnitude and PGA are better than PGV.

CONCLUSIONS

A Sparse Bayesian Regression Model (relevant vector machine) has been adopted to estimate magnitude (JMA), peak ground acceleration and peak ground velocity for the Japanese earthquake records with earthquake magnitudes more than 5.0 and hypocentral distance less than 86 km. The predicted results by this regression model have been found to be accurate to meet with earthquake early warning needs and this regression model can be applied to on-site warning. We believed that with an increased amount of stations receiving new waveform information, the result can be continually revised to become more accurate. The next step for our research is that we will try to use this regression model for the regional warning.

REFERENCES

Allen, R. M. and H. Kanamori (2003). The potential for earthquake early warning in southern California. **300:** 786-789.

Andrew R. Webb, K. D. C. (2011). "Statistical Pattern Recognition, 3rd Edition."

Hagiwara, Y. (2011). "study on the forecast of ground motion parameters from real time earthquake information based on wave form data at the front site " Institute of Social Safety Science **15**: 231-239.

Hsiao, N.-C., et al. (2011). "A new prototype system for earthquake early warning in Taiwan." Soil Dynamics and Earthquake Engineering **31**(2): 201-208.

Kanamori, H. (2005). "Real-Time Seismology and Earthquake Damage Mitigation." Annual Review of Earth and Planetary Sciences **33**(1): 195-214.

M. E. Tipping and A. C.Faul (2003). "Fast marginal likelihood maximisation for sparse Bayesian models." Proceedings of the ninth international workshop on artificial intelligence and statistics: 1-13.

M.Allen, E. L. O. a. R. (2005). "The deterministic nature of earthquake rupture." Nature 438: 212-215.

Nakamura, Y. (1988). "On the urgent earthquake detection and alarm system (UrEDAS)." Proc. of the 9th World Conference on Earthquake ….

Tipping, M. E. and A. C. Faul (2003). "Fast marginal likelihood maximisation for sparse Bayesian models." Proceedings of the ninth …: 1-13.

Toshikazu Odaka, K. A., Shin'ya Tsukada, Shinji Sato,Kazuo Ohtake, and Daisuke Nozaka (2003). "A new method of quickly estimating epicentral distance and magnitude from a single seismic record." Bulletin of the Seismological Society of America **93**: 526-432.

Wu, Y.-M. and H. Kanamori (2005b). "Rapid Assessment of Damaging Potential of Earthquakes in Taiwan from the Beginning of P Waves." Bulletin of the Seismological Society of America **95**: 1181-1185.

Wu, Y.-M., Yen,Hsin-Yi, Zhao,Li, Huang,Bor-Shouh and Liang,Wen-Tzong (2006). "Magnitude determination using initial P waves: A single-station approach." Geophysical Research Letters **33**(5): L05306-L05306.

Zollo, A., Lancieri, Maria and Nielsen, Stefan (2006). "Earthquake magnitude estimation from peak amplitudes of very early seismic signals on strong motion records." Geophysical Research Letters **33**(23): L23312-L23312.