17th World Conference on Earthquake Engineering, 17WCEE

Sendai, Japan - September 13th to 18th 2020

Ground-Motion Prediction Model Considering Spatial Variation of Path Attenuation Effects

T. Hikita[(1)], K. Kasamatsu[(2)], Y. Tomozawa[(3)]

[(1)] *Chief Research Engineer, Kajima Corporation, tomoki@kajima.com*
[(2)] *Senior Research Engineer, Kajima Corporation, kasamake@kajima.com*
[(3)] *Manager, Kobori Research Complex, tomozawy@kajima.com*

Abstract

Improving the accuracy of empirical ground-motion prediction is crucial for input ground-motion evaluation in seismic design and probabilistic seismic hazard analysis. We investigate the improvement of the empirical ground-motion prediction model using the recorded maximum acceleration data for crustal earthquakes around the Fukushima prefecture, Japan. Furthermore, we propose a ground-motion prediction model considering the spatial variation of the attenuation and source characteristics depending on the epicenter location through mixed geographically weighted regression. The proposed model and the conventional multiple regression model that does not consider the spatial variation of the coefficients are applied to the above-mentioned data, and the results are compared. In the proposed model, the differences in the attenuation characteristics depending on the propagation path and those in the source characteristics depending on the epicenter location are modeled by varying the coefficients using the epicenter location as the parameter. The standard deviation of the residual of the proposed model is approximately 20% lesser than that of the conventional one. In particular, the reduction of residuals for large-distance data is significant. Data on the path that crossed the active volcanoes distributed in the target area exhibited significant amplitude attenuation. Using the prediction model, these characteristics could be simulated by varying the coefficients according to the epicenter location of each site. The proposed model is effective in constructing an improved ground-motion prediction model, if sufficient spatially distributed data are available. This method can be used to improve the accuracy of empirical ground-motion prediction at a specific site for rational seismic-hazard analysis.

Keywords: Ground-motion prediction model, Path effect, Multiple regression, Geographically weighted regression

1. Introduction

The ground-motion prediction model is extensively applied as a simple method to estimate the average ground-motion intensity in earthquakes. However, the development of a more accurate prediction model is crucial for input ground-motion evaluation and probabilistic seismic-hazard analysis.

The ground-motion prediction model is generally obtained through regression analysis using vast data on the observed ground motion. The variability of the regression residual is one of indices that indicates the accuracy of the prediction model. There are various causes for residuals, among which the residuals due to poor modeling can be reduced by improving the prediction model. For modeling the attenuation characteristics, the attenuation of the ground-motion intensity is modeled using a simple function with distance index parameters such as the hypocentral distance and closest distance to the fault, in general prediction models. However, due to the spatial heterogeneity of the crust attenuation (e.g., Nakamura and Uetake, 2004; Liu *et al.*, 2014), the attenuation characteristics of different propagation paths may differ, even for the same distance. If there is a difference in the attenuation characteristics depending on the propagation path, it is difficult to accurately simulate data with an average model using the distance index alone as the parameter. If the difference in the attenuation characteristics according to the propagation path can be suitably modeled, the accuracy of ground-motion prediction can be improved.

Although the heterogeneity of attenuation has been extensively studied, only few studies have considered their effect in the ground-motion prediction model. In order to consider the effect of the variation of the attenuation characteristics in the prediction model, this study proposes a prediction model that considers the spatial variation of the regression coefficient according to the propagation path through mixed geographically weighted regression (MGWR) (Fotheringham *et al.*, 2002). The proposed prediction model is validated using the ground-motion data in Japan.

2. Data

Ground-motion data of the shallow crustal earthquakes in Japan were used in this study. We focused on the area around the Fukushima prefecture where several large crustal earthquakes, such as the Mid Niigata prefecture earthquake in 2004 (M_W6.6, M_W indicates the moment magnitude specified by the NIED F-net project), the Niigataken Chuetsu-oki earthquake in 2007 (M_W6.6), and the Fukushima-ken-Hama-Dori earthquake in 2011 (M_W6.6), occurred. We selected crustal earthquakes that occurred in the 36.5°–38.0° N latitude and 138.4°–141.4° E longitude range, with an M_W of 4.0 or more and depth of 20 km or less.

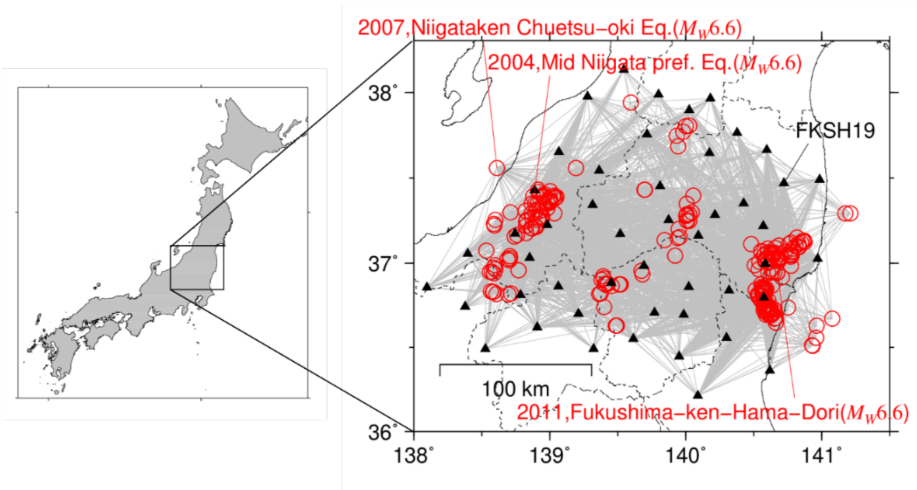

Fig. 1 – Epicenters and data stations. The epicenters are indicated by red circles, whereas the stations are indicated by black triangles. The gray lines connect the epicenters to the respective data observation sites.

In addition, we utilized the borehole records of KiK-net stations (Aoi *et al.*, 2000; Okada *et al.*, 2004.), whose hypocentral distance X was approximately 200 km or less. After eliminating records with low signal-to-noise ratios (S/N), stations with less than 10 records and the earthquakes with less than 10 records were excluded from the selected data. Finally, a total of 4,637 records of 48 stations for 238 earthquakes were used in this study. Fig. 1 shows the epicenters and stations included in the dataset, and Fig. 2(a) depicts the relationship between X and M_W of the dataset. The prediction model of the maximum acceleration of these data was examined as an example. The maximum acceleration is the maximum value of the vector composite amplitude of the two horizontal components of the acceleration waveform bandpass-filtered with a passband of 1–20 Hz. Fig. 2(b) displays X and the maximum acceleration of the dataset. The largest maximum acceleration is approximately 435 cm/s^2, and certain other data exceed 100 cm/s^2. However, as all the target data were borehole records, it was assumed that the effect of the nonlinearity of the surface ground response was less.

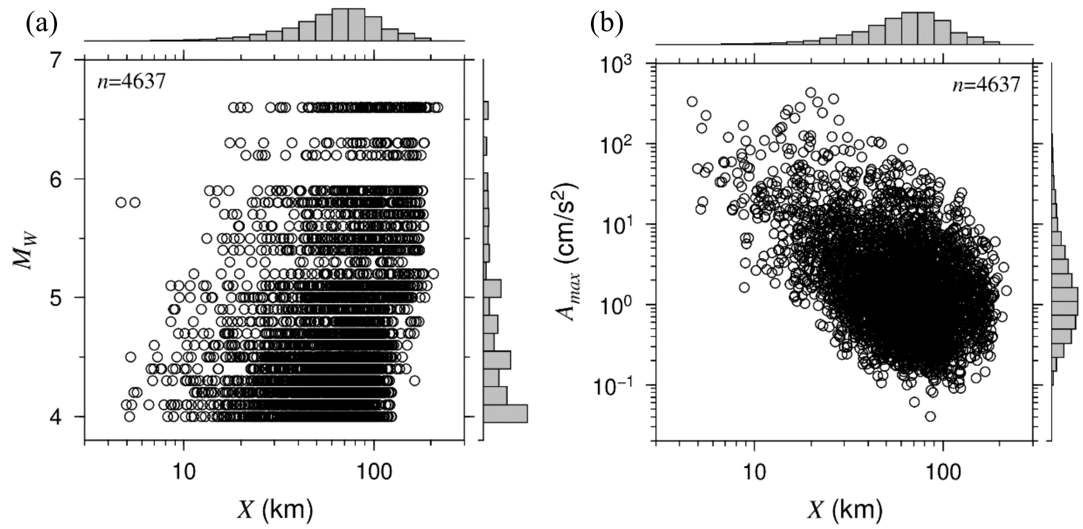

Fig. 2 – (a) Hypocentral distance X and moment magnitude M_W scatterplot of the data, and (b) X and maximum acceleration A_{max} scatterplot of the data. The bar graph outside the box shows the frequency distribution of the data.

3. Models

A conventional prediction model which did not consider the spatial variation of the attenuation characteristics was assumed as the reference. Hereafter, this is referred to as the multiple regression (MR) model. The following MR model was considered:

$$\ln y_{ij} = a_i - \begin{Bmatrix} b_1 \ln X_{ij} & \text{if } X_{ij} < X_C \\ b_1 \ln X_c + b_2 \ln\left(\frac{X_{ij}}{X_C}\right) & \text{if } X_{ij} \geq X_C \end{Bmatrix} - cX_{ij} + d_j + \delta w_{ij}, \tag{1}$$

$$a_i = eM_i + fH_i + g + \delta b_i, \tag{2}$$

where y_{ij} is the maximum acceleration (cm/s^2) of the j-th station due to the i-th earthquake; X_{ij} is the hypocentral distance (km); M_i and H_i are the M_W and depth (km) of the i-th earthquake; $a_i, b_1, b_2, c, d_j, e, f, g$ are the partial regression coefficients; δw_{ij} and δb_i are the residuals, and the overall residual δ_{ij} is defined as $\delta_{ij} = \delta w_{ij} + \delta b_i$. The following two-stage multiple regression analysis was performed. In the first-stage, a_i corresponding to the source characteristics of each earthquake, d_j corresponding to the site characteristics of each station, and b_1, b_2, and c corresponding to the attenuation characteristics were evaluated through MGWR using equation (1). FKSH19 (see Fig. 1) was used as the reference station, and d_j for FKSH19 was constrained to zero. In equation (1), the attenuation characteristics are represented by terms using $\ln X_{ij}$ and X_{ij} as explanatory variables. For the $\ln X_{ij}$ term, different coefficients were allowed, bounded by distance X_C which was set to 34 km, minimizing the standard deviation of δw_{ij} by grid search with a grid of 1 km increments. Second-stage multiple regression analysis was performed using a_i, considering equation (2) to model the source characteristics. In equation (2), the source characteristics are represented by terms using M_i and H_i as explanatory variables, and g is a constant term. In this study, we focused on improving the accuracy of ground-motion prediction at the observation station by utilizing ground-motion records. Therefore, the site characteristics required for prediction in locations with no records were not modeled.

Next, in order to reduce the residuals, a prediction model that considered the spatial variation of the attenuation characteristics and source characteristics was applied. In equation (1), by introducing local coefficients $d_j^{\prime}(u_i, v_i)$, using the latitude u_i and longitude v_i of the epicenter of the i-th earthquake for each station, we considered a model whose attenuation characteristics varied depending on the propagation path:

$$\ln y_{ij} = a_i^{\prime} - \begin{Bmatrix} b_1^{\prime} \ln X_{ij} & \text{if } X_{ij} < X_C \\ b_1^{\prime} \ln X_c + b_2^{\prime} \ln\left(\frac{X_{ij}}{X_C}\right) & \text{if } X_{ij} \geq X_C \end{Bmatrix} - c^{\prime} X_{ij} + d_j^{\prime}(u_i, v_i) + \delta w_{ij}^{\prime}. \tag{3}$$

The difference between equations (1) and (3) is that d_j in equation (1) is replaced by $d_j^{\prime}(u_i, v_i)$ in equation (3). All the other coefficients are the same; however, they are primed to avoid confusion. The local regression coefficients which vary with the epicenter location and the other global regression coefficients were evaluated through MGWR (Fotheringham *et al.*, 2002). In equation (3), $d_j^{\prime}(u_i, v_i)$ can be regarded as a coefficient that combines the spatial variation of the attenuation characteristics and site characteristics of each station.

For modeling the source characteristics, the intercept coefficient g in equation (2) was replaced by local coefficient $g^{\prime}(u_i, v_i)$ that considers the difference in the source characteristics depending on the epicenter location:

$$a_i^{\prime} = e^{\prime} M_i + f^{\prime} H_i + g^{\prime}(u_i, v_i) + \delta b_i^{\prime}. \tag{4}$$

First-stage MGWR was performed using equation (3), after which second-stage MGWR was performed using equation (4) with the obtained coefficients $a_i^{\prime}$. In the first-stage MGWR, $d_j^{\prime}(u_i, v_i)$ for FKSH19 was constrained to zero and X_C was set to 34 km as in the MR model. The overall residual $\delta_{ij}^{\prime}$ was defined as $\delta_{ij}^{\prime} = \delta w_{ij}^{\prime} + \delta b_i^{\prime}$. In MGWR, the following Gaussian kernel function was used:

$$w_n(u_m, v_m) = \exp\left\{-\frac{1}{2}\left(\frac{d_{mn}}{h}\right)^2\right\}. \tag{5}$$

In the above equation, m and n are the number of the earthquakes corresponding to each data; d_{mn} is the distance between the epicenter of the m-th and n-th data; h is a parameter that controls the weighting range and is called bandwidth. Through grid search, h that minimized AIC_C (Hurvich and Tsai, 1989) of the MGWR result was adopted. In the first stage, $h = 30.91$ km and in the second, $h = 11.83$ km. Hereafter, the prediction model obtained using the above procedure is referred to as the MGWR model.

4. Results

Table 1 lists the estimates and standard errors of the MR model regression coefficients, except for a_i and d_j, whereas Table 2 lists those of the MGWR model, except for $a_i^{\prime}$, $d_j^{\prime}(u_i, v_i)$ and $g^{\prime}(u_i, v_i)$. Fig. 3(a) shows the amplitudes of the correction data $\ln y_{ij} - a_i - d_j$ and $\ln y_{ij} - a_i^{\prime} - d_j^{\prime}(u_i, v_i)$ with respect to the hypocentral distance. There were no significant differences in the average attenuation characteristics between the MR and MGWR models. Fig. 3(b) depicts a_i and $a_i^{'}$ obtained through first-stage regression with respect to the magnitude. Comparing a_i and $a_i^{'}$, no significant differences were observed in the average slope of the magnitude.

Table 1 – Regression coefficients (MR model)

Coefficient	Estimate	Standard Error
b_1	1.7546	0.0474
b_2	0.7297	0.0998
c	0.0137	0.0013
e	1.4323	0.0513
f	0.0567	0.0070
g	0.6069	0.2359

Table 2 – Regression coefficients (MGWR model)

Coefficient	Estimate	Standard Error
$b_1^{'}$	1.8759	0.0496
$b_2^{'}$	1.0590	0.1214
$c^{'}$	0.0074	0.0018
$e^{'}$	1.4720	0.0738
$f^{'}$	0.0594	0.0069

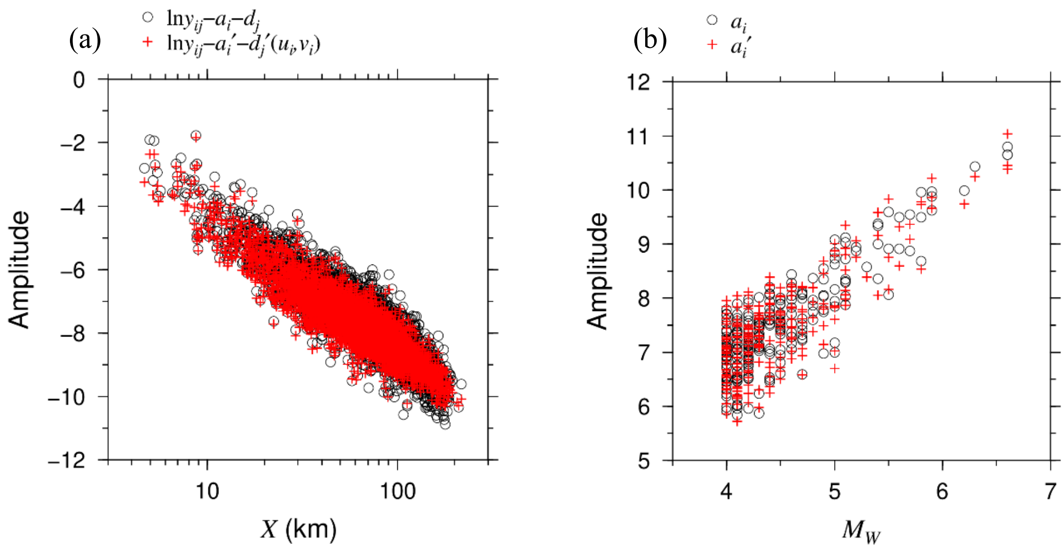

Fig. 3 – (a) Attenuation characteristics of correction data $\ln y_{ij} - a_i - d_j$ and $\ln y_{ij} - a_i^{'} - d_j^{'}(u_i, v_i)$ by the two prediction models, and (b) Comparison of coefficients a_i and a_i' with respect to M_W.

Fig. 4 displays the spatial distributions of the local coefficients $d_j^{'}(u_i, v_i)$ of the MGWR model for two stations NIGH11 and TCGH10; in addition to the epicenter data used for MGWR, hypothetical locations where the display area of the figure is divided into 150 in the longitudinal direction and 100 in the direction of the latitude are considered, and the estimated $d_j^{'}(u_i, v_i)$ are displayed. In ground-motion prediction using the MGWR model, it is possible to predict using the coefficients, according to the epicenter location of the earthquake. $d_j^{'}(u_i, v_i)$ is proportional to the predicted amplitude at the target station of the earthquake that occurs at that location. For an earthquake that occurs in the area where $d_j^{'}(u_i, v_i)$ is large, the attenuation of the path to the target station can be considered smaller than the average, whereas it can be considered larger for an earthquake that occurs in an area where $d_j'(u_i, v_i)$ is small. The red triangles in Fig. 4 indicate active volcanoes (JMA, 2018). NIGH11 was located on the west and TCGH10 was on the east of the active volcanoes distributed linearly from northeast to southwest. $d_j'(u_i, v_i)$ for each station differed east and west of the active volcano distribution. In both cases, the attenuation was large for earthquakes that occurred on the opposite side of the active volcano distribution from the station. This may be caused by the strong attenuation of the amplitude of the seismic wave propagating through the active volcano distribution. However, a more detailed study is needed for the physical interpretation of $d_j'(u_i, v_i)$, focusing on the correspondence with the spatial heterogeneity of the attenuation.

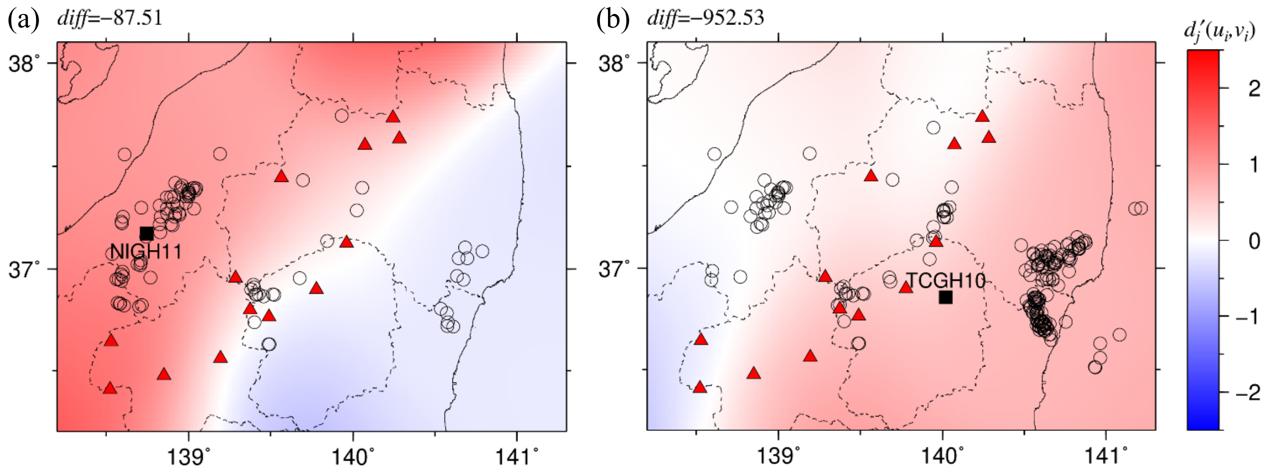

Fig. 4 – Distribution of coefficient $d_j'(u_i, v_i)$ for stations (a) NIGH11 and (b) TCGH10. The red triangles indicate active volcanoes (JMA, 2018). The black square and circles indicate the target station and epicenters of the observed earthquakes at the station. $diff$ indicates the difference between AIC_C and AIC_C' of the MGWR result when the local coefficient $d_j'(u_i, v_i)$ for the target station is changed to a global coefficient.

Fig. 5 shows the spatial distribution of the local coefficient $g'(u_i, v_i)$ representing the spatial variation of the source characteristics. $g'(u_i, v_i)$ for eastern Pacific-coast earthquakes were greater than those of the other earthquakes. There was a bias in the distribution of the fault types for the earthquakes used in this study. Fault-type index $ftype$ (Shearer *et al.*, 2006) evaluated based on the CMT solution by NIED is depicted in Fig. 6. Figs. 5 and 6 were compared to clarify whether the spatial variation of $g'(u_i, v_i)$ reflected the effect of the differences in the fault type or regional differences in the source characteristics. However, if the fault type of the earthquake is limited in the source region, its effect can be evaluated as a spatial variation of coefficient $g'(u_i, v_i)$.

The smoothness of the spatial variation of $d_j'(u_i, v_i)$ and $g'(u_i, v_i)$ was characterized by the bandwidth of the kernel function used in MGWR. The determination of the bandwidth was significantly affected by the epicenter distribution of the data. In future, by increasing the number of observed earthquakes and the density of the epicenter distribution of the data, more local spatial variations can be evaluated.

Fig. 7(a) shows the relationship between the residuals of the two prediction models and the hypocentral distance, whereas Fig. 7(b) shows the that between the residuals and M_W. The standard deviation of the MGWR model was approximately 20% lesser than that of the MR model. Moreover, AIC_C (Table 3) of the MGWR model was smaller, indicating that this model was more effective than the MR model. Figs. 8(a) and 8(b) display the standard deviation of the residuals of X and M_W, respectively, for each bin. As observed in

Fig. 8(a), the standard deviation of the MGWR model reduces for large X. Hence, the MGWR model is expected to be effective in improving the accuracy of ground-motion prediction for distant earthquakes.

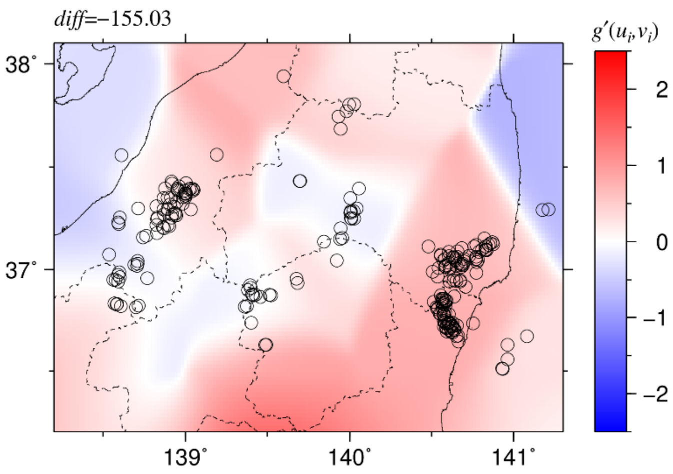

Fig. 5 – Spatial distribution of coefficient $g'(u_i, v_i)$.

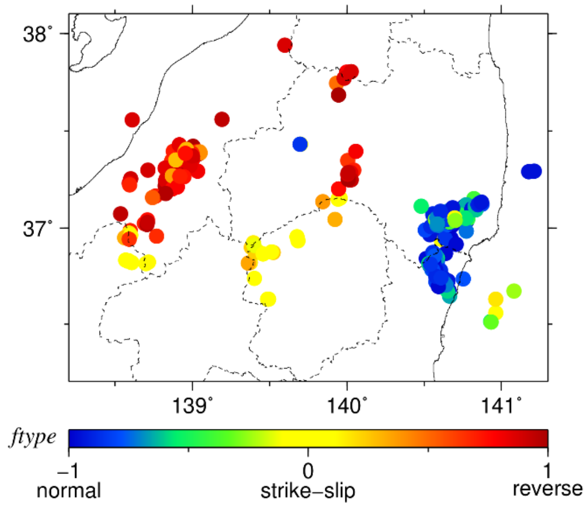

Fig. 6 – Distribution of the earthquake fault-type index (Shearer *et al.*, 2006).

Table 3 – AIC_C in each regression

	MR model	MGWR model
First-stage regression	7782.97	6608.20
Second-stage regression	275.82	187.31

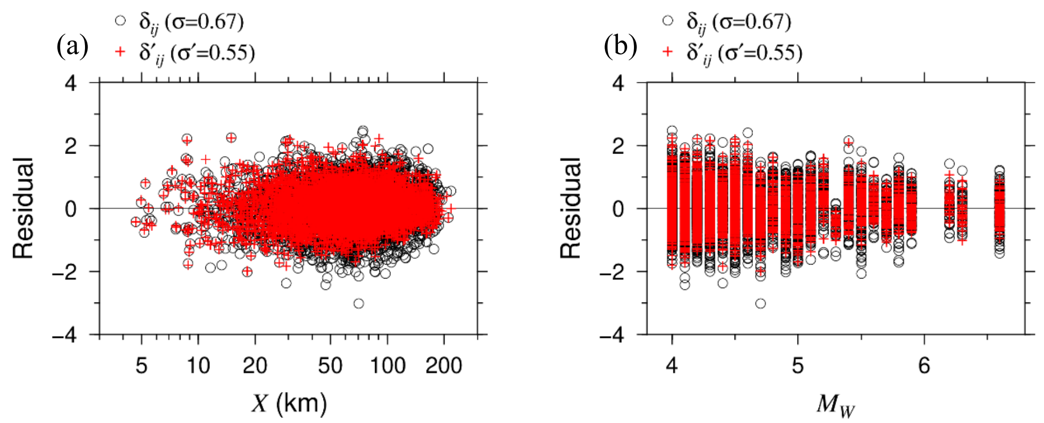

Fig. 7 – (a) Relationship between the overall residual and (a) hypocentral distance X and (b) M_W. The black circles indicate the residuals of the MR model. The red crosses indicate the residuals of the MGWR model. σ and σ' indicate the standard deviation of δ_{ij} and δ'_{ij}, respectively.

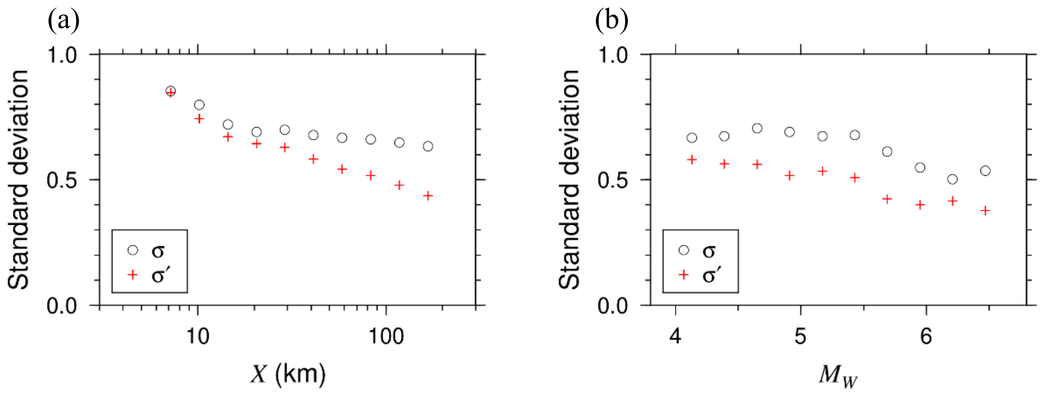

Fig. 8 – Standard deviation of the residuals of (a) X and (b) M_W for each bin.

5. Discussion

5.1 Local-coefficient significance

In order to determine the significance of the spatial variation of local coefficient $d_j^{\prime}(u_i, v_i)$, AIC_C of the MGWR model was compared with $AIC_C^{\prime}$ which is AIC_C of the regression model where $d_j^{\prime}(u_i, v_i)$ for the target station was replaced by a global coefficient . The difference $diff$ ($diff = AIC_C - AIC_C^{\prime}$) is displayed on the upper-left of Fig. 4; in both examples, $diff$ is negative. As the model with local coefficient $d_j^{\prime}(u_i, v_i)$ had a smaller AIC_C, the spatial variation of $d_j^{\prime}(u_i, v_i)$ was considered significant. For coefficients $d_j^{\prime}(u_i, v_i)$ of all stations, $diff$ was negative, except for two stations (FKSH01 and FKSH12). As an example where $diff$ is positive, the spatial distribution of coefficient $d_j^{\prime}(u_i, v_i)$ for FKSH01 is depicted in Fig. 9; the spatial variation of coefficient $d_j^{\prime}(u_i, v_i)$ is small, and the residual does not change considerably, even if this coefficient is replaced by a global coefficient. The reason for this small spatial variation is not clear; however, it may be difficult to model the spatial variation of the attenuation characteristics shown in Fig. 4 for stations where other factors dominate the data variability. In such a case, it is better to change the local coefficient alone for the target station to a global coefficient.

5.2 Reference station in first-stage MGWR

Fig. 10 shows the comparison between d_j and $d_j^{\prime}(u_i, v_i)$ for each station. For $d_j^{\prime}(u_i, v_i)$, the mean and standard deviation of the epicenter values of the dataset are presented. The mean of $d_j^{\prime}(u_i, v_i)$ was often close to d_j. The standard deviation of $d_j^{\prime}(u_i, v_i)$ differed depending on the station. In first-stage MGWR, coefficient $d_j^{\prime}(u_i, v_i)$ for reference station FKSH19 was constrained to zero without considering the spatial variation. However, as shown in Fig. 10, as the spatial variation of $d_j^{\prime}(u_i, v_i)$ varies depending on the station, it is better to select the reference station with a small spatial variation in $d_j^{\prime}(u_i, v_i)$. Therefore, we performed MGWR using an arbitrary reference station, and compared the spatial variations of $d_j^{\prime}(u_i, v_i)$ for each station. Further, the station with a small spatial variation in $d_j^{\prime}(u_i, v_i)$ was re-selected as the reference station, and first-stage

MGWR was performed. When comparing the spatial variations of coefficients $d_j^{'}(u_i, v_i)$, it should be noted that these variations may not be appropriately evaluated for a station with less data.

5.3 Applicability to seismic hazard analysis

The MGWR model could reduce the regression residual compared to the conventional MR model. When considering the uncertainty of ground-motion prediction using each model, the MR model needs to consider a larger uncertainty than the MGWR model. This is important in seismic hazard analysis considering the uncertainty of ground-motion prediction. By applying the MGWR model for ground-motion prediction in seismic hazard analysis, it is possible to avoid overestimating the uncertainty, and realize a more reasonable seismic hazard analysis. Based on the results of the previous section, the MGWR model can be considered effective in reducing the uncertainty of ground-motion prediction, particularly for distant earthquakes. In general, there are many sites where the effect of distant large earthquakes on the seismic hazard are dominant. Therefore, by applying the MGWR model, rational the seismic hazard analysis can be performed at these sites.

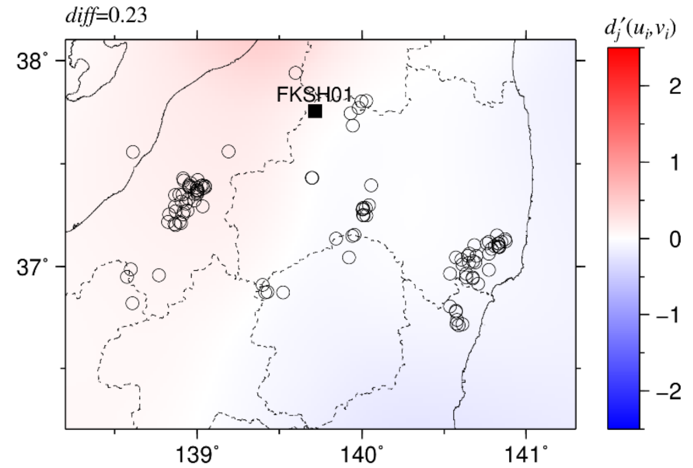

Fig. 9 – Distribution of coefficient $d_j^{'}(u_i, v_i)$ for station FKSH01. The black square indicates the station, and the circles indicate the epicenters of the observed earthquakes at the station. $diff$ indicates the difference between AIC_C and $AIC_C^{'}$ of the MGWR result when the local coefficient $d_j^{'}(u_i, v_i)$ of the target station is changed to a global coefficient.

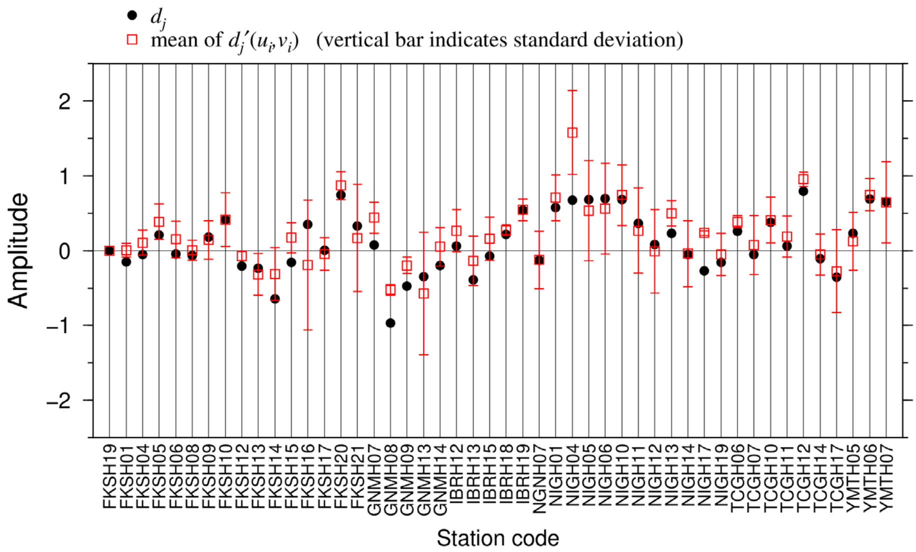

Fig. 10 – Coefficient d_j for all stations and average of $d_j^{'}\ (u_i, v_i)$.

6. Conclusions

In order to improve the accuracy of empirical ground-motion prediction, we proposed a ground-motion prediction model considering the spatial variation of the attenuation and source characteristics, through MGWR. On applying this model to data on crustal earthquakes around the Fukushima prefecture, Japan, the standard deviation of the residuals was reduced by approximately 20% compared to the conventional multiple regression model. The proposed method is effective for obtaining the ground-motion prediction model, if sufficient data are available to appropriately evaluate the spatial variations of the coefficients. The obtained ground-motion prediction model may be used for rational seismic hazard analysis. In future, it is intended to verify the validity of the proposed method using data from other regions and different types of earthquakes, and consider further improvements in modeling.

7. Acknowledgements

The strong-motion records of K-NET and KiK-net operated by NIED were used in this study. We appreciate all the concerned personnel.

8. References

[1] Nakamura R, Uetake T (2004): Three dimensional attenuation structure beneath the Tohoku district by using seismic strong motion records, *ZISIN, Journal of the Seismological Society of Japan*, **56**, 447-455. (in Japanese)

[2] Liu X, Zhao D, Li S (2014): Seismic attenuation tomography of the Northeast Japan arc: Insight into the 2011 Tohoku earthquake (M_W9.0) and subduction dynamics, *Journal of Geophysical Research, Solid Earth*, **119**, 1094-1118.

[3] Fotheringham AS, Brunsdon C, Charlton M (2002): *Geographically weighted regression: The analysis of spatially varying relationships*, Wiley.

[4] National Research Institute for Earth Science and Disaster Resilience (NIED), F-net, http://www.fnet.bosai.go.jp/top.php, viewed 18 Oct. 2018.

[5] Aoi S, Obara K, Hori S, Kasahara K, Okada Y (2000): New strong-motion observation network: KiK-net, *Eos Trans. AGU*, **81**, 329.

[6] Okada Y, Kasahara K, Hori S, Obara K, Sekiguchi S, Fujiwara H, Yamamoto A (2004): Recent progress of seismic observation networks in Japan -Hi-net, F-net, K-NET and KiK-net-, *Earth Planets Space*, **56**, XV-XVIII.

[7] Hurvich CM, Tsai CL (1989): Regression and time series model selection in small samples, *Biometrika*, **76**, 297-307.

[8] Japan Meteorological Agency (2018): National catalogue of the active volcanoes in Japan, https://www.data.jma.go.jp/ svd/vois/data/tokyo/STOCK/souran_eng/menu.htm, viewed 24 May 2019.

[9] Shearer PM, Prieto GA, Hauksson E (2006): Comprehensive analysis of earthquake source spectra in southern California, *Journal of Geophysical Research*, **111**, B06303.

