17th World Conference on Earthquake Engineering, 17WCEE
Sendai, Japan - September 13th to 18th 2020

STUDY ON GROUND MOTION MODELS FOR KANTO REGION, JAPAN, UTILIZING MACHINE LEARNING

Atsuko Oana[1], Toru Ishii[2], Kensuke Wada[3]

[1] *Research Engineer, Dr.Eng., Shimizu Corporation, a.oana@shimz.co.jp*
[2] *Chief Research Engineer, Dr.Eng., Shimizu Corporation, tokyo@shimz.co.jp*
[3] *Research Engineer, M.Eng., Shimizu Corporation, wada@shimz.co.jp*

Abstract

As a preliminary study for constructing the future earthquake ground motion prediction model in which all the ground motion records will be accumulated every day, will be utilized and will be updated, we try to create the future earthquake ground motion evaluation models by machine learning using the past ground motion records obtained at the observation stations in the Kanto region, Japan.

The ground motion records obtained at 138 observation stations of the K-NET of NIED (National Research Institute for Earth Science and Disaster Prevention) deployed in the Kanto region are used as the training data. We identify noises and signals of different earthquakes mixed in the records, then exclude them and select the appropriate records for this study.

As the target variables of the machine learning, the peak ground acceleration PGA [cm/s^2], the 5% damped pseudo velocity response spectra $_{\mathrm{P}}S_{\mathrm{V}}$ [cm/s] and the 5% damped velocity response duration spectra TS_{V} [s] (parameter $p1$=0.03, $p2$=0.95) of several periods, are examined, respectively. Since there are few large data of PGA nor $_{\mathrm{P}}S_{\mathrm{V}}$, both are changed into common logarithmic input data ($\log_{10} PGA$ and $\log_{10} {}_{\mathrm{P}}S_{\mathrm{V}}$) for the machine learning to raise the accuracy of the evaluation models. As the features of the machine learning, the moment magnitude M_{W}, the hypocentral depth H [km], the hypocentral distance X [km], the epicentral direction Λ [degree], the top depth of the seismic bedrock $D28$ [m], and the averaged S-wave velocity in the surface layers of total thickness 30 m $AVS30$ [m/s] are selected. Λ is set zero degrees to due north and is defined clockwise. However, a set of sin Λ and cos Λ are inputted for the machine learning because Λ is discontinuous at due north. The "Gradient Boosting Decision Trees" is used as the machine learning method.

The feature impact on the target variable is examined based on the degree of changing the evaluation accuracy of the target variable when only one specific feature is rearranged at random and the other features remain unchanged. The feature impact of M_{W} increase with T, and the feature impact of X and that of H on TS_{V} decrease with T. The feature impact of $AVS30$ is relatively large for the period of 1 second, that of $D28$ is large for the period of 1 second or more, that of Λ is relatively large for the period of 1 second or less. The averages of the ratios of the evaluated earthquake ground motion indexes to the observed ones are almost 1, and the common logarithmic standard deviation is about 0.2 for PGA and $_{\mathrm{P}}S_{\mathrm{V}}$s, and about 0.1 for TS_{V}s. Most of the evaluated earthquake ground motion indexes are staying between twice the average of observed values and half the average.

The attenuation relations of PGA and M_{W} are examined using the proposed ground motion evaluation model. The relations of PGA and M_{W} or X is modeled similar to the ones by the conventional ground motion prediction equations. As a result of examining attenuation relations of $_{\mathrm{P}}S_{\mathrm{V}}$ (T=1 s) and Λ, it is found that the difference in the amplitude was caused by Λ even in the same M_{W} and X, which suggests that it is possible to establish an advanced ground motion model which takes into account much more regional characteristics than the ones in the previous studies.

Keywords: earthquake ground motion, Kanto region, machine learning, response spectrum, epicentral direction

1. Introduction

Recently, since ground motion observation stations have been deployed nationwide (e.g. [1]), we can obtain high-quality and huge data in real time. In addition, since high-speed calculations are becoming possible due to the remarkable evolution of computers, it is expected that knowledge on earthquakes and ground motions will improve in both quality and quantity. On the other hand, although many practical ground motion prediction equations (GMPEs) (e.g. [2]) have been developed and utilized, the data includes a bias in the number of earthquakes in each region and those of ground motion observation records at each station. Therefore, models constructed based on such data also include imbalance. Recent observation data growth may lead to an improvement in this problem.

In the future, we would like to realize an environment in which ground motion evaluation models (GMMs) are automatically generated by AI (Artificial Intelligence) using observation data obtained every time an earthquake occurs and are upgraded. The pioneering studies on constructing GMPEs using machine learning have also begun (e.g. [3], [4]).

The aim of this paper is to obtain clues to obtain new insights about earthquakes and ground motions. First, the data for this study were selected manually, and then preliminary studies on the construction of GMMs utilizing machine learning were performed. Here, not only the amplitude index of the ground motion but also the duration index [5] of the ground motion which was hardly considered in the conventional GMPEs was also evaluated. The epicentral direction was adopted as a feature because the previous studies (e.g. [6, 7]) have pointed out that the differences in the characteristics of ground motions due to the epicentral directions appeared in the observation records. Next, the influence of data quality (soundness of time history) on the evaluation results by machine learning was examined. Furthermore, we tried to visualize the characteristics such as magnitude dependency, hypocentral distance dependency, and epicentral direction dependency of the GMMs constructed by machine learning. Finally, the future issues to be examined were described.

2. Method and data for constructing GMMs

2.1 Outline of method for constructing GMMs

In this study, we tried to make preliminary GMMs by supervised machine learning using the ground motion observation records obtained in the past. The parameters describing the characteristic of a ground motion regarding the source, path, and site were used as features. The ground motion amplitude and duration indexes calculated from the observation records were used as the target variables.

As machine learning algorithm, a "Gradient Boosting Decision Tree [8]" that combined "Gradient Boosting" and "Decision Tree" was used. Gradient Boosting is a method of creating a strong classifier (high-performance machine learning model) by combining some weak classifiers (low-performance machine learning models). Decision Tree is a method for creating a classifier by branching criteria. Gradient Boosting Decision Tree is a method of combining weak classifier created by the decision tree by applying the gradient boosting [9]. Fig. 1 shows the schematic explanation of Gradient Boosting Decision Tree. In this study, XGBoost (eXtreme Gradient Boosting) [10], an open source library, was used.

2.2 Ground motion records and target variables

We used the horizontal ground motion records observed at 138 stations deployed in Kanto region (Tokyo, Kanagawa, Chiba, Saitama, Ibaraki, Tochigi, and Gumma) of the strong motion observation network K-NET of the National Research Institute for Earth Science and Disaster Resilience [1]. Fig. 2 shows the observation stations.

The target variables of the GMMs are the peak ground acceleration (PGA in cm/s^2), the 5% damped pseudo velocity response spectra (${}_pS_V$ in cm/s) for 5 periods (0.1, 0.5, 1, 3, 5 sec), and the 5% damped velocity response duration spectra (TS_V in sec) [6] for 5 periods (0.1, 0.5, 1, 3, 5 sec). The parameters defining the start and the end of TS_V are $p1$ = 0.03 and $p2$ = 0.95 [6]. PGA and ${}_pS_V$ were given as the common logarithm ($\log_{10}PGA$ and $\log_{10}{}_pS_V$).

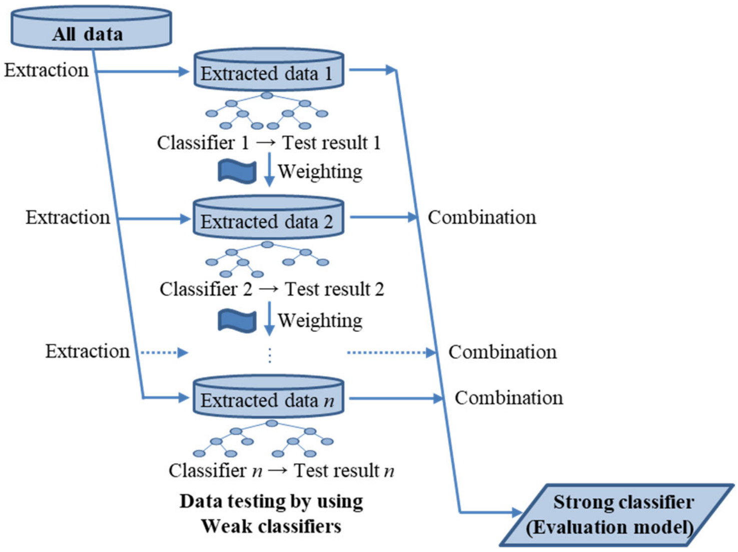


Fig. 1 – Schematic explanation of the Gradient Boosting Decision Tree

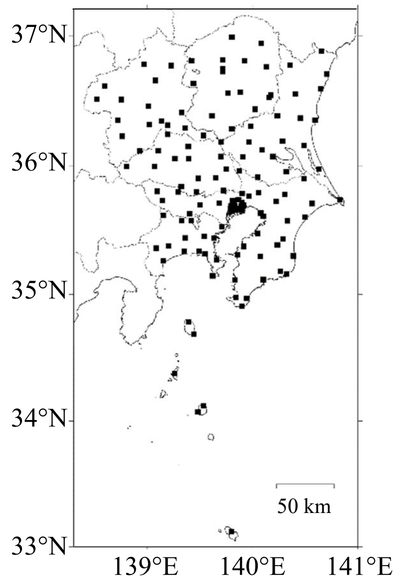


Fig. 2 – Locations of observation stations used in this study

2.3 Features of GMMs

We selected the features of the GMM based on the results of our trial and error preliminary studies. Specifically, the parameters that determine the ground motion characteristics, the moment magnitude (M_W), the hypocentral depth (H in km), the hypocentral distance (X in km), the epicentral direction (Λ in degree), the top depth of the seismic bedrock ($D28$ in m), and the averaged S-wave velocity in the surface layers of total thickness 30 m ($AVS30$ in m/s) were selected. These features were assumed to be independent of each other. Only those earthquakes whose M_W were determined by the NIED's broadband seismograph network F-net [1] were selected. As data needed to determine H, X, and Λ, we obtained the locations of hypocenters from the JMA [11], and the locations of stations from the K-NET [1].

Λ is set zero degrees to due north and is defined clockwise. Then, a set of sin Λ and cos Λ are inputted for the machine learning models because Λ is discontinuous at due north. $D28$ is the bottom depth of the 28th layer of the underground velocity structure model by the Japan Seismic Hazard Information Station (J-SHIS) [12] (i.e. the top depth of the layer of 5000 m/s for P-wave velocity and 2700 m/s for S-wave velocity). When the underground velocity structures up to a depth of 20 m are available at the observation station, the $AVS30$ is calculated by the equation by Morikawa and Fujiwara (2013) [2] using the average S-wave velocity of the surface layers with a total thickness 20 m. When the underground velocity structures are not available at the observation station, the $AVS30$ is replaced with the $AVS30$ disclosed by the J-SHIS [12].

The feature impact, the degree of deterioration of the estimation accuracy of the target variable when using the GMM created using training data which the specific feature column is rearranged at random, was used in order to determine which features were considered important in each GMM.

3. GMM for Kanto region, Japan

3.1 Earthquake and ground motion records

We selected the surface earthquake ground motion records obtained at 138 observation stations of the K-NET (Fig. 2) deployed in the Kanto region based on the following criteria:

1) Utilizing the JMA database, we extracted earthquakes with both a seismic intensity of 2 or more in Chiyoda-ku, Tokyo and a maximum seismic intensity of 4 or more nationwide from 1996 to January 15,

2019. Then, we manually selected earthquakes from them, considering both the distribution of magnitudes and that of source locations. Here, earthquakes with very few stations observed in Kanto region were excluded.

2) Magnitude of the earthquake (JMA magnitude M_J) was considered in three categories as follows: $3.0 < M_J \le 5.0$ (As a result, there was no data satisfying above-mentioned condition 1 for $M_J < 4.0$.), $5.0 < M_J \le 5.5$, and $5.5 < M_J$. The number of earthquakes in each category was also considered.

3) The deep earthquakes were also selected if the hypocentral depth varied much in the same hypocentral area.

4) The 2011 off the Pacific coast of Tohoku earthquake (M_J 9.0) was excluded because of its extremely large source area.

As shown in Fig. 3, the epicenters of the 74 selected earthquakes spread mainly in Kanto region, and those of some earthquakes are located far from Kanto region. From the time histories of surface ground motion records of the selected earthquakes, the horizontal ground motions (NS and EW components) at the observation station where the maximum acceleration combined with the three components is more than 1 cm/s^2 were selected. As a result, a total of 14104 records were selected for Data-set A.

3.2 Data-set for machine learning (Data-set A)

Fig. 4 shows the heatmaps of the number of the training data for the couples of features in Data-set A and the histograms corresponding those features. There are few data of earthquakes which $X > 300$ km and $M_W < 6$. Also, there are few data for earthquakes which $X < 200$ km and $M_W > 7$. It seems that the data seem to consist of three groups according to H. Their hypocentral depths appear to be dominated by the continental and oceanic plate depths. Λ for each observation station is generally distributed isotropically for $X < 150$ km, but is mostly distributed in the northeastern direction (Λ of around 45 degrees) for $X > 150$ km. Most of the observation stations where *D28* is deeper than 2500 m have *AVS30* less than 400 m/s. This is probably due to the deposition of the soft ground by the concentration of major rivers in the geological process that forms the bedrock around northern Tokyo Bay. Therefore, *D28* may not be independent of *AVS30*.

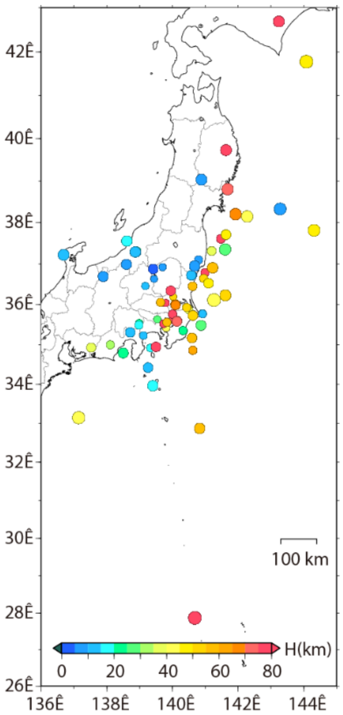
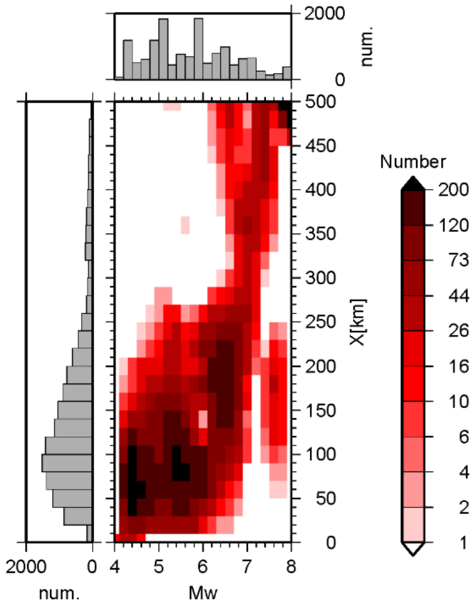
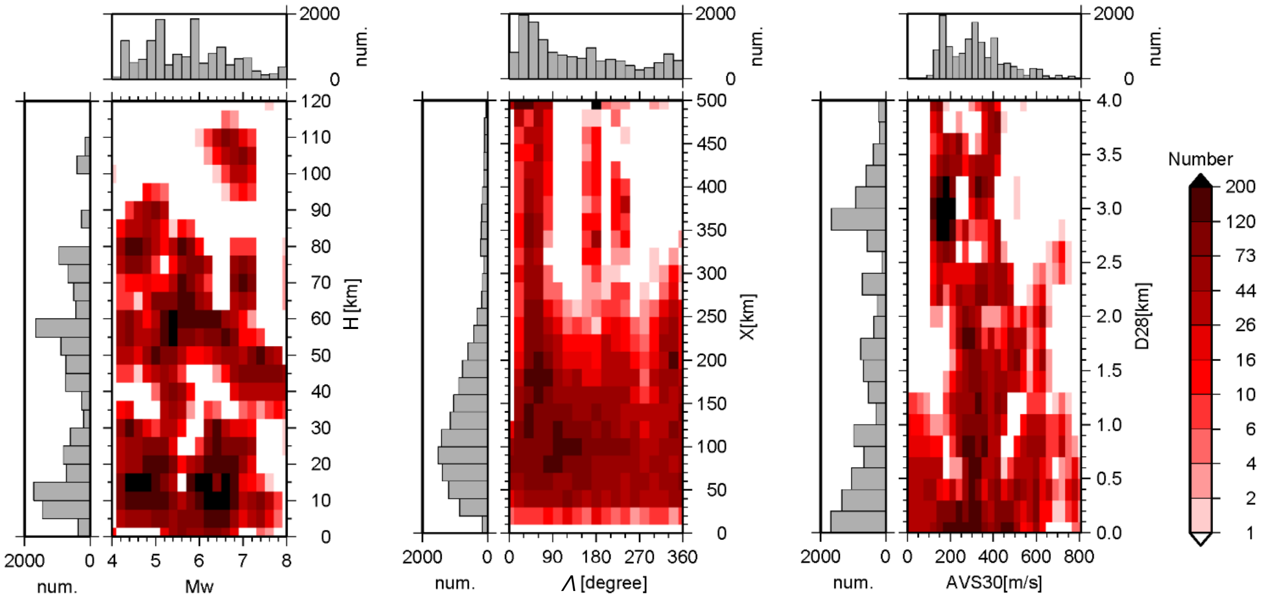


Fig. 3 – Epicenters of earthquakes used in this study



(a) M_W and X

Fig. 4 – Heatmaps of number of training data for the couples of features in Data-set A and histograms corresponding those features



(b) M_W and H (c) Λ and X (d) $AVS30$ and $D28$

Fig. 4 – Heatmaps of number of training data for the couples of features in Data-set A and histograms corresponding those features

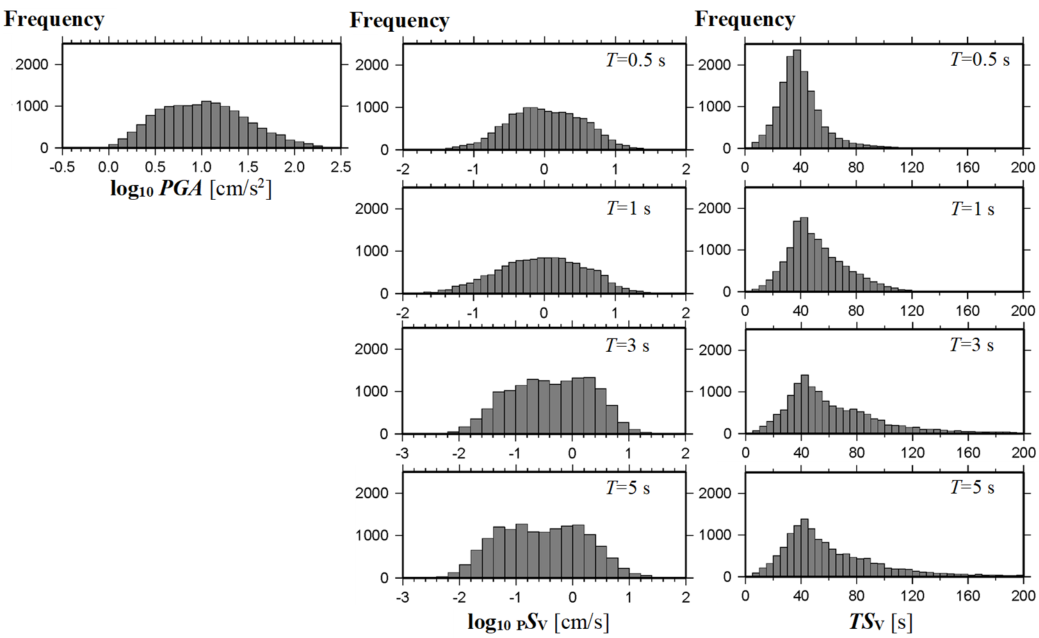


Fig. 5 – Histogram examples of the target variables ($\log_{10} PGA$, $\log_{10 p}S_V$, TS_V) of Data-set A

Fig. 5 shows the histograms examples of the target variables. The histograms of $\log_{10}PGA$ and $\log_{10p}S_V$ seem like a normal distribution. The histograms of TS_V are distributions biased to the short-duration side. Based on the above, as a loss function in the machine learning, the root mean squared error (normal distribution) was applied to $\log_{10}PGA$ and $\log_{10p}S_V$ and the Poisson deviance (Poisson distribution) was applied to TS_V.

3.3 GMM (Model A5 and Model A6)

Model A (A5 and A6) were created by supervised machine learning using Data-set A. Here, in order to investigate the effects of Λ, Model A6 considering all features (M_W, X, H, $D28$, $AVS30$, Λ) and Model A5

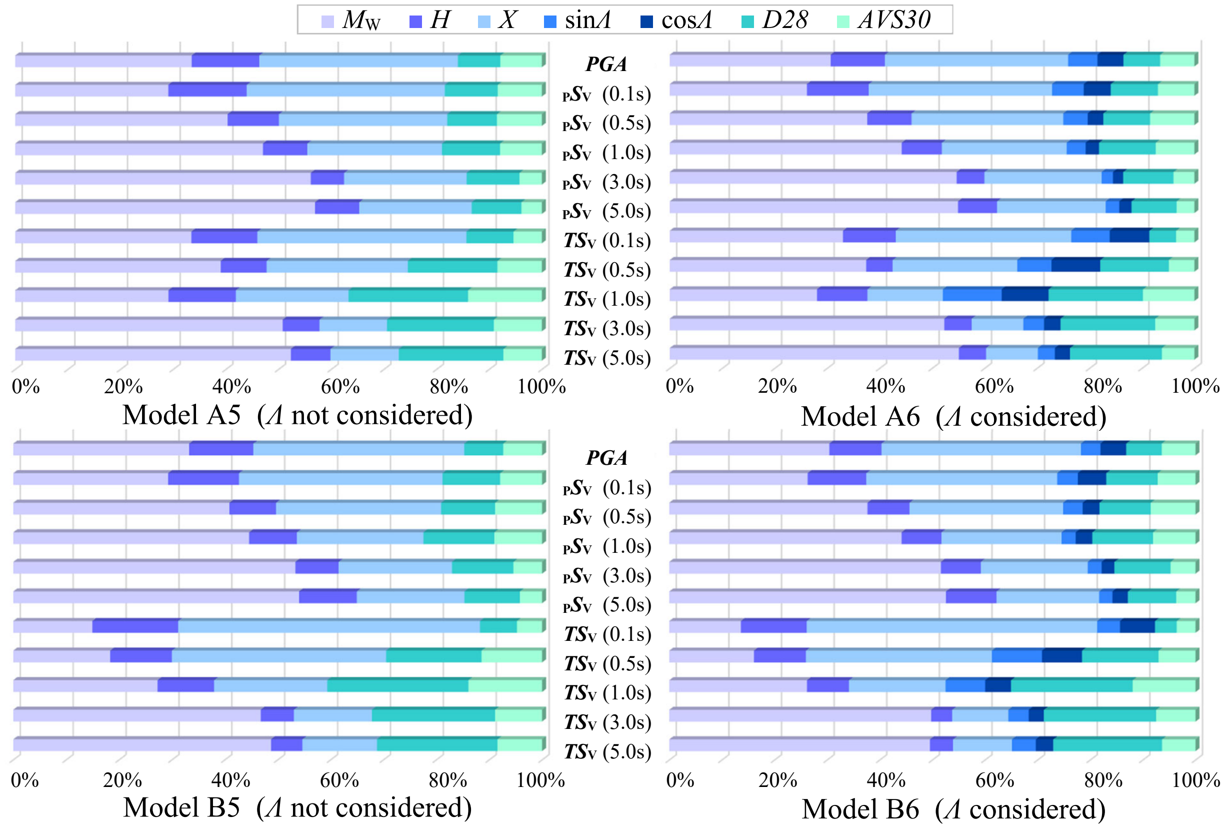


Fig. 6 – Feature impacts on the target variables of the GMMs

considering five features except for $\mathit{\Lambda}$ (M_W, X, H, $D28$, $AVS30$) were created and compared each other. In each model, both the index of the cross-validation results for the training data (64% of all data) and those for the validation data (36% of all data) were good and were almost same value, which indicated high versatility.

Fig. 6 shows the feature impacts on the target variables in Model A5 and Model A6. For the target variables except for TS_V for the period of 1 second, the feature impacts of M_W increased with period, and those of X decreased with period. The feature impacts of H increased slightly for the short period range, and those of $D28$ on TS_V increased for the period more than 1 second. The feature impact of $AVS30$ was relatively large for TS_V for the period of 1 second. The feature impact of $\mathit{\Lambda}$ is evaluated by the sum of sin$\mathit{\Lambda}$ and cos$\mathit{\Lambda}$. The feature impact of $\mathit{\Lambda}$ on ${}_pS_V$ and those on TS_V for the period of 1 second were relatively large. There were several target variables whose feature impacts of $\mathit{\Lambda}$ were comparable to that of H, that of $D28$, and that of $AVS30$. In general, it seems that the tendency of the feature impact changes in the period around 1 second.

Fig. 7 shows examples of the relationship between the observed value target variables and the evaluated ones based on Model A6. Here, the black plots indicate the training data used for constructing the model, and the red plots indicate the validation data. A solid line indicates that an evaluated value is equal to the observed one, and a broken line indicates that an evaluated value is double or half of the observed one. It can be seen that the model which can reproduce the observed values well has been constructed since most of the evaluated values are within double to half of the observed ones. Specifically, for PGA and ${}_pS_V$, the evaluated values were slightly smaller than the observed ones in the large-amplitude range, and slightly larger than the observed ones in the small-amplitude range. For TS_V, the evaluated values were slightly larger than the observed ones in the short-duration range, and the variation in the short-duration range was slightly larger than that in the long-duration range.

Fig. 8 shows examples of histograms of the ratio of the evaluated earthquake ground motion indexes to the observed ones. The ratio for all earthquake ground motion indexes were distributed like a normal distribution with an average of 1. The variations of TS_V were smaller than that of PGA and that of ${}_pS_V$. Table

1 shows the averages and the common logarithmic standard deviations of the ratios of the evaluated values to the observed ones. For *PGA* and ${}_{\mathrm{p}}S_{\mathrm{V}}$, the averages of the ratios were almost 1, and the common logarithmic standard deviations were around 0.18 to 0.19. For TS_{V}, the averages of the ratios were almost 1, and the

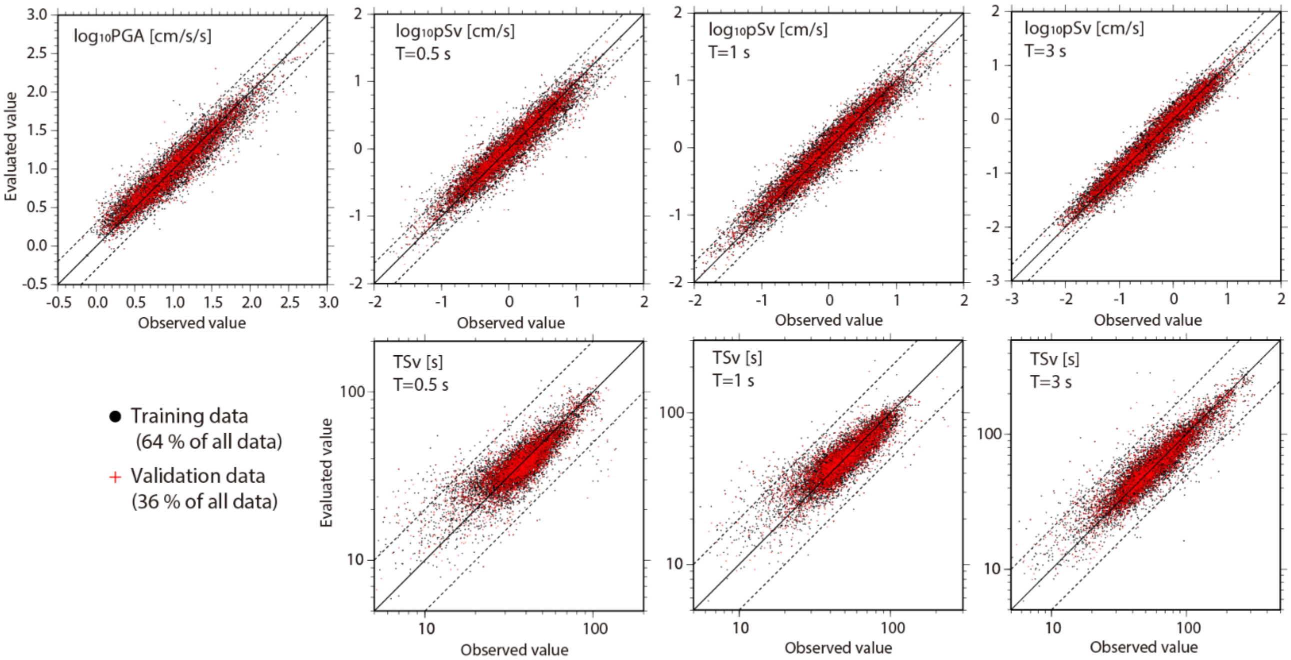


Fig. 7 – Examples of relationship between the observed target variables and the evaluated ones ($\log_{10} PGA$, $\log_{10}{}_{\mathrm{p}}S_{\mathrm{V}}$, TS_{V}) based on Model A6

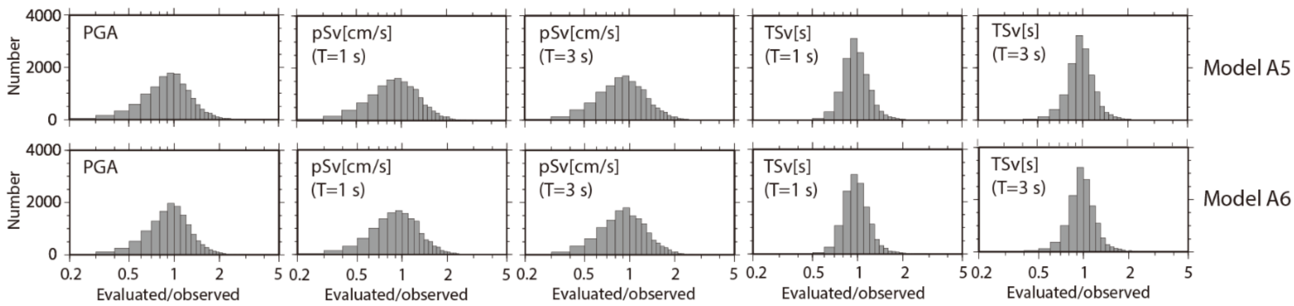


Fig. 8 – Examples of histograms of the ratios of the evaluated earthquake ground motion indexes (*PGA*, ${}_{\mathrm{p}}S_{\mathrm{V}}$, and TS_{V}) to the observed ones

Table 1 – Ratios of the evaluated earthquake ground motion indexes (*PGA*, ${}_{\mathrm{p}}S_{\mathrm{V}}$, TS_{V}) to the observed ones

Earthquake ground motion index	***PGA***	${}_{\mathrm{p}}S_{\mathrm{V}}$					TS_{V}				
Period [s]	—	0.1	0.5	1	3	5	0.1	0.5	1	3	5
Average μ of ratios of the evaluated to the observed											
Model A5 (Δ not considered)	1.08	1.08	1.09	1.08	1.07	1.07	1.06	1.05	1.05	1.05	1.05
Model A6 (Δ considered)	1.07	1.07	1.08	1.07	1.07	1.07	1.07	1.05	1.05	1.05	1.05
Model B5 (Δ not considered)	1.08	1.08	1.08	1.08	1.08	1.08	1.05	1.05	1.04	1.04	1.04
Model B6 (Δ considered)	1.07	1.07	1.08	1.07	1.07	1.08	1.04	1.05	1.04	1.04	1.03
Common logarithmic standard deviation σ of ratios of the evaluated to the observed											
Model A5 (Δ not considered)	0.18	0.19	0.20	0.19	0.18	0.19	0.12	0.11	0.11	0.11	0.11
Model A6 (Δ considered)	0.17	0.17	0.19	0.18	0.17	0.18	0.12	0.11	0.10	0.11	0.11
Model B5 (Δ not considered)	0.19	0.20	0.20	0.19	0.19	0.19	0.11	0.11	0.10	0.11	0.10
Model B6 (Δ considered)	0.17	0.18	0.19	0.18	0.18	0.18	0.11	0.10	0.10	0.10	0.10

common logarithmic standard deviations were around 0.11. The evaluation accuracies of most target variables in Model A6, which included *Λ* as a feature, were improved slightly compared to those in Model A5.

4. Necessity of data quality examination

4.1 Data-set considering data quality (Data-set B)

In this section, we examined the influences of training data quality on the results of the machine learning. In general, it is necessary to check the soundness of time histories as a preliminary study before ground motion analysis. In the future, we would like to automate this work with AI. However, since we didn't have such measures, we created Data-set B by manually selecting wholesome time histories from Data-set A. Model B constructed using Data-set B and Model A constructed using Data-set A were compared. The Data-set B time histories were selected from Data-set A focusing on the followings:

1) No abnormality is included in the time history. (As a result, there was no such record in this study.)
2) The time history, the Fourier amplitude spectrum (in the period range of 0.05-10 seconds), and the response spectrum (in the period range of 0.05-10 seconds) don't contain obvious noise.
3) The time history is recorded at the latest from the S-wave onset and is not interrupted in the middle of coda waves.
4) The time history does not include other events (e.g. reflections, refractions, aftershocks, etc.).

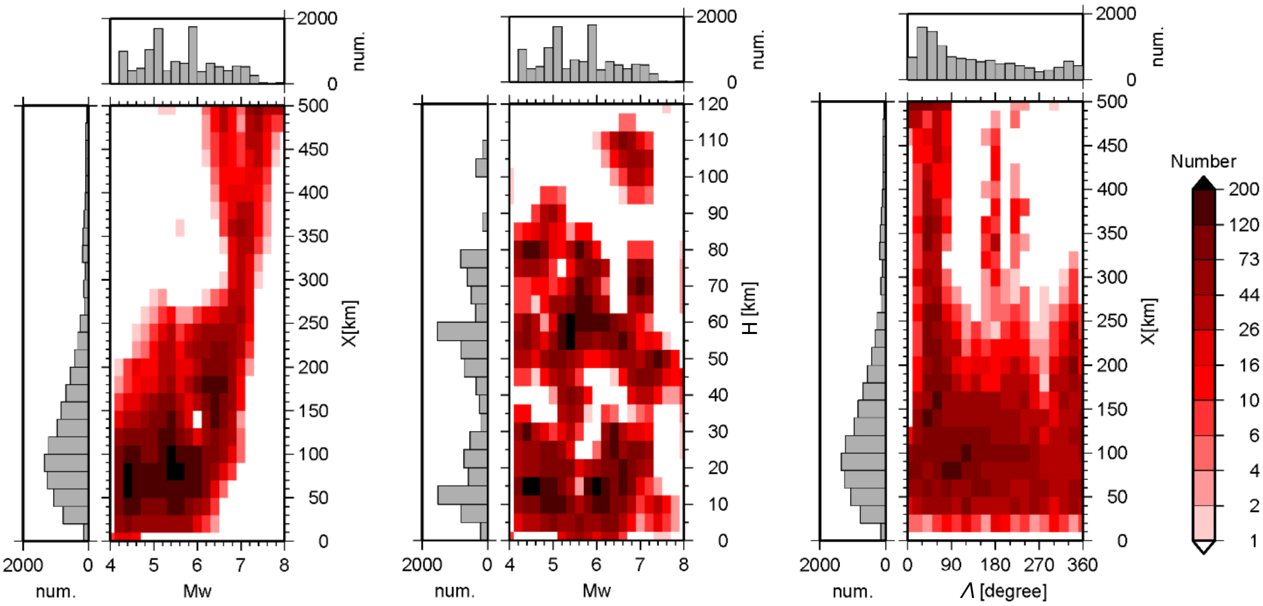
Although it also should be necessary to consider the velocity time histories, the filtered acceleration time histories, etc., only the nonfiltered acceleration time histories were considered in this study. As a result, (1) For M_W 4 class earthquakes, there were many ground motion records in which the noises included in the components of the period of 5 seconds or more. (2) The ground motion records with PGA < 2 cm/s^2 included large noises in their Fourier amplitude spectra. (3) There were many ground motion records in which the S-wave onset were not recorded or the coda waves were interrupted. (4) There were ground motion records which made it difficult to separate and exclude other earthquakes from the target earthquake. (5) There were ground motion records in which included multiple aftershocks. (6) There were ground motion records whose coda waves included another earthquake or the effects of suspected reflections or refractions. These data may have adversely affected the model A. In particular, the effects on the evaluations of TS_V (for all periods) and ${}_pS_V$ (for the period of 5 seconds) may not be negligible.

After appropriate processing, it is desirable that ground motion records with these problems will also be considered, but we excluded all these data from Data-set B in this preliminary study. As a result, a total of 11488 records were selected for Data-set B. Fig. 9 shows the heatmaps of the number of the training data for the couples of features in Data-set B and the histograms corresponding those features. The training data of $M_W \cong 7\sim8$ decreased from Data-set A (Fig. 4). The other heatmaps of Data-set B have not changed much from Data-set A.

4.2 GMM (Model B5 and Model B6)

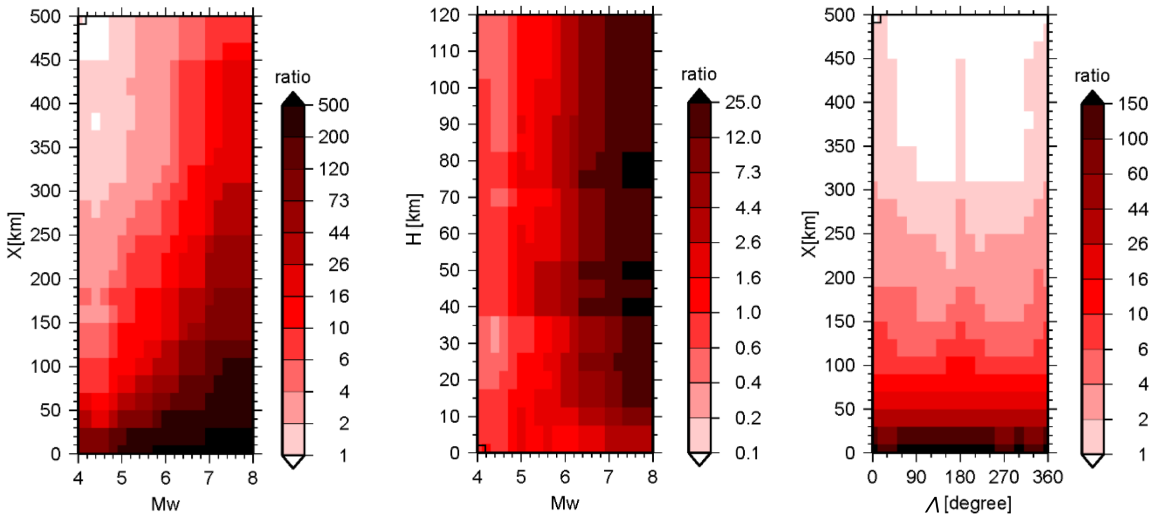
Model B (B5 and B6) were created by supervised machine learning using Data-set B. As in Model A, in order to investigate the effects of *Λ*, Model B6 considering all features (M_W, *X*, *H*, *D28*, *AVS30*, *Λ*) and Model B5 considering five features except for *Λ* (M_W, *X*, *H*, *D28*, *AVS30*) were created and compared each other. In each model, both the index of the cross-validation results for the training data (64% of all data) and those for the validation data (36% of all data) were good and were almost same value, which indicated high versatility.

Fig. 6 shows the feature impacts on the target variables in Model B5 and Model B6. As shown in the figure, the periodic characteristics of the feature impacts on TS_V in Model B became clearer than those in Model A. For ${}_pS_V$ and TS_V, as with Model A, the feature impacts of M_W increased with period, and those of *X* decreased with period. The feature impact of *AVS30* was relatively large on TS_V for the period of 1 second. Compared with Model A, the feature impacts of *D28* increased further on TS_V for the period more than 1



(a) M_W and X (b) M_W and H (c) Λ and X

Fig. 9 – Heatmap of number of training data for the couples of features in Data-set B and histograms corresponding those features



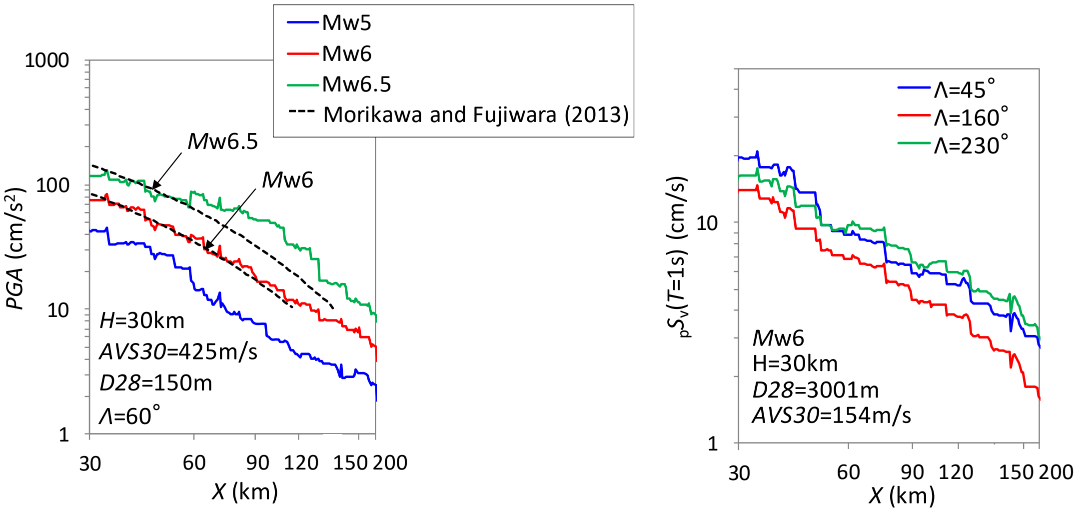
(a) M_W and X (b) M_W and H (c) Λ and X

Fig. 10 – Heatmaps of the ratio of the evaluated *PGA* to the reference *PGA* (The reference *PGA* is indicated by a square in each figure.)

second. The feature impacts of Λ were relatively large on $_{p}S_V$ and TS_V for the period of 1 second.

As shown in Table 1, for *PGA* and $_{p}S_V$, the averages of the ratios of the earthquake ground motion indexes were almost 1, and the common logarithmic standard deviations of the ones were around 0.18 to 0.19. For TS_V, the averages of the ratios were almost 1, and the common logarithmic standard deviations were around 0.10. As with Model A, the evaluation accuracies of most target variables in Model B6, which included Λ as a feature, were improved slightly compared to those in Model B5. Compared to the averages and the common logarithmic standard deviations of the earthquake ground motion indexes of Model A, those of TS_V of Model B improved slightly, but those of *PGA* and $_{p}S_V$ of Model B were almost the same.

From the above, it can be pointed out that in order to improve or stabilize the accuracy of a GMM, it is important to confirm the soundness of each time history in advance.



(a) Relationships of *PGA* and *X* or M_W (b) Relationships of $_{p}S_V$(T=1s) and *X* or *Λ*

Fig. 11 – Examples of the attenuation relationships of the earthquake motion indexes (*PGA* and $_{p}S_V$) and *X* in Model B6

5. Visualization of characteristics and estimated performance of Model B6

Generally speaking, since a repetition of non-linear calculations based on a machine learning algorithm cannot be represented by a clear physical equation, there is a problem that the reason why the model constructed by a machine learning derives the prediction result is a black box. However, it is possible to grasp the characteristics of the model from the relationships between a target variable and features based on the machine learning model. In this section, we tried to visualize the model characteristics and the estimation performance by illustrating the distributions of the evaluation results of the target variables when the two-dimensional features were given to Model B6.

5.1 Magnitude dependency and hypocentral distance dependency

Fig. 10(a) shows a heatmap of the ratio of the evaluated *PGA* when M_W is changed in increments of 0.2 and *X* is changed in increments of 20 km. Here, the conditions of the other specified features are $H \cong 30$ km, $AVS30 \cong$ 400 m/s, $D28 \cong 400$ m, and $\Lambda \cong 45$ degrees (northeast direction). Since the training data of the features used for the machine learning were sparse, the average of the evaluated values when *H*, *AVS30*, *D28*, and *Λ* were changed around the above-mentioned values were adopted to the evaluated model. The reference *PGA* as a denominator of the ratio was the evaluated *PGA* when $M_W \cong 4$ and $X \cong 500$ km. As shown in Fig. 9(a), there is almost no training data with $M_W \cong 7$~8 and $X \leq 300$ km. However, regardless of the presence or absence of data, the tendency that the evaluated values increase as M_W increases (i.e. magnitude dependency) and the tendency that the evaluated values increase as *X* decreases (i.e. hypocentral distance dependency) were expressed in Fig. 10. It is inferred that the characteristics learned based on the data of $M_W < 7$ were extrapolated to the range of $M_W \cong 7$~8. It is necessary to verify the certainty of the evaluated values in the extrapolated range in the future.

Fig. 11(a) shows examples of the attenuation relationships of *PGA* and M_W or *X* in the model B6. The given features were determined on the assumption that plate boundary earthquakes with the epicenter in the northeast direction would be observed at a bedrock site in Kanto. The evaluated values based on the GMPE (in case of M_W6 and M_W6.5) by Morikawa and Fujiwara (2013) [2] were also shown in Fig. 11(a) for reference. Here, since the original equation is using the root of the sum of squares of the maximum amplitude of horizontal two components, it was divided by $\sqrt{2}$. The evaluated values based on the equation were illustrated according to the applicable range of the equation ($M_W > 5.5$ and $PGA \geq 10$ cm/s^2). The attenuation relationships of the evaluated *PGA*s and M_W or *X* appear to be modeled in the roughly same as the conventional GMPEs. However, in the case of M_W 6.5, the tendency of attenuation seems to have changed at $X \cong 60$ km, which suggests that the possibility of creating a new GMM that reflects detailed regional characteristics by limiting the target region or considering Λ.

5.2 Hypocentral depth dependency

Fig. 10(b) shows a heatmap of the ratio of the evaluated *PGA* when M_W is changed in increments of 0.2 and H is changed in increments of 5 km. Here, the conditions of the other specified features are $X \cong 100$ km, $AVS30 \cong$ 400 m/s, $D28 \cong 400$ m, and $\Lambda \cong 45$ degrees (northeast direction). The procedure of creating the figure is the same as in the previous section. The reference *PGA* as a denominator of the ratio was the evaluated *PGA* when $M_W \cong 4$ and $H \cong 0$ km. As shown in Fig. 9(b), there is very little training data with $M_W \cong 7$~8. However, as shown in Fig. 10(b), the relationship between M_W and H for $M_W \leq 7$ seems to be reflected in the evaluated *PGA*s of $M_W \cong 7$~8. While the training data with $H \leq 100$ km were limited to $M_W \cong 6$~7, it is inferred that they were derived with reference to the evaluation results around the range where the features were missing. No clear hypocentral depth dependency appeared in this study. It is also necessary to consider the classification of earthquake types and additional training data in the future.

5.3 Epicentral direction dependency

Fig. 10(c) shows a heatmap of the ratio of the evaluated *PGA* when Λ is changed in increments of 20 degrees and X is changed in increments of 20 km. Here, the conditions of the other specified features are $M_W \cong 5$, $H \cong 30$, $AVS30 \cong 400$ m/s, and $D28 \cong 400$ m. The procedure of creating the figure is the same as in the previous section. The reference *PGA* as a denominator of the ratio was the evaluated *PGA* when $\Lambda \cong 0$ degree and $X \cong 500$ km. As shown in Fig. 10(c), while the epicentral direction dependency is not clearly seen for $X \leq 100$ km, is seen for $X > 100$ km. Although the bias of the training data as shown in Fig. 9(c) may be affected the evaluation results, it may be interpreted that the locational relationship between the Kanto Plain and the epicenter was reflected in the model.

Fig. 11(b) shows examples of the attenuation relationships of ${}_pS_V$ (for the period of 1 second) and X or Λ in Model B6. It is evaluated that the amplitude differs depending not only on M_W and X but also on Λ, which suggests that the possibility of creating a new GMM which reflects detailed regional characteristics by limiting the target region or considering Λ that has not been taken into account in previous attenuation relation formulas.

6. Conclusions

In this paper, the preliminary studies on the construction of GMMs utilizing machine learning were performed. For *PGA* and ${}_pS_V$, the averages of the ratios of the evaluated values to the observed ones were almost 1, and the common logarithmic standard deviations of the evaluated values to the observed ones were around 0.2. For TS_V, the averages of the ratios were almost 1, and the common logarithmic standard deviations were around 0.1. It can be seen that the GMMs which can reproduce the observed values well has been constructed since most of the evaluated values are within double to half of the observed ones. In particular, TS_V could be evaluate more accurately than *PGA* or ${}_pS_V$.

The evaluation accuracies of most target variables in the model considering Λ as a feature were improved slightly compared to those in the model not considering Λ as a feature. Therefore, it is significant to include Λ as a feature. Compared to the variations of TS_V of Model A in which the soundness of each time history was not confirmed, those of Model B in which the soundness of each time history was confirmed decreased slightly. It was pointed out that in order to improve or stabilize the accuracy of a GMM, it was important to confirm the soundness of each time history in advance.

We tried to visualize the model characteristics and the estimation performance. As a result, while the magnitude dependency, the hypocentral distance dependency, and the epicentral direction dependency could be confirmed, the hypocentral depth dependency was not apparent in this study. The attenuation relationships between *PGA*s and M_W or X, as evaluated by the machine learning, appeared to be modeled in the roughly like the conventional GMPEs. On the other hand, it also appeared the trends that differed from the conventional

GMPE, which suggests that the possibility of creating a new GMM which reflects detailed regional characteristics by machine learning.

In the future, it is necessary to take measures to improve the imbalance of the model accuracy caused by the training data density. In other words, it is necessary to study the method for extrapolation to huge magnitude earthquakes, large-amplitude ground motions, and long-duration ground motions that are hardly observed. One way to solve this problem is to use the simulation results from existing earthquake ground motion evaluation methods, and we plan to study them.

7. Acknowledgements

In this research, we used the earthquake information of Japan Meteorological Agency, the public data of MOWLAS (Monitoring of Waves on Land and Seafloor) of the National Research Institute for Earth Science and Disaster Resilience, and the machine learning tool DataRobot [13]. Some figures were created using GMT (The Generic Mapping Tools). We would like to thank everyone involved in their development, construction, operation and management.

8. References

[1] National Research Institute for Earth Science and Disaster Resilience: Monitoring of waves on land and seafloor; http://www.mowlas.bosai.go.jp/network/

[2] Morikawa N, Fujiwara H (2013): A new ground motion prediction equation for Japan applicable up to M9 mega-earthquake, *Journal of Disaster Research*, **8** (5), 878-888.

[3] Kubo H, Kunugi T, Suzuki S, Suzuki W, Aoi S (2018): Construction of ground motion prediction equation using random forest, *The 32nd annual conference of the Japanese Society for Artificial Intelligence*, 4Pin1-35 (in Japanese with English abstract).

[4] Kubo H, Kunugi T, Suzuki S, Suzuki W, Aoi S (2019): Attempt to reduce the effect of biased data-set on ground-motion prediction using machine learning, *The 33rd annual conference of the Japanese Society for Artificial Intelligence*, 4K2-J-13-02 (in Japanese with English abstract).

[5] Ishii T (2012): Response duration time spectra of earthquake motions, *Journal of structural and construction engineering*, Architectural Institute of Japan, **77** (676), 843-850 (in Japanese with English abstract).

[6] Ishii T (2008): A Study on response duration time spectra of earthquake motions in Tokyo, *The 14th World Conference on Earthquake Engineering*, Beijing, China, 02-0020.

[7] Arai K, Ishii T, Hirata N (2015): Study on spatial distribution of response duration time of earthquake motions in Tokyo metropolitan area, *Summaries of technical papers of annual meeting*, Architectural Institute of Japan, 15-16 (in Japanese).

[8] Friedman J.H (2001): Greedy function approximation: A gradient boosting machine, *The Annals of Statistics*, **29** (5), 1189-1232.

[9] Japan Deep Learning Association (2018): Official textbook for deep learning G (Generalist) test, Shoeisha Co., Ltd. (in Japanese).

[10] Chen T, Guestrin C (2016): XGBoost: A Scalable Tree Boosting System, *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 785-794.

[11] Japan Meteorological Agency: *Earthquake Information*; http://www.jma.go.jp/en/quake/

[12] National Research Institute for Earth Science and Disaster Resilience: Japan Seismic Hazard Information Station J-SHIS; http://www.j-shis.bosai.go.jp/en/

[13] DataRobot; https://app.datarobot.com/

