

Implementing horizontal-to-vertical Fourier spectral ratios and spatial correlation in a ground-motion prediction equation to predict site effects

Earthquake Spectra

1–30

© The Author(s) 2020

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/8755293020952449

journals.sagepub.com/home/eqs

Shu-Hsien Chao¹, Che-Min Lin¹, Chun-Hsiang Kuo^{1,2},
Jyun-Yan Huang¹, Kuo-Liang Wen², and
Yi-Hau Chen³

Abstract

We propose a methodology to implement horizontal-to-vertical Fourier spectral ratios (HVRs) evaluated from strong ground motion induced by earthquake (EHVRs) or ambient ground motion observed from microtremor (MHVRs) individually and simultaneously with the spatial correlation (SC) in a ground-motion prediction equation (GMPE) to improve the prediction accuracy of site effects. We illustrated the methodology by developing an EHVRs-SC-based model which supplements Vs30 and Z1.0 with the SC and EHVRs collected at strong motion stations, and a MHVRs-SC-based model that supplements Vs30 and Z1.0 with the SC and MHVRs observed from microtremors at sites which were collocated with strong motion stations. The standard deviation of the station-specific residuals can be reduced by up to 90% when the proposed models are used to predict site effects. In the proposed models, the spatial distribution of the predicted station terms for peak ground acceleration (PGA) from MHVRs at 3699 sites is consistent with that of the predicted station terms for PGA from EHVRs at 721 strong motion stations. Prediction accuracy for stations with inferred Vs30 is similar to that of stations with measured Vs30 with the proposed models. This study provides a methodology to simultaneously implement SC and EHVRs, or SC and MHVRs in a GMPE to improve the prediction accuracy of site effects for a target site with available EHVRs or MHVRs information.

¹National Center for Research on Earthquake Engineering, Taipei, Taiwan

²National Central University, Taoyuan, Taiwan

³Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

Corresponding author:

Shu-Hsien Chao, National Center for Research on Earthquake Engineering, Taipei 106, Taiwan.

Email: shchao@ncree.narl.org.tw

Keywords

Horizontal-to-vertical Fourier spectral ratios, spatial correlation, ground-motion prediction equation, site effects, non-ergodic

Date received: 6 April 2020; accepted: 27 July 2020

Introduction

Most ground-motion prediction equations (GMPEs) use average shear-wave velocity to 30 m layers (V_{s30}) and depth to shear-wave velocity to 1 or 2.5 km/s (Z1.0 or Z2.5) as predictor variables to quantify site effects. However, the standard deviation of the station-specific residuals representing the site-to-site variability of ground-motion intensity remains substantial when these GMPEs are used to predict the site effects. For example, the standard deviation of the GMPE developed for Taiwan (Chao et al., 2019a, 2020) ranges from 0.35 to 0.5. Many studies have demonstrated that the variability of ground-motion intensity, which removes site-to-site variability, can be significantly reduced (Al-Atik et al., 2010; Chen and Tsai, 2002; Lin et al., 2011; Morikawa et al., 2008; Rodriguez-Marek et al., 2013). This is one reason that those involved in site-specific projects often evaluate hazard curves for a reference rock site condition using the single-station sigma, which removes site-to-site variability, when conducting site-specific site response analysis to derive the site-specific site amplification regarding reference rock site condition (GeoPentech, 2015; National Center for Research on Earthquake Engineering (NCREE), 2015; Rodriguez-Marek et al., 2014). This indicates that site effect characteristics cannot be fully represented by the site parameters V_{s30} and Z1.0.

Various studies have illustrated that horizontal-to-vertical Fourier spectral ratios (HVRs) evaluated using either strong ground motion induced by earthquake (EHVRs) or ambient ground motion observed using microtremor (MHVRs) can be used to classify the condition of a target site (Kuo et al., 2015; Nakamura, 2019; Yamazaki and Ansary, 1997) and to represent site amplification characteristics (Kuo et al., 2018; Lermo et al., 1993). When the V_{s30} and Z1.0 parameters observed through borehole drilling or noninvasive geophysical measurements are compared, EHVRs and MHVRs are more easily derived. Some researchers have used HVRs to quantify the site effects of observed ground motion (Chao et al., 2019b; Hassani and Atkinson, 2016, 2017a, 2017b; Kwak et al., 2017). In these studies, researchers extracted several predictor variables from the observed HVRs to quantify the site effects of observed spectral acceleration of ground motion (e.g. the peak frequency of the HVRs, HVRs at certain frequencies, and shape of the HVRs), but the reduction of aleatory uncertainty and improvement of prediction accuracy remained limited. Furthermore, the methodology for applying the HVRs in GMPEs for site effect prediction has not been widely discussed.

Numerous studies have revealed the spatial correlation (SC) of ground-motion intensity (Goda and Hong, 2008; Jayaram and Baker, 2009; Kawakami and Sharma, 1999; Sokolov et al., 2010; Wang and Takada, 2005). Several have used the SC of ground-motion intensity to develop nonergodic GMPEs (Kuehn and Abrahamson, 2019; Kuehn and Scherbaum, 2016; Landwehr et al., 2016) and apply them to probabilistic seismic hazard analysis (Abrahamson et al., 2019). They have demonstrated that the aleatory uncertainty of ground motion can be significantly reduced using a nonergodic GMPE, and prediction accuracy can be improved. However, with the application of a nonergodic GMPE, epistemic uncertainty may increase and require additional attention. The idea

underlying the development of these nonergodic GMPEs is that one random variable can be better estimated from the observed values of the other one random variable through a nonparametric modeling approach when these two random variables are correlated. Similar idea was used to develop the conditional mean spectrum (Baker, 2011), in which the known spectral acceleration of a particular period is used to estimate the spectral accelerations of other periods through the correlation between the spectral accelerations of the periods (Baker and Cornell, 2006; Baker and Jayaram, 2008; Jaimes and Candia, 2019; Jayaram et al., 2011).

Because site characteristics can be appropriately represented by the HVRs, in this study, we attempted to apply the HVRs of a wide-frequency range as a vectorized predictor variable in a GMPE to predict site effects using the concept of the correlation, which is similar to the development of a nonergodic GMPE. The EHVRs available at strong motion stations and the MHVRs available at several sites in Taiwan were first compiled. Then, a methodology was proposed to implement the HVRs and SC simultaneously in a GMPE. An EHVRs-SC-based model and a MHVRs-SC-based model were proposed to quantify the site effect of ground motion using the results of a Taiwan GMPE proposed by Chao et al. (2019a, 2020), referred to here as the NCREE19 GMPE and the compiled EHVRs and MHVRs observed at stations and sites in Taiwan. We illustrated the performance of the proposed models by comparing the standard deviation of the station-specific residuals, the spatial distribution of the predicted station terms, and the predicted spectral accelerations of two stations with similar V_{s30} values but different HVRs for a particular ground-motion scenario. The prediction accuracy of the proposed models for stations with inferred V_{s30} and measured V_{s30} was also compared. The results indicate the prediction accuracy of the site effect for a target site with available EHVRs or MHVRs can be substantially improved using the proposed EHVRs-SC-based and MHVRs-SC-based models regardless of how the V_{s30} value of the target site is derived.

Uncertainty was present in the observed EHVRs and MHVRs of each station as well as in other predictors (e.g. magnitude, depth, and V_{s30}). However, the methodology of translating this uncertainty into that of the predicted site terms was not considered in this study. Several studies (Dellaportas and Stephens, 1995; Girard et al., 2003; Iba and Akaho, 2010; Rasmussen and Williams, 2006) may provide guidance in addressing this problem.

HVRs observed in Taiwan

Seismic motion data

EHVRs were evaluated using the seismic motion records of earthquakes collected in strong motion stations in Taiwan (Liu et al., 1999; Shin et al., 2013). The procedure for determining the EHVRs from strong ground-motion records is summarized as follows:

1. Only ground-motion records with a maximum of 3-axis peak ground accelerations (PGA) smaller than 80 gal were selected to avoid the nonlinear site effect influencing the observation of the EHVRs.
2. Only ground-motion records with clear S-wave and P-wave arrival times that were manually picked were selected. A clear P-wave arrival time can help us confirm the existence of the S-wave. Ground-motion records with low signal-to-noise ratios and low data quality were excluded through this approach.

3. The window of ground motion in 3-axes time histories after S-waves arrived for each record was extracted. If the end of the ground-motion window was difficult to identify, the delay time between S-wave and P-wave arrival times multiplied by 3 was the assumed length of the ground-motion window.
4. Fourier amplitudes of 3-axes time histories for each ground-motion window were calculated using 8192-point fast Fourier transform. If the number of data points of a ground-motion window were fewer than 8192, additional data points with values of 0 were added. If a ground-motion window had more than 8192 data points, only the front 8192 data points were used to calculate its Fourier amplitude.
5. The vector sum of Fourier amplitudes in two horizontal components (east–west and north–south directions) were calculated as the horizontal Fourier amplitudes, then the ratio of horizontal and vertical Fourier amplitudes was calculated to be the HVRs for the ground-motion record.
6. The geometric mean of the EHVRs of each ground-motion record collected at the same station was calculated to be the EHVRs of that station. Only stations with at least five available ground-motion records were selected for EHVRs evaluation.

The sampling rate of the ground-motion records that were used to calculate the EHVRs was 200 Hz. Furthermore, 8192 points (40.92 s window length) were used to determine EHVRs because doing so entails using the same window length as the MHVRs, which promotes consistency. The window lengths were shorter than 40.92 s for most of the used seismic motion data. Most energy from the ground motion was contained in the beginning of 40.92 s, even for the used seismic motion data with a longer ground-motion window length. Thus, we use a ground-motion window of 8192 points to determine the EHVRs.

A total of 721 strong motion stations with available EHVRs information are included, and all have Vs30 and Z1.0 information available (Kuo et al., 2017). All EHVRs of each station were resampled from 0.1 to 30 Hz with an equal spacing of 0.02 Hz to improve computational efficiency. The frequency range of 0.1–30 Hz was selected because the variability of the observed EHVRs is stable in this range. As a result, a total of 1496 values of EHVRs were available to act as a vectorized predictor variable for each station. Figure 1a demonstrates the spatial distribution of the 721 stations with available EHVRs. The stations are color-coded based on their Vs30 values. Figure 2a portrays the individual EHVRs of each station and the geometric mean of the EHVRs for different Vs30 bins. The predominant frequency corresponding to the highest geometric mean of EHVRs increases as Vs30 increases (Ghofrani and Atkinson, 2014; Hassani and Atkinson, 2016), and the difference of EHVRs for different Vs30 bins is most significant for frequencies between 0.5 and 5 Hz. Most EHVRs values range from 0.5 to 10.

Ambient motion data

MHVRs are evaluated using the ambient motion data collected through microtremor measurement at 3699 sites in Taiwan (Wen and Huang, 2012). The instruments used to perform the measurements are the three-component velocity-type seismometer VSE-311C (or VSE-315D) and the digital recorder SAMTAC-801B. The measurement duration at each site is approximately 18 min, with a sampling frequency of 200 Hz. The sites are spaced approximately 1–2 km apart in each sub-region in Taiwan. Their locations were all carefully chosen to avoid artificial vibrations and effects from underground construction. In the analyses of the MHVRs, the three-component waveforms of ambient motion data were divided by parting them in moving windows with a length of 8192 points (40.96 s)

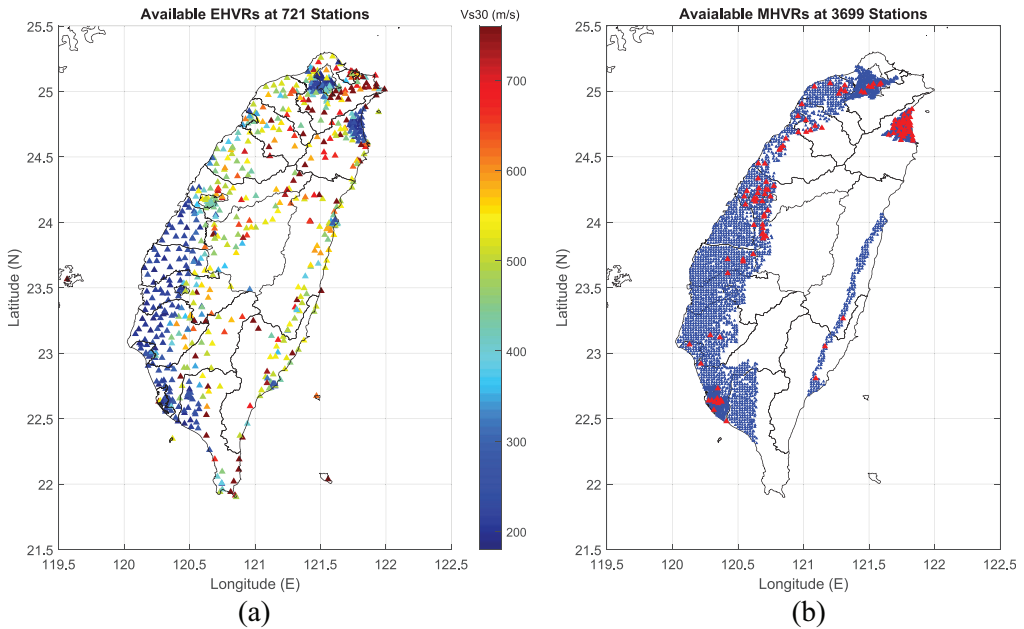


Figure 1. (a) Spatial distribution of stations with available EHVrs. The stations are color-coded based on their Vs30 values. (b) Spatial distribution of sites with available MHVrs. The red markers identify 117 sites which are collocated with stations with available EHVrs.

without overlap. The windows that contained obvious abnormal transient signals that were identified through visual inspection, which may contaminate the results, were eliminated. Fast Fourier transform was conducted for each component to derive Fourier amplitudes for them. One horizontal Fourier amplitude was derived as the root mean square of the Fourier amplitudes of two horizontal components and was divided by the vertical Fourier amplitude to attain the MHVrs of each window. The MHVrs of each station is the arithmetic mean of the MHVrs of at least 20 windows.

A total of 3699 sites in Taiwan with available MHVrs were included in this study, and most of them do not have Vs30 or Z1.0 information available. As with the EHVrs, the MHVrs of each site were resampled from 0.1 to 30 Hz with an equal spacing of 0.02 Hz to improve computational efficiency. The range 0.1–30 Hz is selected because MHVrs variability is stable in this range. A total of 1496 values of MHVrs at different frequency points are available to act as a vectorized predictor variable for each site. Figure 1b shows the spatial distribution of all 3699 sites with available MHVrs. Figure 2a displays the individual MHVrs of each site and geometric mean of MHVrs. The range of MHVrs values is wider than that for EHVrs values. In all, 117 sites have corresponding strong motion stations located within 200 m of the site. We assumed that they were collocated with strong motion stations. It is a reasonable assumption which is necessary for the application of MHVrs in this study. Variability in the MHVrs may be observed in the 200 m range, but we expect this to be non-significant. The locations of these sites are represented in Figure 1b as red markers. Their MHVrs are depicted in Figure 2a as green lines.

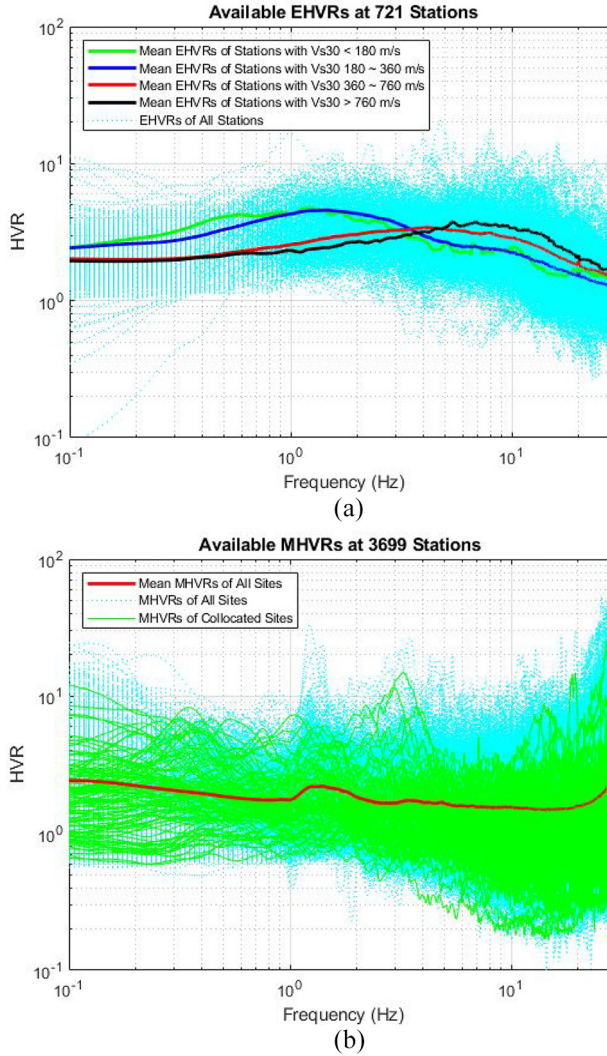


Figure 2. (a) Individual EHVRs of all stations and averaged EHVRs for different V_{s30} bins. (b) Individual MHVRs of all sites and averaged MHVRs.

Site effect of ground motion in Taiwan

The station term derived from a Taiwan GMPE proposed by Chao et al. (2019a, 2020), NCREE19 GMPE, was used to quantify the linear site effect of Taiwan ground motion in this study. This section summarizes the procedure for deriving the station term of each station from regression analysis of the GMPE. The NCREE19 GMPE comprises the predicted reference ground-motion intensity for the reference ground-motion scenario, and the scaling relationships describing the difference in ground-motion intensity between the target and reference scenarios are as follows

$$\ln S_a = \ln S_a^{ref} + S_{source} + S_{path} + S_{site, lin} + S_{site, non} + \delta_e + \delta_s + \delta_r \quad (1)$$

where $\ln S_a$ is the log value of the observed spectral acceleration (in g), $\ln S_a^{ref}$ is the log value of the predicted median reference spectral acceleration for the reference ground-motion scenario (in g), S_{source} is the source scaling that describes the source effect on the median ground-motion intensity, and S_{path} is the path scaling that describes the path effect on the median ground-motion intensity. Furthermore, $S_{site,lin}$ and $S_{site,non}$ are linear and nonlinear site effects on the median ground-motion intensity, respectively, and δ_e , δ_s , and δ_r are event-, station-, and record-specific residuals, respectively.

The source, path scaling, linear and nonlinear site scaling, and model coefficients of the NCREE19 GMPE are determined using the revised two-step maximum-likelihood method (Chao and Chen, 2019; Joyner and Boore, 1993). In the first step, the functional form of the proposed ground-motion model is rearranged as follows

$$\ln S_a = S_{path} + S_{site,non} + E_e + S_s + \delta_r \quad (2)$$

where E_e and S_s are the event term of each event and the station term of each station, respectively. This step can determine the predictions of the event term E_e and station term S_s , model coefficients relating to the path scaling term S_{path} , nonlinear site effect term $S_{site,non}$, and for each record, the record-specific residual together with its standard deviation. In the second step, the predictions of the event and station terms are used to determine other model coefficients as follows

$$E_e = E^{ref} + S_{source} + \delta_e + \delta_{E_e} \quad (3)$$

and

$$S_s = S^{ref} + S_{site,lin} + \delta_s + \delta_{S_s} \quad (4)$$

where E^{ref} and S^{ref} are constants representing the event and station terms for the reference ground-motion scenario, respectively, and δ_{E_e} and δ_{S_s} are the errors of the predicted event and station terms, respectively. In this step, the predictions of the event and station terms for the reference ground-motion scenario (E^{ref} and S^{ref} , respectively) and the model coefficients related to the source scaling, linear site scaling, and event-specific residual (with its standard deviation) of each event as well as the station-specific residual (and its standard deviation) for each station can be determined. Finally, the spectral acceleration of the reference ground-motion scenario can be derived as follows

$$\ln S_a^{ref} = E^{ref} + S^{ref} \quad (5)$$

In this study, we attempted to apply HVRs and SC to improve the prediction of S_s for the linear site effect. Nonlinear site effects were not considered in this study. The linear site effect on ground-motion intensity was predicted using only two predictor variables, V_{s30} (in unit m/s) and $Z_{1.0}$ (in unit m), in the NCREE19 GMPE as follows

$$S_{site,lin} = c_{24} \ln \left(\frac{V_{s30}}{V_{s30}^{ref}} \right) + c_{25} \ln \left(\frac{Z_{1.0}}{Z_{1.0}^{ref}} \right) \quad (6)$$

where

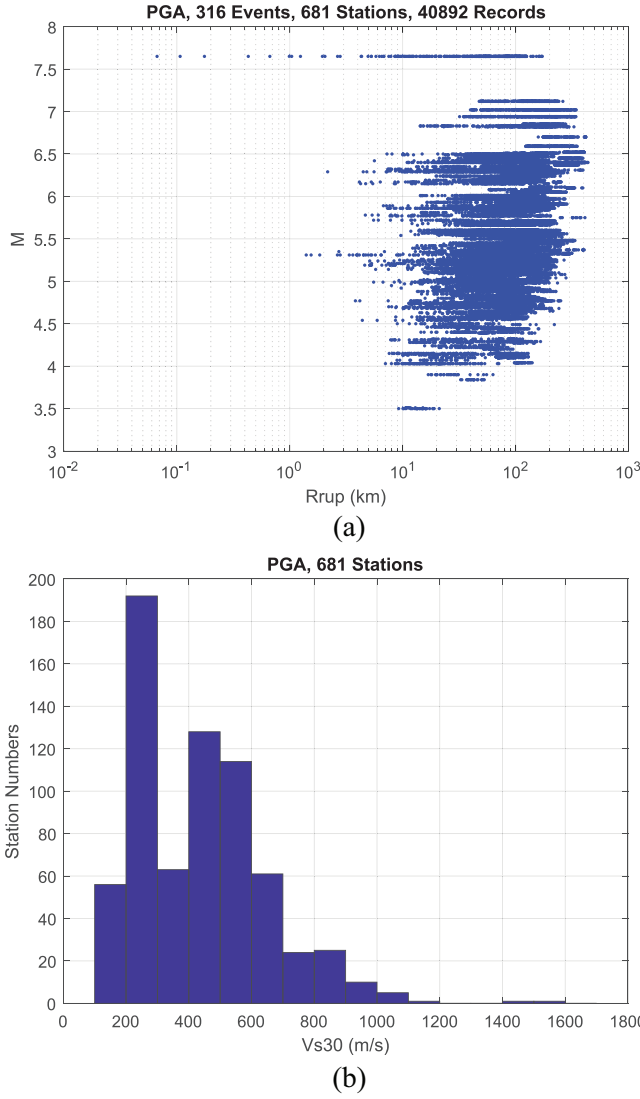


Figure 3. (a) Magnitude–distance pairs of used ground-motion records in NCEE19 GMPE at PGA. (b) Station numbers in different V_{s30} bins used in NCEE19 GMPE at PGA.

$$Z_{1.0}^{ref} = \exp\left(\frac{-4.08}{2} \ln\left(\frac{V_{S30}^2 + 355.4^2}{1750^2 + 355.4^2}\right)\right) \quad (7)$$

where c_{24} and c_{25} are model coefficients relating to linear site effects. The magnitude–distance distribution of the selected data in NCEE19 GMPE at PGA is shown in Figure 3a. Station numbers in different V_{s30} bins used in NCEE19 GMPE at PGA is shown in Figure 3b. The V_{s30} of most stations ranges from 200 to 300 m/s and from 400 to 600 m/s.

In summary, the linear site effect on ground-motion intensity can be quantified using the station term S_s of each station. A higher S_s value represents higher site amplification. The methodology for applying HVRs and SC in a GMPE to improve the prediction of the

station term S_s is introduced in the next section. Some factors that may also influence the linear site amplification such as different earthquake source types, travel paths, angles of incidence, magnitudes, and distances of scenarios (Stafford et al., 2017), have not been considered in evaluations of the station terms S_s . The proposed methodology and models may derive a biased prediction, while the dependence of the site amplification on these factors is significant for a target site.

Methodology for applying HVRs and SC

This section describes the proposed methodology for implementing the HVRs and SC to predict the station terms during the development of the GMPE. The proposed methodology was developed based on the framework of two-step maximum-likelihood regression analysis (Chao and Chen, 2019; Joyner and Boore, 1993). All symbols used in this study follow those used in the study of Chao and Chen.

In step 2 of the revised two-step maximum-likelihood method, the station term vector S_s , comprising the station term of each station, and its hessian matrix H_{S_s} , obtained in step 1, are used to solve model coefficients relating to linear site scaling. We improved the methodology by implementing the HVRs and SC to predict the station term vector. The model equation in step 2 (Chao and Chen, 2019) can be expressed as

$$S_s = X_s c_s + \delta_{s2} = X_s c_s + \delta_s + \delta_{S_s} \quad (8)$$

where S_s is a vector comprising the station term of each station, X_s is a matrix corresponding to the station-specific term of the median model, c_s is a vector comprising model coefficients of the station-specific terms of the median model related to linear site effect as S^{ref} , c_{24} , and c_{25} in the NCREE19 GMPE, and δ_{s2} is the residual term vector involving station-specific residual vector δ_s , comprising the station-specific residual of each station as well as the prediction error of station term vector δ_{S_s} derived in step 1, which comprises the prediction error of the station term of each station. The covariance matrix of the residual term δ_{s2} can be expressed as

$$V_{s2} = V_s + H_{S_s}^{-1} \quad (9)$$

where V_s is the covariance matrix of δ_s , and H_{S_s} is the hessian matrix of S_s which is equal to the hessian matrix of δ_{S_s} (Joyner and Boore, 1993); V_{s2} are full-matrices representing that the residual terms of each station are correlated with each other. As a result, the least-squares error method cannot be used to derive accurate results. In this case, the maximum-likelihood method should be used to derive unbiased results. If we assume that all residual terms are zero-mean normal random variables, the likelihood function of the equation can be expressed as

$$\ln L_s = -0.5n_s \ln(2\pi) - 0.5 \ln(|V_{s2}|) - 0.5(S_s - X_s c_s)^T H_{s2} (S_s - X_s c_s) \quad (10)$$

where L_s is the likelihood value of the equation, and the hessian matrix H_{s2} can be expressed as

$$H_{s2} = H_{S_s} - (I + H_{S_s} V_s)^{-1} H_{S_s} V_s H_{S_s} \quad (11)$$

where \mathbf{I} is the identity matrix. The log values of $|\mathbf{V}_{s2}|$ can be calculated as

$$\begin{aligned}\ln(|\mathbf{V}_{s2}|) &= \ln\left(\left|\left(\mathbf{V}_s + \mathbf{H}_{S_s}^{-1}\right)\right|\right) = \ln\left(\left|\mathbf{H}_{S_s}^{-1}(\mathbf{H}_{S_s} \mathbf{V}_s + \mathbf{I})\right|\right) \\ &= -\ln(|\mathbf{H}_{S_s}|) + \ln(|\mathbf{H}_{S_s} \mathbf{V}_s + \mathbf{I}|)\end{aligned}\quad (12)$$

The model coefficients related to linear site effects can be derived such that they maximize the likelihood values with known matrix \mathbf{S}_s and \mathbf{H}_{S_s} derived in step 1. If a large matrix must be inverted when data sets are large, some studies (Bauer et al., 2016; Rasmussen and Williams, 2006) have provided guidance regarding how to do so.

In the NCREE19 GMPE, the station-specific residuals of each station are assumed to be uncorrelated and have a diagonal covariance matrix. In this study, both the HVRs and SC were implemented to predict the station term by revising the covariance matrix of the station-specific residual. We illustrated this by developing four different models with four covariance matrices as follows

$$\text{Model 0 : } \mathbf{V}_s = \phi_{s2s}^2 \mathbf{I} \quad (13)$$

$$\text{Model 1 : } \mathbf{V}_s = \phi_{f1}^2 \exp(-a_1^2 \mathbf{\Delta}_{HVR}) + \phi_{n1}^2 \mathbf{I} \quad (14)$$

$$\text{Model 2 : } \mathbf{V}_s = \phi_{f2}^2 \exp(-a_2^2 \mathbf{\Delta}_t) + \phi_{n2}^2 \mathbf{I} \quad (15)$$

$$\text{Model 3 : } \mathbf{V}_s = \alpha_1 \phi_{f1}^2 \exp(-a_1^2 \mathbf{\Delta}_{HVR}) + \alpha_2 \phi_{f2}^2 \exp(-a_2^2 \mathbf{\Delta}_t) + \phi_{n3}^2 \mathbf{I} \quad (16)$$

with the following constraints

$$\phi_{s2s}^2 = \phi_{f1}^2 + \phi_{n1}^2 = \phi_{f2}^2 + \phi_{n2}^2 = \alpha_1 \phi_{f1}^2 + \alpha_2 \phi_{f2}^2 + \phi_{n3}^2 \quad (17)$$

The setup of Model 0 is identical to that of NCREE19 GMPE and assumes that the station-specific residuals of each station are uncorrelated with themselves. Model 1 assumes that the station terms between two stations are highly correlated when their HVRs are similar. Model 2 assumes that the station terms of two stations are highly correlated when they are nearby. Model 3 assumes that the correlation of the station terms between two stations depends not only on their spatial distance but also the similarity of their HVRs. Model coefficients ϕ_{f1} and a_1 quantify the correlation as a function of HVRs difference between two stations; model coefficients ϕ_{f2} and a_2 quantify the SC as a function of the distance between two stations; α_1 and α_2 quantify the contribution of the correlation matrices from the HVRs and SC; model coefficients $\phi_{s2s,0}$, ϕ_{n1} , ϕ_{n2} , and ϕ_{n3} are the standard deviations for the residuals uncorrelated with each other which are contributed by noise; $\mathbf{\Delta}_t$ is a matrix describing the spatial distance between each station, and its element at i th row and j th column represents the spatial distance between i th and j th stations in kilometers; $\mathbf{\Delta}_{HVR}$ is a matrix describing the difference of HVRs between each station, and its element can be expressed as

$$\Delta_{HVR,ij} = \sqrt{\frac{\sum_{k=1}^{n_f} \{\ln(HVR_{i,k}) - \ln(HVR_{j,k})\}^2}{n_f}} \quad (18)$$

where $\Delta_{HVR,ij}$ is the element of matrix $\mathbf{\Delta}_{HVR}$ at i th row and j th column describing the difference of the HVRs between the i th and j th stations; n_f is the total number of frequency points of the HVRs; $HVR_{i,k}$ and $HVR_{j,k}$ are HVR at k th frequency at i th and j th stations, respectively. In the next section, we illustrate the correlation of station terms with respect to spatial distance and similarity of HVRs using semivariograms. The constraints were defined to ensure the consistency of total site-to-site variability between each model.

The methodology for simultaneously implementing the HVRs and SC in a GMPE has been introduced. All model coefficients can be determined through finding their values which maximize the likelihood function with the constraints. The proposed methodology can account for the difference of the prediction errors of the station term \mathbf{S}_s among stations due to different numbers of available ground-motion records among stations being used in the regression analysis. When applying the developed models to conduct ground-motion prediction for a target site, the following equation (Rasmussen and Williams, 2006) can be used to derive the predicted station term \mathbf{S}_s^* for the site

$$\mathbf{S}_s^* = \mathbf{S}_{s,0}^* + \mathbf{S}_{s,s}^* \quad (19)$$

where

$$\mathbf{S}_{s,0}^* = \mathbf{X}_S^* \mathbf{c}_s \quad (20)$$

$$\mathbf{S}_{s,s}^* = \mathbf{V}_s^* \mathbf{H}_{s2} (\mathbf{S}_s - \mathbf{X}_S \mathbf{c}_s) \quad (21)$$

and the prediction error of predicted station term $\mathbf{S}_{s,s}^*$ can be derived from its covariance matrix as follows (Rasmussen and Williams, 2006)

$$\mathbf{V}_{S_{s,s}^*} = \mathbf{V}_s^{**} - \mathbf{V}_s^* \mathbf{H}_{s2} \mathbf{V}_s^{*T} \quad (22)$$

where \mathbf{S}_s^* is a vector involving the predicted station term of each specific target site; $\mathbf{S}_{s,0}^*$ is the station term predicted from the vector \mathbf{c}_s , which is equal to the predicted station term of Model 0 (NCREE19 GMPE); $\mathbf{S}_{s,s}^*$ is the predicted station term evaluated from the similarity of the observed HVRs and the locations of the target sites to the HVRs and locations of the stations we observed and used to develop the model; \mathbf{X}_S^* is a matrix corresponding to the station-specific term of the median model for each target site; \mathbf{V}_s^* is the covariance matrix between the target sites and the stations with observed station terms used in the regression analysis; and \mathbf{V}_s^{**} is the covariance matrix between the target sites. The symbol “*” is employed to distinguish the target sites and the stations with station terms used in the regression analysis. The covariance matrices \mathbf{V}_s^* and \mathbf{V}_s^{**} of Model 1, Model 2, and Model 3 can be expressed as

$$\begin{aligned} \text{Model 1 : } \mathbf{V}_s^* &= \phi_{f1}^2 \exp(-a_1^2 \delta_{HVR}^*) \\ \mathbf{V}_s^{**} &= \phi_{f1}^2 \exp(-a_1^2 \mathbf{\Delta}_{HVR}^{**}) + \phi_{n1}^2 \mathbf{I} \end{aligned} \quad (23)$$

$$\begin{aligned} \text{Model 2 : } \mathbf{V}_s^* &= \phi_{f2}^2 \exp(-a_2^2 \mathbf{\Delta}_{t_s}^*) \\ \mathbf{V}_s^{**} &= \phi_{f2}^2 \exp(-a_2^2 \mathbf{\Delta}_{t_s}^{**}) + \phi_{n2}^2 \mathbf{I} \end{aligned} \quad (24)$$

$$\begin{aligned} \text{Model 3 : } \mathbf{V}_s^* &= \alpha_1 \phi_{f_1}^2 \exp(-a_1^2 \boldsymbol{\Delta}_{HVR}^*) + \alpha_2 \phi_{f_2}^2 \exp(-a_2^2 \boldsymbol{\Delta}_{t_s}^*) \\ \mathbf{V}_s^{**} &= \alpha_1 \phi_{f_1}^2 \exp(-a_1^2 \boldsymbol{\Delta}_{HVR}^{**}) + \alpha_2 \phi_{f_2}^2 \exp(-a_2^2 \boldsymbol{\Delta}_{t_s}^{**}) + \phi_{n_3}^2 \mathbf{I} \end{aligned} \quad (25)$$

where $\boldsymbol{\Delta}_{t_s}^*$ is a matrix describing the spatial distance between target sites and stations with observed station terms used in the regression analysis, and its element at i th row and j th column represents the spatial distance in kilometers between the i th target site and j th station with observed station terms used in the regression analysis; $\boldsymbol{\Delta}_{t_s}^{**}$ is a matrix describing the spatial distance between target sites; $\boldsymbol{\Delta}_{HVR}^*$ is a matrix describing the difference of HVRs between target sites and stations with station terms used in the regression analysis; and $\boldsymbol{\Delta}_{HVR}^{**}$ is a matrix describing the difference of HVRs between target sites. Elements in $\boldsymbol{\Delta}_{HVR}^*$ and $\boldsymbol{\Delta}_{HVR}^{**}$ can be expressed as

$$\Delta_{HVR,ij}^* = \sqrt{\frac{\sum_{k=1}^{n_f} \left\{ \ln(HVR_{i,k}^*) - \ln(HVR_{j,k}) \right\}^2}{n_f}} \quad (26)$$

$$\Delta_{HVR,ij}^{**} = \sqrt{\frac{\sum_{k=1}^{n_f} \left\{ \ln(HVR_{i,k}^*) - \ln(HVR_{j,k}^*) \right\}^2}{n_f}} \quad (27)$$

where $\Delta_{HVR,ij}^*$ is the element of matrix $\boldsymbol{\Delta}_{HVR}^*$ at i th row and j th column describing the difference of the HVRs between the i th target site and j th station used in the regression analysis; $\Delta_{HVR,ij}^{**}$ is the element of matrix $\boldsymbol{\Delta}_{HVR}^{**}$ at i th row and j th column describing the difference of the HVRs between the i th target site and j th target site; $HVR_{i,k}^*$ is HVR at k th frequency of the i th target site. In this study, we just selected the exponential covariance function which has mainly been used in the past studies to describe the SC of the ground motion. It is a stationary and isotropic covariance function. Improvement of the covariance function may improve the predicted station terms.

A key challenge of developing a GMPE is controlling the model behavior and constraining the prediction result for a ground-motion scenario with no available or insufficient ground-motion data. One advantage of the proposed methodology is that the model behavior and prediction are well constrained when the methodology is employed for the prediction of the station terms with the HVRs and SC. When the observed HVRs of a target site markedly differ from the HVRs previously observed and used to develop the model, or the location of a target site is distant from the stations used to develop the model, the matrix \mathbf{V}_s^* will be almost equal to a zero matrix with all elements equal to zero, and matrix \mathbf{V}_s^{**} will be almost equal to the matrix \mathbf{V}_s of Model 0. In that case, the predictions of Model 1, Model 2, and Model 3 will be close to the prediction of Model 0 with the prediction error as original site-to-site variability. This means that the observed HVRs and the location of the target site are entirely new, and the prediction accuracy of its station terms cannot benefit from past observations.

Illustration of applying HVRs and SC

This section illustrates how Vs30 and Z1.0 with HVRs and SC are implemented to predict site effects. The methodology proposed in the previous section was used to develop four models with EHVRs from seismic data and four other models with MHVRs from ambient

data. The standard deviations of the station-specific residuals were compared to evaluate the performance of each model. The results are individually introduced in this section.

Seismic motion data

The stations that have available station terms from NCREE19 GMPE and available EHVRs were used to illustrate the proposed methodology. The number of stations used for each period of spectral acceleration is presented in Table 2. Approximately 620 stations have both available EHVRs and station terms from NCREE19 GMPE. The correlations of the station terms with regard to the distances Δ_{HVR} and Δ_{t_s} are illustrated first using semivariograms, which are measurements of the average similarity of data for different distance bins (Goovaerts, 1997). The y-axes of semivariograms for distances Δ_{HVR} and Δ_{t_s} are calculated as

$$\gamma(\Delta_{HVR}) = \frac{1}{2N(\Delta_{HVR})} \sum_{i=1}^{N(\Delta_{HVR})} \Delta\delta_{s2,i}^2 \quad (28)$$

and

$$\gamma(\Delta_{t_s}) = \frac{1}{2N(\Delta_{t_s})} \sum_{i=1}^{N(\Delta_{t_s})} \Delta\delta_{s2,i}^2 \quad (29)$$

where $N(\Delta_{HVR})$ and $N(\Delta_{t_s})$ are the number of station pairs with distances between $\Delta_{HVR} - d\Delta_{HVR} < \Delta_{HVR} \leq \Delta_{HVR} + d\Delta_{HVR}$ and $\Delta_{t_s} - d\Delta_{t_s} < \Delta_{t_s} \leq \Delta_{t_s} + d\Delta_{t_s}$, respectively; $\Delta\delta_{s2,i}$ are the differences between the station-specific residuals of the i th station pair. The derived semivariograms with respect to the distances are shown in Figure 4a and b. Results of different station pairs within different distance ranges are also shown. The station-specific residuals derived from NCREE19 GMPE (Model 0) are used for these plots, and the values of $d\Delta_{HVR}$ and $d\Delta_{t_s}$ are set as 0.1 and 1 km, respectively, for these plots. The values of γ are higher for larger Δ_{t_s} or Δ_{HVR} values, as these plots indicate. This illustrates that the station terms between two stations with shorter Δ_{t_s} and Δ_{HVR} are more similar to each other, providing a basis for the assertion that better prediction can be derived by considering Δ_{t_s} or Δ_{HVR} simultaneously, as in Model 3.

The model coefficients of each model are determined in a stepwise manner. Model 0 is developed first, for which the station terms are predicted using only Vs30 and Z1.0 as in NCREE19 GMPE. The model coefficients in vector c_s and standard deviation ϕ_{s2s} which maximizes the likelihood function are derived. They are identical to the results of NCREE19 GMPE. To simplify the analysis procedure and improve the stability of the regression analysis result, the model coefficients in vector c_s of Model 0 are used for Models 1, 2, and 3. Next, model coefficients ϕ_{f1} , a_1 , ϕ_{f2} , a_2 , ϕ_{n1} , and ϕ_{n2} , which maximize the likelihood functions with constraints, were derived for Models 1 and 2 individually. Again, model coefficients ϕ_{f1} , a_1 , ϕ_{f2} , and a_2 are used to develop Model 3. The model coefficients α_1 and α_2 and ϕ_{n3} , which maximize the likelihood function with the constraints, are derived for Model 3. Two constraints were used in this study to maximize the likelihood function. One constraint is displayed in Equation 17. This constraint is used to ensure that the site-to-site variability of different models is the same. The second constraint is that the values of ϕ_{n1}^2 , ϕ_{n2}^2 , and ϕ_{n3}^2 should be larger than 0.004. This constraint

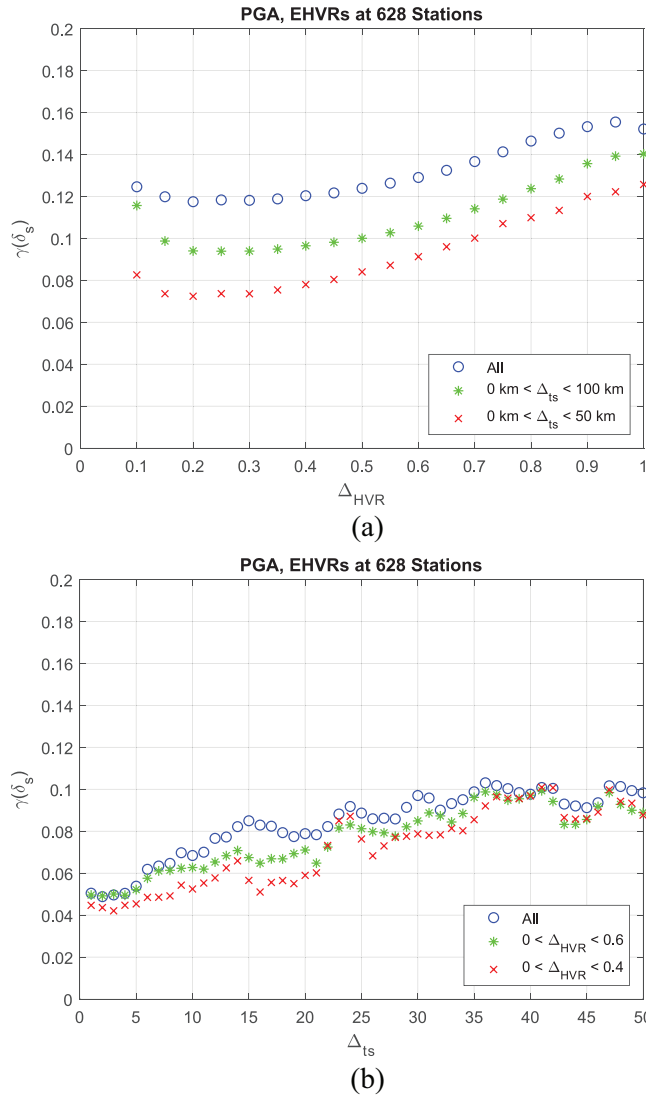


Figure 4. (a) Semivariogram with respect to distance Δ_{HVR} for the data sets with different Δ_{ts} bins for PGA. (b) Semivariogram with respect to distance Δ_{ts} for the data sets with different Δ_{HVR} bins for PGA.

is used to avoid overfitting the data with no-existence or a very low level of noise. The MATLAB command `fmincon` is used to derive the model coefficients which maximizing the likelihood function with constraints. The derived model coefficients of Models 1, 2, and 3, which maximize the likelihood with constraints, are presented in Table 1. Figure 5 illustrates the dependence of the station terms on two distances from the proposed models. Figure 5a presents the correlation coefficient of the proposed Models 1 and 3 with respect to distance Δ_{HVR} for PGA. Figure 5b shows the correlation coefficient of the proposed Models 2 and 3 with respect to distance Δ_{ts} for PGA. Basically, Model 3 provides a more refined description of the dependence of the station terms than Models 1 and 2.

After we derived the model coefficients of the four models, we calculated the predicted station terms and the station-specific residuals for the stations used to develop the models.

Table 1. Model coefficients of EHVRs-SC-based model and MHVRs-SC-based model

T (s)	α_1	α_2	α_1	α_2	ϕ_{f1}	ϕ_{f2}	ϕ_{n1}	ϕ_{n2}	ϕ_{n3}
PGA (g)	1.8339	0.2080	0.7405	0.9049	0.2105	0.2910	0.2716	0.1827	0.0928
0.01	1.8356	0.2080	0.7374	0.9055	0.2109	0.2911	0.2715	0.1828	0.0930
0.02	1.8412	0.2091	0.7261	0.9061	0.2134	0.2939	0.2716	0.1814	0.0893
0.03	1.7693	0.2069	0.7227	0.8979	0.2189	0.3028	0.2791	0.1847	0.0940
0.05	1.6020	0.1981	0.7060	0.8652	0.2597	0.3424	0.3013	0.2026	0.0960
0.075	1.5809	0.1893	0.7199	0.8267	0.3282	0.3989	0.3290	0.2384	0.0829
0.1	1.6499	0.1848	0.7790	0.8339	0.3384	0.4009	0.3370	0.2596	0.0696
0.15	1.9278	0.1760	0.7473	0.8989	0.3146	0.3697	0.3245	0.2600	0.0863
0.2	2.0594	0.2064	0.8327	0.9480	0.2569	0.3272	0.3211	0.2492	0.1127
0.25	2.0370	0.2068	0.9188	0.9510	0.2379	0.2805	0.2981	0.2583	0.1364
0.3	2.1707	0.2210	0.9583	0.9381	0.2447	0.2679	0.2683	0.2452	0.0848
0.4	2.3085	0.2374	0.7007	0.9641	0.2832	0.2471	0.1972	0.2408	0.0633
0.5	2.2881	0.2411	0.7001	0.9246	0.2853	0.2375	0.1781	0.2382	0.0632
0.75	2.1978	0.2411	0.7447	0.8396	0.2936	0.2362	0.1698	0.2434	0.0632
1	2.2855	0.2414	0.6102	0.8070	0.3278	0.2438	0.1004	0.2411	0.0633
1.5	2.1775	0.2531	0.5535	0.7597	0.3467	0.2803	0.1000	0.2273	0.0633
2	2.1433	0.2462	0.4832	0.7964	0.3501	0.3048	0.1211	0.2105	0.0633
3	2.2248	0.2271	0.4166	0.8709	0.3288	0.3142	0.1721	0.1976	0.0820
4	2.1202	0.2065	0.2818	0.9292	0.3016	0.3309	0.2307	0.1863	0.1297
5	2.1795	0.1880	0.1804	0.9614	0.3061	0.3483	0.2391	0.1720	0.1317
PGV (cm/s)	2.2945	0.2092	0.6219	0.9137	0.1883	0.2108	0.1991	0.1751	0.1116
PGD (cm)	2.0297	0.1980	0.4639	0.8764	0.2426	0.2651	0.2132	0.1845	0.1242

PGA: peak ground acceleration; PGD: peak ground displacement; PGV: peak ground velocity.

A comparison of the standard deviations of the station-specific residuals of the models is presented in Table 2 and Figure 6a. The site-to-site variability can be reduced dramatically, by 65%–90%, for various spectral periods when Vs30 and Z1.0 are implemented with the EHVRs and SC simultaneously. This means that implementing Vs30 and Z1.0 with the EHVRs and SC simultaneously can considerably improve the prediction accuracy of the linear site effect. Figure 7a to d presents a comparison of the spatial distribution of the station-specific residuals for each model. Significant regional differences can be observed in the spatial distribution of the station-specific residuals for Model 0. This illustrates how using SC can improve prediction accuracy. Implementing EHVRs can also improve the prediction accuracy of the station term for local stations, but regional differences and SCs remain. The best prediction accuracy can be achieved by implementing EHVRs and SC simultaneously. The measured EHVRs at new target sites can be directly used to predict the station terms of the new sites using the proposed model.

Ambient motion data

For ground-motion prediction, one challenge is deriving the accurate prediction of the station term for a target site without any available seismic data. In the previous section, we illustrate that implementing Vs30 and Z1.0 with the SC and EHVRs simultaneously can improve the accuracy of the predicted station terms. However, in a real situation, a target site may have no strong motion instrument or ground-motion records with which to evaluate the EHVRs. In this case, a possible alternative is to evaluate the MHVRs from the ambient motion data of microtremors. Numerous studies have demonstrated that both the EHVRs and the MHVRs can be used for site classification; however, for the same station, EHVRs and MHVRs can differ.

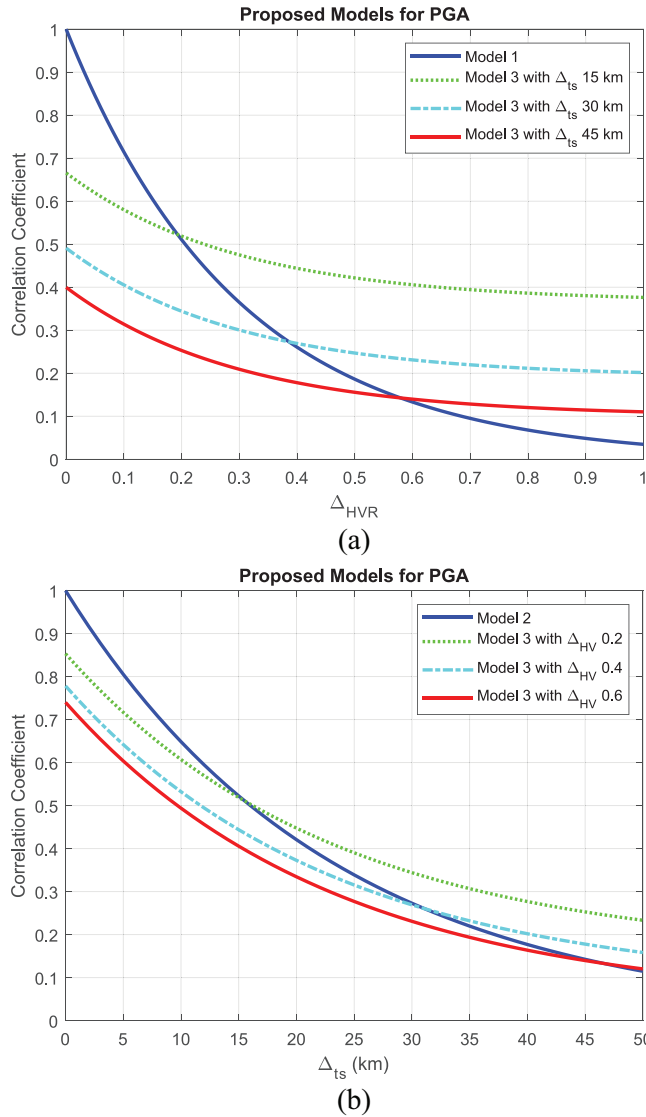


Figure 5. (a) Correlation coefficient of the proposed Models 1 and 3 with respect to distance Δ_{HVR} plot for PGA and (b) correlation coefficient of the proposed Models 2 and 3 with respect to distance Δ_{ts} for PGA.

Kawase et al. (2018) determined that MHVRs and EHVRs share similarities, especially from low frequency until their first peak frequency but exhibit significant differences in the higher-frequency range. They argued that this is because microtremors mainly consist of surface waves, so that peaks associated with higher modes would not be prominent, whereas seismic motions mainly consist of upwardly propagating body waves, so that higher mode resonances can be observed in high-frequency domain (Kawase et al., 2018). They also developed empirical amplitude ratios between MHVRs and EHVRs. These ratios can be used to infer the EHVRs from the MHVRs for a target site. However, the uncertainty of these empirical amplitude ratios is substantial and should be considered when implementing EHVRs transferred from MHVRs to predict the station terms.

Table 2. List of the number of stations and standard deviations of the station-specific residuals of each model when implementing SC and EHRs, or SC and MHVRs

T (s)	EHRs			MHVRs						
	Stations	Model no.	Stations	Model no.	Stations	Model no.				
	0	1	2	3	0	1	2	3		
PGA (g)	628	0.3570	0.2532	0.1619	0.0656	112	0.3100	0.1986	0.2929	0.0466
0.01	628	0.3572	0.2530	0.1621	0.0657	112	0.3098	0.1982	0.2928	0.0466
0.02	628	0.3589	0.2524	0.1603	0.0629	112	0.3127	0.1984	0.2947	0.0439
0.03	628	0.3677	0.2605	0.1629	0.0663	112	0.3207	0.2035	0.3010	0.0459
0.05	628	0.4079	0.2776	0.1769	0.0650	112	0.3503	0.2061	0.3224	0.0424
0.075	628	0.4750	0.2911	0.2068	0.0499	112	0.4058	0.2083	0.3615	0.0287
0.1	628	0.4900	0.2946	0.2273	0.0407	112	0.4175	0.2127	0.3728	0.0239
0.15	628	0.4647	0.2768	0.2330	0.0488	112	0.4053	0.2138	0.3704	0.0337
0.2	628	0.4244	0.2852	0.2210	0.0679	112	0.3808	0.2370	0.3571	0.0517
0.25	628	0.3944	0.2674	0.2372	0.0898	112	0.3589	0.2230	0.3454	0.0699
0.3	628	0.3746	0.2306	0.2233	0.0509	112	0.3418	0.1927	0.3302	0.0391
0.4	628	0.3572	0.1441	0.2201	0.0382	112	0.3217	0.1144	0.3142	0.0327
0.5	628	0.3490	0.1266	0.2199	0.0397	112	0.3008	0.0942	0.2946	0.0330
0.75	628	0.3501	0.1190	0.2267	0.0400	112	0.2536	0.0727	0.2487	0.0279
1	627	0.3493	0.0551	0.2223	0.0368	112	0.2313	0.0313	0.2263	0.0241
1.5	625	0.3616	0.0537	0.2005	0.0359	112	0.2693	0.0333	0.2599	0.0268
2	622	0.3692	0.0677	0.1809	0.0369	112	0.2887	0.0443	0.2746	0.0295
3	617	0.3719	0.1116	0.1703	0.0503	111	0.3233	0.0819	0.3044	0.0401
4	607	0.3801	0.1773	0.1632	0.0949	111	0.3636	0.1443	0.3357	0.0775
5	590	0.3899	0.1827	0.1506	0.1012	108	0.4101	0.1673	0.3713	0.0872
PGV (cm/s)	628	0.2840	0.1731	0.1653	0.0853	112	0.2242	0.1251	0.2192	0.0619
PGD (cm)	628	0.3268	0.1806	0.1734	0.0987	112	0.2619	0.1240	0.2508	0.0653

PGA: peak ground acceleration; PGD: peak ground displacement; PGV: peak ground velocity.

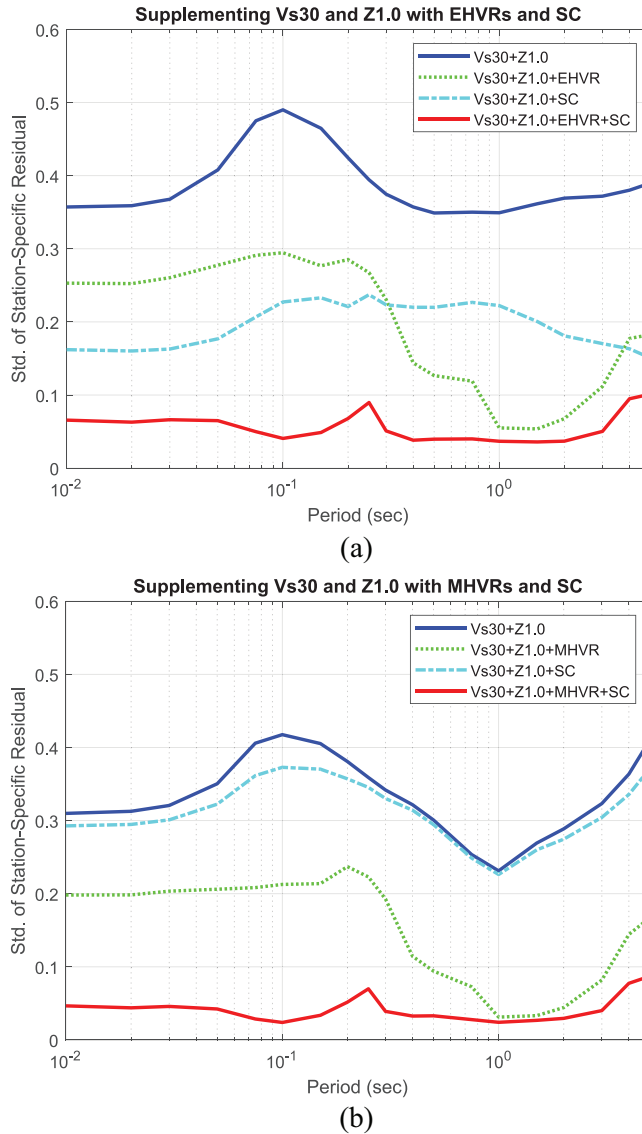


Figure 6. Standard deviations of station-specific residuals before and after supplementing Vs30 and Z1.0 with SC and (a) EHVRs and (b) MHVRs.

In this study, we sought to implement Vs30 and Z1.0 with the SC and MHVRs directly to predict the station terms. Stations that have available station terms from the NCREE19 GMPE and available MHVRs collocated with strong motion stations were used to illustrate the proposed methodology. The number of stations used in this study for each period of spectral acceleration is presented in Table 2. Approximately 112 sites with available MHVRs and available station terms from the NCREE19 GMPE were included. Due to the similarity between the EHVRs and MHVRs at the same site, we assumed that the similarity of the station terms with identical Δ_{HVR} values obtained from the EHVRs and MHVRs were identical. As a result, model coefficients derived from the EHVRs (as can be observed in Table 1) were used directly for applying the MHVRs to predict the station

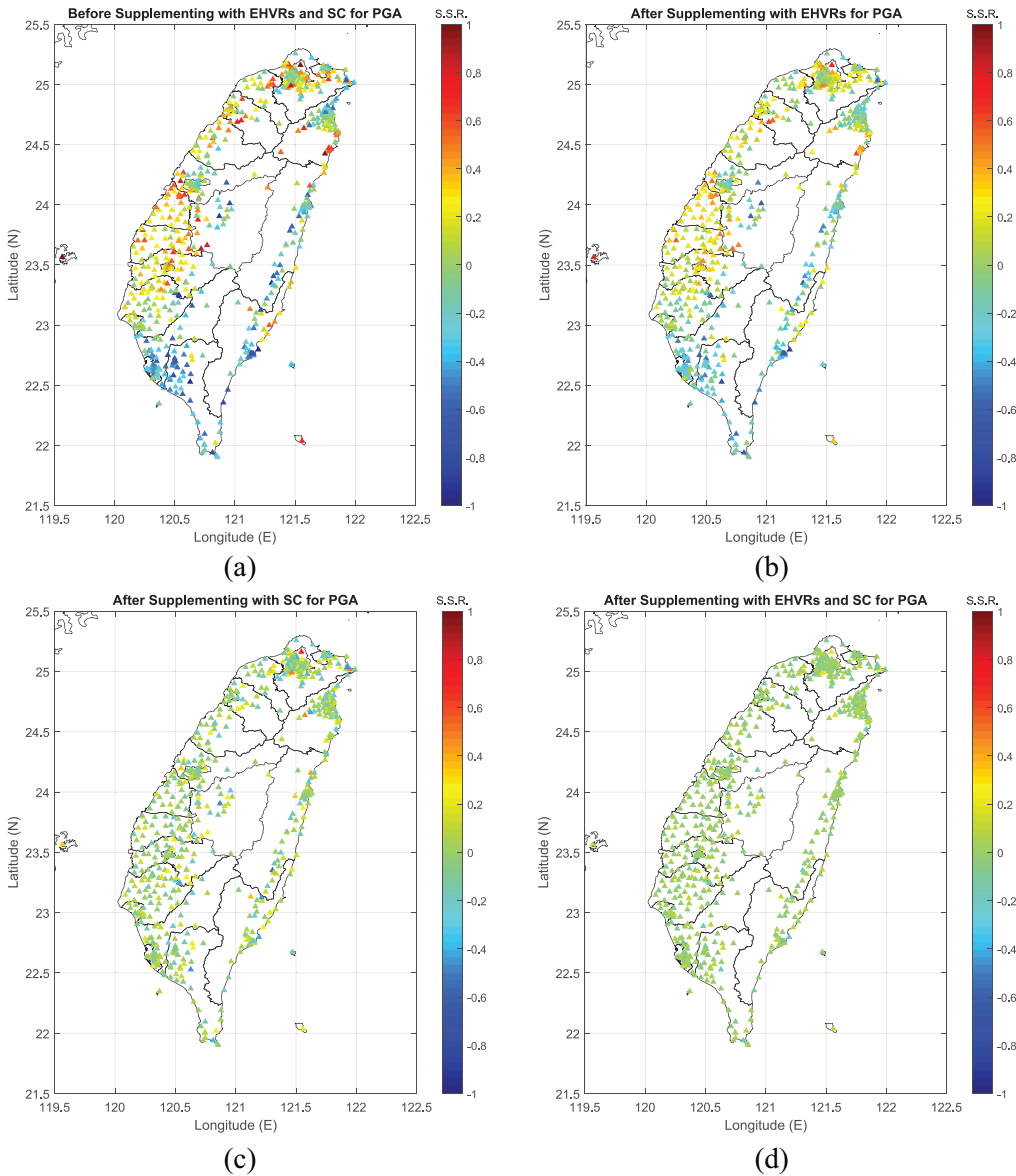


Figure 7. Spatial distribution of the station-specific residuals for PGA (a) before supplementation with EHVrs and SC (Model 0), (b) after supplementation with EHVrs (Model 1), (c) after supplementation with SC (Model 2), and (d) after supplementation with EHVrs and SC (Model 3).

terms. This assumption is a compromise because deriving an accurate model using only MHVRs is difficult because of its current lower station numbers and sparse distribution. We suggest developing a new model for MHVRs individually for the forward prediction of ground motion when more MHVRs measurements that are collocated with strong motion stations are available in the future.

We calculated the predicted station term and the station-specific residual for the stations that had available station terms from the NCREE19 GMPE and available MHVRs

Table 3. List of the standard errors of the estimated standard deviations of the station-specific residuals of each model when implementing SC and EHVRs, or SC and MHVRs

T (s)	EHVRs				MHVRs			
	Model no.				Model no.			
	0	1	2	3	0	1	2	3
PGA (g)	0.0101	0.0072	0.0046	0.0019	0.0208	0.0133	0.0197	0.0031
0.01	0.0101	0.0071	0.0046	0.0019	0.0208	0.0133	0.0197	0.0031
0.02	0.0101	0.0071	0.0045	0.0018	0.0210	0.0133	0.0198	0.0029
0.03	0.0104	0.0074	0.0046	0.0019	0.0215	0.0137	0.0202	0.0031
0.05	0.0115	0.0078	0.0050	0.0018	0.0235	0.0138	0.0216	0.0028
0.075	0.0134	0.0082	0.0058	0.0014	0.0272	0.0140	0.0243	0.0019
0.1	0.0138	0.0083	0.0064	0.0011	0.0280	0.0143	0.0250	0.0016
0.15	0.0131	0.0078	0.0066	0.0014	0.0272	0.0143	0.0249	0.0023
0.2	0.0120	0.0081	0.0062	0.0019	0.0256	0.0159	0.0240	0.0035
0.25	0.0111	0.0076	0.0067	0.0025	0.0241	0.0150	0.0232	0.0047
0.3	0.0106	0.0065	0.0063	0.0014	0.0229	0.0129	0.0222	0.0026
0.4	0.0101	0.0041	0.0062	0.0011	0.0216	0.0077	0.0211	0.0022
0.5	0.0099	0.0036	0.0062	0.0011	0.0202	0.0063	0.0198	0.0022
0.75	0.0099	0.0034	0.0064	0.0011	0.0170	0.0049	0.0167	0.0019
1	0.0099	0.0016	0.0063	0.0010	0.0155	0.0021	0.0152	0.0016
1.5	0.0102	0.0015	0.0057	0.0010	0.0181	0.0022	0.0174	0.0018
2	0.0105	0.0019	0.0051	0.0010	0.0194	0.0030	0.0184	0.0020
3	0.0106	0.0032	0.0049	0.0014	0.0218	0.0055	0.0205	0.0027
4	0.0109	0.0051	0.0047	0.0027	0.0245	0.0097	0.0226	0.0052
5	0.0114	0.0053	0.0044	0.0029	0.0280	0.0114	0.0254	0.0060
PGV (cm/s)	0.0080	0.0049	0.0047	0.0024	0.0151	0.0084	0.0147	0.0042
PGD (cm)	0.0092	0.0051	0.0049	0.0028	0.0176	0.0083	0.0168	0.0044

PGA: peak ground acceleration; PGD: peak ground displacement; PGV: peak ground velocity.

using the model coefficient (see Table 1). A comparison of the standard deviations of the station-specific residuals of the four models is presented in Table 2 and Figure 6b. Figure 8a to d presents a comparison of the station-specific residuals of each model for PGA. The standard deviations of the station-specific residuals can also be reduced dramatically, by 70%–95%, for various spectral periods when Vs30 and Z1.0 are implemented with the MHVRs and SC simultaneously, even though only approximately 110 sites with available MHVRs and station terms from the NCREE19 GMPE were used. The model implementing the MHVRs has better prediction accuracy than that implementing only SC. This result may be explained by the number and density of the stations with available MHVRs and station terms from the NCREE19 GMPE as well as the SC not being as high as in the case of EHVRs, as detailed in the previous section. This also illustrates that applying the MHVRs (or EHVRs) and SC simultaneously can also improve the prediction accuracy of the station terms in a sparse strong motion network situation. Table 3 presents the standard errors of the estimated standard deviations. The standard errors of the estimated standard deviations from MHVRs are approximately twice that from EVHRs because of the lower number of stations.

Application to a new site

The proposed EHVRs-SC-based and MHVRs-SC-based models can be used to predict the station terms for a new site with available EHVRs or MHVRs information. For most

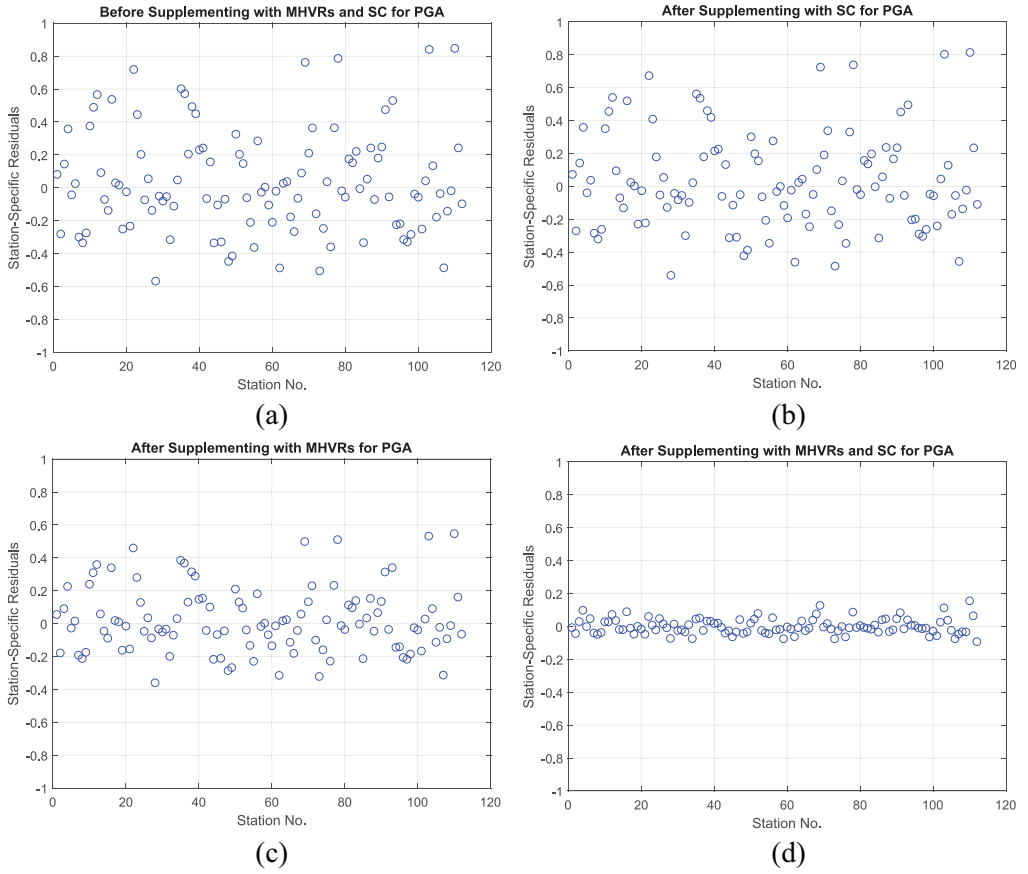


Figure 8. Station-specific residuals for PGA (a) before supplementation with MHVRs and SC (Model 0), (b) after supplementation with MHVRs (Model 1), (c) after supplementation with SC (Model 2), and (d) after supplementation with MHVRs and SC (Model 3).

cases, MHVRs information from microtremor measurements is easier to derive for a new site. However, the MHVRs-SC-based model was developed using fewer stations than the EHVRs-SC-based model. To demonstrate the performance of the proposed MHVRs-SC-based models, we use all MHVRs collected at 3699 sites to calculate the predicted station term $\mathcal{S}_{s,s}^*$ for PGA using the proposed MHVRs-SC-based model and comparing it to the predicted station terms $\mathcal{S}_{s,s}^*$ for PGA using all EHVRs collected at 721 stations and the proposed EHVRs-SC-based model. A comparison of the spatial distribution of the predicted station terms $\mathcal{S}_{s,s}^*$ for PGA using the EHVRs-SC-based and MHVRs-SC-based models is presented in Figure 9a and b. Similar spatial distributions of the predicted station terms of the two models can be observed in these plots. This means that the same regional characteristics can be appropriately represented by the MHVRs-SC-based model even when the number of stations used to develop the model is limited. Figure 10a and b illustrates the spatial distribution of the standard deviation of the predicted station terms $\mathcal{S}_{s,s}^*$ for PGA using the EHVRs-SC-based and MHVRs-SC-based models. Generally, higher prediction errors can be found at the stations that are far away from the stations used to develop the models. Prediction errors of the predictions are also higher for the MHVRs-

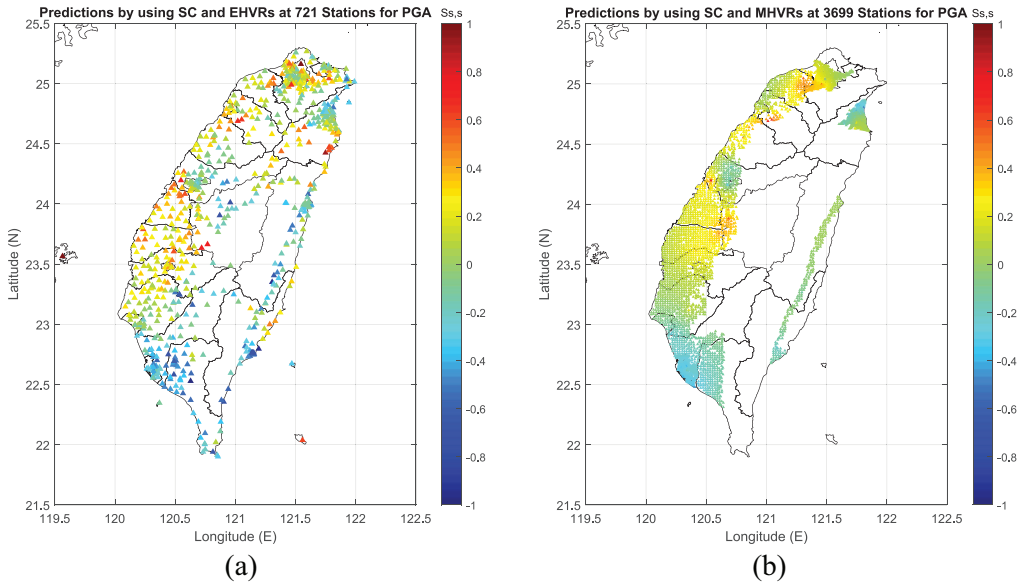


Figure 9. Spatial distribution of the predicted station terms $S_{s,s}^*$ for PGA using (a) SC and EHVRs at 721 stations and (b) SC and MHVRs at 3699 sites.

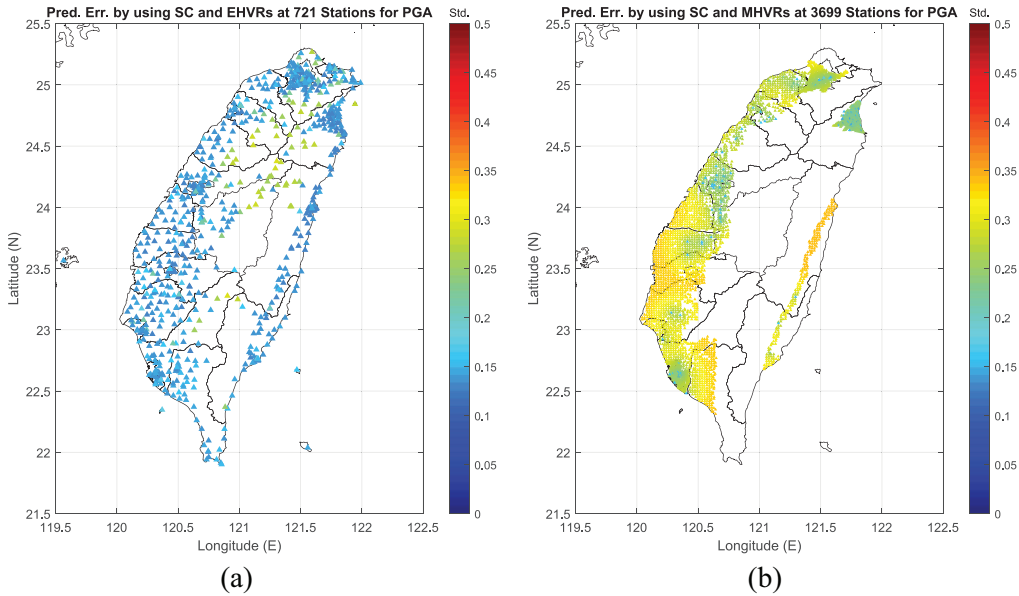


Figure 10. Spatial distribution of the standard deviation of the predicted station terms $S_{s,s}^*$ for PGA using (a) SC and EHVRs at 721 stations and (b) SC and MHVRs at 3699 sites.

SC-based model than the EHVRs-SC-based model due to fewer number of stations being used to develop the model.

When the proposed models are used with a new target site to predict the site effect, the site parameters V_{s30} and $Z_{1.0}$ remain necessary. As mentioned, all stations with available

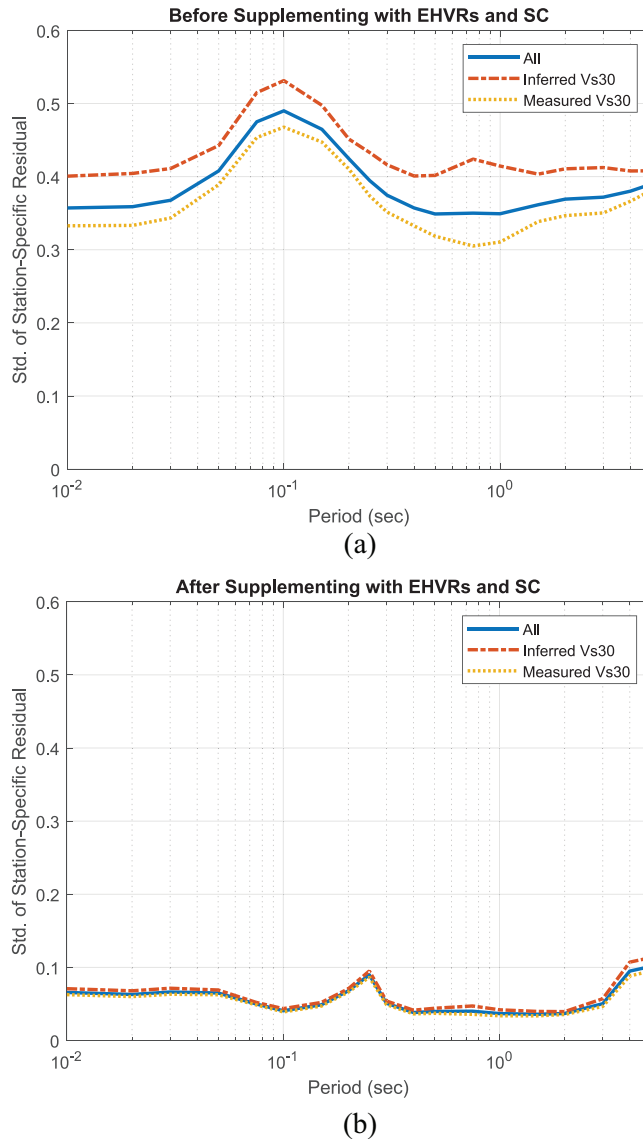


Figure 11. Comparison of the standard deviations of the station-specific residuals of all stations; stations with inferred Vs30 and stations with measured Vs30 (a) before supplementation with EHVRs and SC and (b) after supplementation with EHVRs and SC.

EHVRs have Vs30 and Z1.0 information. Nearly all Z1.0 values are inferred through receiver function analysis using seismic data (Lin et al., 2018). Some Vs30 values are determined from direct measurements (Kuo et al., 2012, 2016), and some are inferred from multiple proxies (Kwok et al., 2018) or through receiver function analysis using seismic data (Lin et al., 2018). A comparison of the standard deviations of the station-specific residuals before supplementation with the EHVRs and SC (Model 0) for all stations, stations with inferred Vs30, and stations with measured Vs30 is presented in Figure 11a, and the standard deviations of the station-specific residuals after supplementation with the EHVRs

and SC (Model 3) are presented in Figure 11b. The value of the standard deviation depends on how the Vs30 value is derived when using the model before it is supplemented with the EHVRs and SC. Higher prediction accuracy can be achieved for measured Vs30 when using the Vs30- and Z1.0-based model. After supplementation with the EHVRs and SC, the standard deviation is no longer dependent on how the Vs30 value is derived. This means that regardless of whether the Vs30 value is measured or inferred, almost identical prediction accuracy can be achieved. In real applications, an efficient approach can be employed to derive Vs30 information for a new target site when using the proposed EHVRs-SC-based or MHVRs-SC-based models.

We conducted a test case to develop Models 1, 2, and 3 but without considering Vs30 and Z1.0 to predict the station terms, and the standard deviations of the station-specific residuals of two cases (with and without Vs30 and Z1.0) were very close to each other. It seems that Vs30 and Z1.0 are not necessary if SC and HVRs are used. However, for a new target site that is far away from the stations with the available station terms, or with an observed HVRs that differs greatly from the available HVRs, Vs30 and Z1.0 remain necessary predictors for knowing site condition and capturing the site effects. Otherwise, a predicted station term can only be derived as the average of all stations, and this is not reasonable. Moreover, the nonlinear site effect term in NCREE19 GMPE uses Vs30 as a predictor. According to these points, we believe that Vs30 and Z1.0 currently remain necessary predictors for capturing site effects.

We supplement the proposed EHVRs-SC-based and MHVRs-SC-based models by adding the $S_{s,s}^*$ term to the NCREE19 GMPE as new GMPEs to illustrate their performance. Figure 12a and b shows comparison plots of the median predictions for stations CHY005 and ILA034, respectively, produced by the NCREE19 GMPE, EHVRs-SC-based model, and MHVRs-SC-based model proposed in this study for a vertical dip crustal strike-slip fault with a moment magnitude (M) of 6 and closest distance from site to rupture plane (R_{rup}) of 10 km. Comparisons of the EHVRs and MHVRs used in the EHVRs-SC-based model and MHVRs-SC-based model for these two stations are presented in Figure 13a and b. The NCREE19 GMPE produces almost identical predictions for these two stations because their Vs30 values are similar. The Z1.0 values of these two stations are 385 and 783 m, respectively. This difference in Z1.0 does not cause considerable differences in the NCREE19 GMPE predictions. However, both the EHVRs-SC-based and MHVRs-SC-based models can represent the difference in the site amplification of these stations with similar Vs30 values. Both the EHVRs-SC-based and MHVRs-SC-based models have consistent prediction results for a station, even when the EHVRs and MHVRs of the station are extremely different, as can be observed in Figure 13a and b. This means that similar prediction accuracy can be achieved by both models. The proposed models indicate higher site amplification of a short period spectral acceleration for the CHY005 station and a lower site amplification of a short-period spectral acceleration for the ILA034 station. This may be due to the higher EHVRs and MHVRs (from 0.1 to 10 Hz) observed at CHY005 station. Local peaks at frequencies of approximately 0.1 to 0.2 Hz can be observed in the EHVRs and MHVRs of ILA034 station. However, no significant site amplification is identified for a long period of spectral acceleration at station ILA034. This indicates that the observed higher HVR at a certain frequency do not directly relate to the observed higher spectral acceleration for a given period. It also means that using the HVRs of a wide frequency range as a vectorized predictor variable to quantify the site effect allows better results than using only one or two predictors (e.g. a peak frequency or HVR at a certain frequency) extracted from the HVRs. This case study demonstrates that the proposed

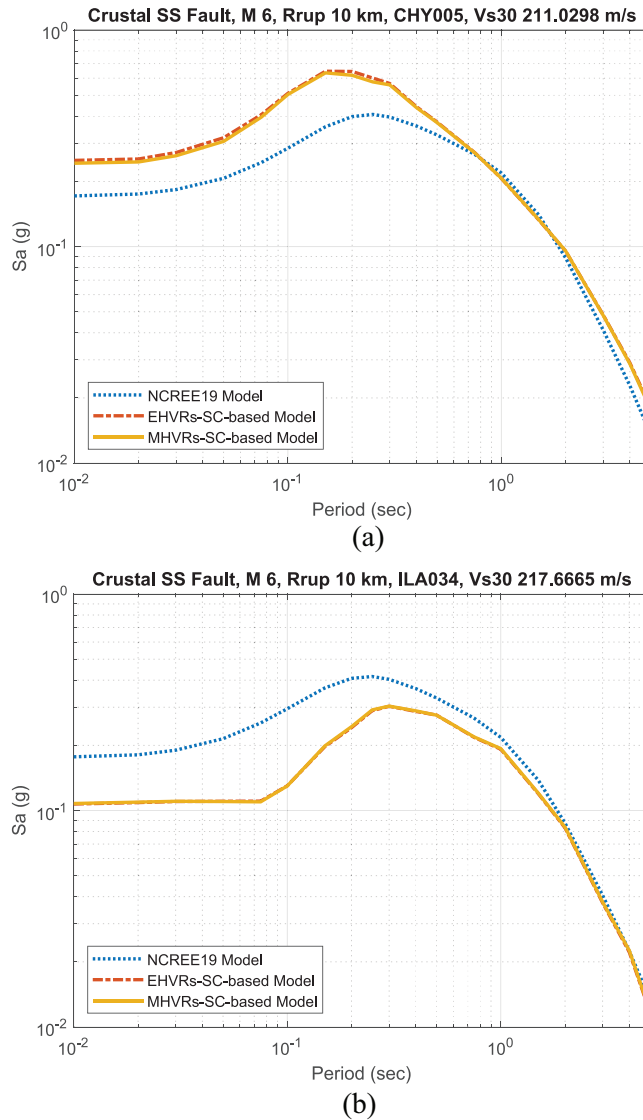


Figure 12. Comparison of the median prediction of the NCEE19 GMPE, EHVRs-SC-based model and MHVRs-SC-based model proposed in this study for two individual stations for vertical dip crustal strike-slip fault, M 6, Rrup 10 km; (a) station CHY005 and (b) station ILA034.

methodology can help with implementing observed EHVRs and MHVRs in GMPEs to improve the prediction accuracy of the site effect for a target site.

Conclusion

In this study, we propose a methodology that involves using HVRs calculated from strong ground motion induced by earthquake (EHVRs) and from ambient ground motion observed from microtremor (MHVRs) individually and with SC in a GMPE to improve the prediction accuracy of the site effect. The proposed methodology was developed using

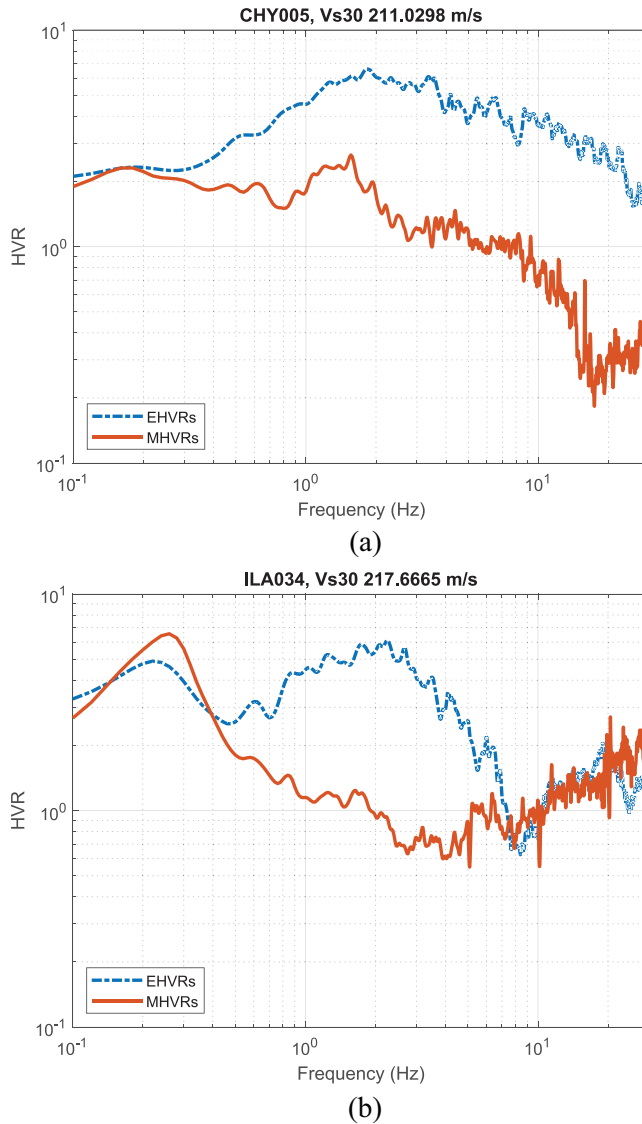


Figure 13. Comparison of the EHVRs and MHVRs for (a) station CHY005 and (b) station ILA034.

the framework of two-step maximum-likelihood regression analysis. The prediction errors of the station term resulting from different available ground-motion record numbers of stations can be considered in the regression analysis. The model behavior and prediction can be well constrained when it is implemented with the SC and EHVRs, or SC and MHVRs through the proposed methodology to predict the station terms even when the observed EHVRs or MHVRs of a target site differs entirely from those previously been observed.

We illustrate the proposed methodology by developing a Vs30- and Z1.0-based model supplemented with the SC and EHVRs collected at 628 strong motion stations and another Vs30- and Z1.0-based model supplemented with the SC and MHVRs collected at

112 stations collocated with those of 112 strong motion stations. The standard deviations of the station-specific residuals can be reduced up by to 90% with the supplementation of SC and EHVRs, or SC and MHVRs. The spatial distribution of the station terms predicted from the MHVRs at 3699 sites is consistent with the predicted station terms predicted from the EHVRs at 721 strong motion stations when the proposed models are applied. This means that the same regional characteristics can be well represented by the MHVRs-SC-based model even when the number of stations used to develop the model is limited. The proposed MHVRs-SC-based model can be improved when additional stations with available MHVRs information and available station terms from the NCREE19 GMPE are collected and included.

We supplement the proposed EHVRs-SC-based and MHVRs-SC-based models by adding the $S_{v,s}^*$ term to the NCREE19 GMPE as new GMPEs to illustrate their performance. Both models can represent the difference of site amplification for stations with similar Vs30 values. Both models have consistent prediction results for the same station even when the EHVRs and MHVRs of the station are markedly different.

From using the proposed EHVRs-SC-based model or MHVRs-SC-based model, the prediction accuracy for stations with inferred Vs30 values is similar to that for stations with measured Vs30. This means that whether the Vs30 value is measured or inferred, similar prediction accuracy can be achieved. For a real application, an efficient approach can be selected to derive Vs30 information for a new target site when using the proposed model.

In summary, this study provides a methodology that involves using SC and EHVRs, or SC and MHVRs in a GMPE to improve the prediction accuracy of the site effect for a target site with available EHVRs or MHVRs. The accuracy of ground-motion prediction can be improved for a target site with available HVRs information by using the proposed EHVRs-SC-based model or MHVRs-SC-based model. This will contribute to site-specific probability seismic hazard analysis for a target site with available HVRs information.

Author's Note

Chun-Hsiang Kuo is now affiliated with National Central University, Taoyuan, Taiwan.

Acknowledgments

The authors appreciate the management of Taiwan SSHAC Level 3 PSHA Project for providing us with the Taiwan ground-motion database flatfile. We also extend our deepest thanks to the Seismology Center at the Central Weather Bureau of Taiwan for compiling and providing the strong ground-motion records for this study.


Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by the Ministry of Science and Technology of Taiwan (grant no. MOST 108-2116-M-492-007).

ORCID iD

Shu-Hsien Chao  <https://orcid.org/0000-0002-1465-3032>

References

- Abrahamson NA, Kuehn NM, Walling M and Landwehr N (2019) Probabilistic seismic hazard analysis in California using nonergodic ground-motion models. *Bulletin of the Seismological Society of America* 109(4): 1235–1249.
- Al-Atik L, Abrahamson N, Bommer JJ, Scherbaum F, Cotton F and Kuehn N (2010) The variability of ground-motion prediction models and its components. *Seismological Research Letters* 81: 794–801.
- Baker JW (2011) Conditional mean spectrum: Tool for ground motion selection. *Journal of Structural Engineering* 137(3): 322–331.
- Baker JW and Cornell CA (2006) Correlation of response spectral values for multi-component ground motions. *Bulletin of the Seismological Society of America* 96(1): 215–227.
- Baker JW and Jayaram N (2008) Correlation of spectral acceleration values from NGA ground motion models. *Earthquake Spectra* 24(1): 299–317.
- Bauer M, van der Wilk M and Rasmussen CE (2016) *Understanding Probabilistic Sparse Gaussian Process Approximations*. In: *Proceedings of the 30th International Conference on Neural Information Processing Systems*, December 2016, pp. 1533–1541.
- Chao SH and Chen YH (2019) A novel regression analysis method for randomly truncated strong motion data. *Earthquake Spectra* 35(2): 977–1001.
- Chao SH, Chiou B, Hsu CC and Lin PS (2019a) *Development of horizontal and vertical ground motion model for crustal earthquakes and subduction earthquakes in Taiwan*. NCREE Report No. NCREE-19-003. Taipei, Taiwan: National Center for Research on Earthquake Engineering.
- Chao SH, Chiou B, Hsu CC and Lin PS (2020) A horizontal ground-motion model for crustal earthquakes and subduction earthquakes in Taiwan. *Earthquake Spectra*. Epub ahead of print 13 March. DOI: 10.1177/8755293019891711.
- Chao SH, Lin CM, Kuo CH and Huang JY (2019b) Applying H/V Fourier spectral ratios for predicting the site effect of ground motion. In: *International conference in commemoration of 20th anniversary of the 1999 Chi-Chi earthquake*, Taipei, Taiwan, 15–19 September.
- Chen Y-H and Tsai C-CP (2002) A new method for estimation of the attenuation relationship with variance components. *Bulletin of the Seismological Society of America* 92(5): 1984–1991.
- Dellaportas P and Stephens DA (1995) Bayesian analysis of errors-in-variables regression models. *Biometrics* 51: 1085–1095.
- GeoPentech (2015) *Southwestern United States ground motion characterization SSHAC Level 3*. Technical Report Rev.2, March. Irvine, CA: GeoPentech.
- Ghofrani H and Atkinson GM (2014) Site condition evaluation using horizontal-to-vertical response spectral ratios of earthquakes in the NGA-West 2 and Japanese databases. *Soil Dynamics and Earthquake Engineering* 67: 30–43.
- Girard A, Rasmussen CE, Quinonero-Candela J and Murray-Smith R (2003) Gaussian process priors with uncertain inputs: Application to multiple-step ahead time series forecasting. In: Becker S, Thrun S and Obermayer K (eds) *Advances in Neural Information Processing Systems*, vol. 15. Cambridge, MA: MIT Press, pp. 545–552.
- Goda K and Hong HP (2008) Spatial correlation of peak ground motions and response spectra. *Bulletin of the Seismological Society of America* 98(1): 354–365.
- Goovaerts P (1997) *Geostatistics for Natural Resources Evaluation*. Oxford and New York: Oxford University Press.
- Hassani B and Atkinson GM (2016) Applicability of the site fundamental frequency as a VS30 proxy for central and eastern North America. *Bulletin of the Seismological Society of America* 106: 653–664.

- Hassani B and Atkinson GM (2017a) Application of a site-effects model based on peak frequency and average shear-wave velocity to California. *Bulletin of the Seismological Society of America* 108(1): 351–357.
- Hassani B and Atkinson GM (2017b) Site-effects model for central and eastern North America based on peak frequency and average shear-wave velocity. *Bulletin of the Seismological Society of America* 108(1): 338–350.
- Iba Y and Akaho S (2010) Gaussian process regression with measurement error. *IEICE Transactions on Information and Systems* E93.D(10): 2680–2689.
- Jaimes MA and Candia G (2019) Interperiod correlation model for Mexican interface earthquakes. *Earthquake Spectra* 35(3): 1351–1365.
- Jayaram N and Baker JW (2009) Correlation model for spatially-distributed ground-motion intensities. *Earthquake Engineering and Structural Dynamics* 38(15): 1687–1708.
- Jayaram N, Baker JW, Okano H, Ishida H, McCann MW and Mihara Y (2011) Correlation of response spectral values in Japanese ground motions. *Earthquakes and Structures* 2(4): 357–376.
- Joyner WB and Boore DM (1993) Methods for regression analysis of strong-motion data. *Bulletin of the Seismological Society of America* 83(2): 469–487.
- Kawakami H and Sharma S (1999) Statistical study of spatial variation of response spectrum using free field records of dense strong ground motion arrays. *Earthquake Engineering and Structural Dynamics* 28: 1273–1294.
- Kawase H, Mori Y and Nagshima F (2018) Difference of horizontal-to-vertical spectral ratios of observed earthquakes and microtremors and its application to S-wave velocity inversion based on the diffuse field concept. *Earth Planets Space* 70: 1.
- Kuehn NM and Abrahamson NA (2019) Spatial correlations of ground motion for non-ergodic seismic hazard analysis. *Earthquake Engineering and Structural Dynamics* 49: 4–23.
- Kuehn NM and Scherbaum F (2016) A partially non-ergodic ground-motion prediction equation for Europe and the Middle East. *Bulletin of Earthquake Engineering* 14: 2629–2642.
- Kuo CH, Chen CT, Lin CM, Wen KL, Huang JY and Chang SC (2016) S-wave velocity structure and site effect parameters derived by microtremor arrays in the Western Plain of Taiwan. *Journal of Asian Earth Sciences* 128: 27–41.
- Kuo CH, Lin CM, Chang SC, Wen KL and Hsieh HH (2017) *Site database for Taiwan strong motion stations*. NCREE Report No. NCREE-17-004. Taipei, Taiwan: National Center for Research on Earthquake Engineering.
- Kuo CH, Wen KL, Hsieh HH, Lin CM, Chang TM and Kuo KW (2012) Site Classification and Vs30 estimation of free-field TSMIP stations using the logging data of EGDT. *Engineering Geology* 129–130: 68–75.
- Kuo CH, Wen KL, Lin CM, Hsiao NC and Chen DY (2018) Site amplifications and the effect on local magnitude determination at stations of the surface-downhole network in Taiwan. *Soil Dynamics and Earthquake Engineering* 104: 106–116.
- Kuo CH, Wen KL, Lin CM, Wen S and Huang JY (2015) Investigating near surface S-wave velocity properties using ambient noise in Southwestern Taiwan. *Terrestrial, Atmospheric and Oceanic Sciences* 26(2): 205–211.
- Kwak DY, Stewart JP, Mandokhail SJ and Park D (2017) Supplementing VS30 with H/V spectral ratios for predicting site effects. *Bulletin of the Seismological Society of America* 107(5): 2028–2042.
- Kwok OLA, Stewart JP, Kwak DY and Sun PL (2018) Taiwan-specific model for VS30 prediction considering between-proxy correlations. *Earthquake Spectra* 34(4): 1973–1993.
- Landwehr N, Kuehn NM, Scheffer T and Abrahamson N (2016) A nonergodic ground-motion model for California with spatially varying coefficients. *Bulletin of the Seismological Society of America* 106(6): 2574–2583.
- Lermo J and Chavez-Garcia F (1993) Site effect evaluation using spectral ratios with only one station. *Bulletin of the Seismological Society of America* 83(5): 1574–1594.
- Lin CM, Kuo CH, Huang JY, Hsieh HH, Si CC and Wen KL (2018) *Shallow shear-wave velocity structures of TSMIP stations in Taiwan*. NCREE Report No. NCREE-18-019. Taipei, Taiwan: National Center for Research on Earthquake Engineering.

- Lin PS, Chiou B, Abrahamson N, Walling M, Lee CT and Cheng CT (2011) Repeatable source, site, and path effects on the standard deviation for empirical ground-motion prediction models. *Bulletin of the Seismological Society of America* 101(5): 2281–2295.
- Liu KS, Shin TC and Tsai YB (1999) A free-field strong motion network in Taiwan: TSMIP. *Terrestrial, Atmospheric and Oceanic Sciences* 10(2): 377–396.
- Morikawa N, Kanno T, Narita A, Fujiwara H, Okumura T, Fukushima Y and Guerpinar A (2008) Strong motion uncertainty determined from observed records by dense network in Japan. *Journal of Seismology* 12(4): 529–546.
- Nakamura Y (2019) What is the Nakamura method? *Seismological Research Letters* 90(4): 1437–1443.
- National Center for Research on Earthquake Engineering (NCREE) (2015) Web page for reevaluation of probabilistic seismic hazard of nuclear facilities in Taiwan using SSHAC Level 3 Methodology Project. Available at: <http://sshac.ncee.org.tw> (accessed 31 December 2019).
- Rasmussen CE and Williams CKI (2006). *Gaussian Processes for Machine Learning*. Cambridge: MIT Press.
- Rodriguez-Marek A, Rathje EM, Bommer JJ, Scherbaum F and Stafford PJ (2014) Application of single-station sigma and site-response characterization in a probabilistic seismic-hazard analysis for a new nuclear site. *Bulletin of the Seismological Society of America* 104(4): 1601–1619.
- Rodriguez-Marek A, Cotton F, Abrahamson NA, Akkar S, Al Atik L, Edwards B, Montalva GA and Dawood HM (2013) A model for single-station standard deviation using data from various tectonic regions. *Bulletin of the Seismological Society of America* 103(6): 3149–3163.
- Shin TC, Chang CH, Pu HC, Lin HW and Leu PL (2013) The geophysical database management system in Taiwan. *Terrestrial, Atmospheric and Oceanic Sciences* 24(1): 11–18.
- Sokolov V, Wenzel F, Jean WY and Wen KL (2010) Uncertainty and spatial correlation of earthquake ground motion in Taiwan. *Terrestrial, Atmospheric and Oceanic Sciences* 21: 905–921.
- Stafford PJ, Rodriguez-Marek A, Edwards B, Kruiver PP and Bommer JJ (2017) Scenario dependence of linear site-effect factors for short-period response spectral ordinates. *Bulletin of the Seismological Society of America* 107(6): 2859–2872.
- Wang M and Takada T (2005) Macrospectral correlation model of seismic ground motions. *Earthquake Spectra* 21: 1137–1156.
- Wen KL and Huang JY (2012) Dense microtremor survey for site effect study in Taiwan. In: *15th world conference of earthquake engineering*, Lisbon, 24–28 September.
- Yamazaki F and Ansary M (1997) Horizontal-to-vertical spectrum ratio of earthquake ground motion for site characterization. *Earthquake Engineering and Structural Dynamics* 26(7): 671–689.