14th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP14
Dublin, Ireland, July 9-13, 2023

A Framework of Data-driven Seismic Performance Assessment of Structures using Earthquake Ground Motion Time Histories Generated by Machine Learning Techniques

Yuma Matsumoto
PhD student, Dept. of Architecture, The University of Tokyo, JSPS Research Fellow, Tokyo, Japan

Taro Yaoyama
Project Assist. Prof., Dept. of Architecture, The University of Tokyo, Tokyo, Japan

Sangwon Lee
PhD student, Dept. of Architecture, The University of Tokyo, Tokyo, Japan

Takenori Hida
Assoc. Prof., Major in Urban and Civil Engineering, Ibaraki University, Hitachi, Ibaraki

Tatsuya Itoi
Assoc. Prof., Dept. of Architecture, The University of Tokyo, Tokyo, Japan

ABSTRACT: This study proposes a surrogate model for seismic response analysis and an efficient sampling method for sampling critical ground motion time histories. Both methods are based on a deep generative model and Gaussian process regression, which enables utilization and management of high-dimensional time series of ground motions. Numerical experiments demonstrated that the surrogate model successfully predicted response analysis results with reasonable accuracy. In addition, by combining the above models with a Markov chain Monte Carlo method, it was possible to efficiently sample ground motion time histories capable of causing significant damage to structures.

1. INTRODUCTION

Dynamic response analysis using earthquake ground motion is an important part in the evaluation of seismic performance of buildings (Applied Technology Council (2018)). Considering the uncertainty of ground motions, input ground motion datasets are sometimes compiled with a large number of ground motion data obtained from a Monte Carlo simulation based method (Nishida et al. (2015)). The building performance is then evaluated against the compiled dataset in a probabilistic manner. Such an evaluation presents two main difficulties: (1) if the model of the structure is complex, performing response analysis on a large number of ground motion data is difficult because of computational cost, and (2) for computing the probability distribution of response, sampling input ground motions randomly from the dataset is inefficient because significant damage-causing ground motions to the structure are very limited. For reducing computational cost, a surrogate model has been proposed to replace the simulation of response analysis with machine learning techniques (Bucher and Bourgund (1990)). While superior in terms of computational cost, input variables of surrogate models are often low-dimensional and do not directly deal with the ground motion data itself. For sampling efficiency, Engelund and Rackwitz (1993) studied the

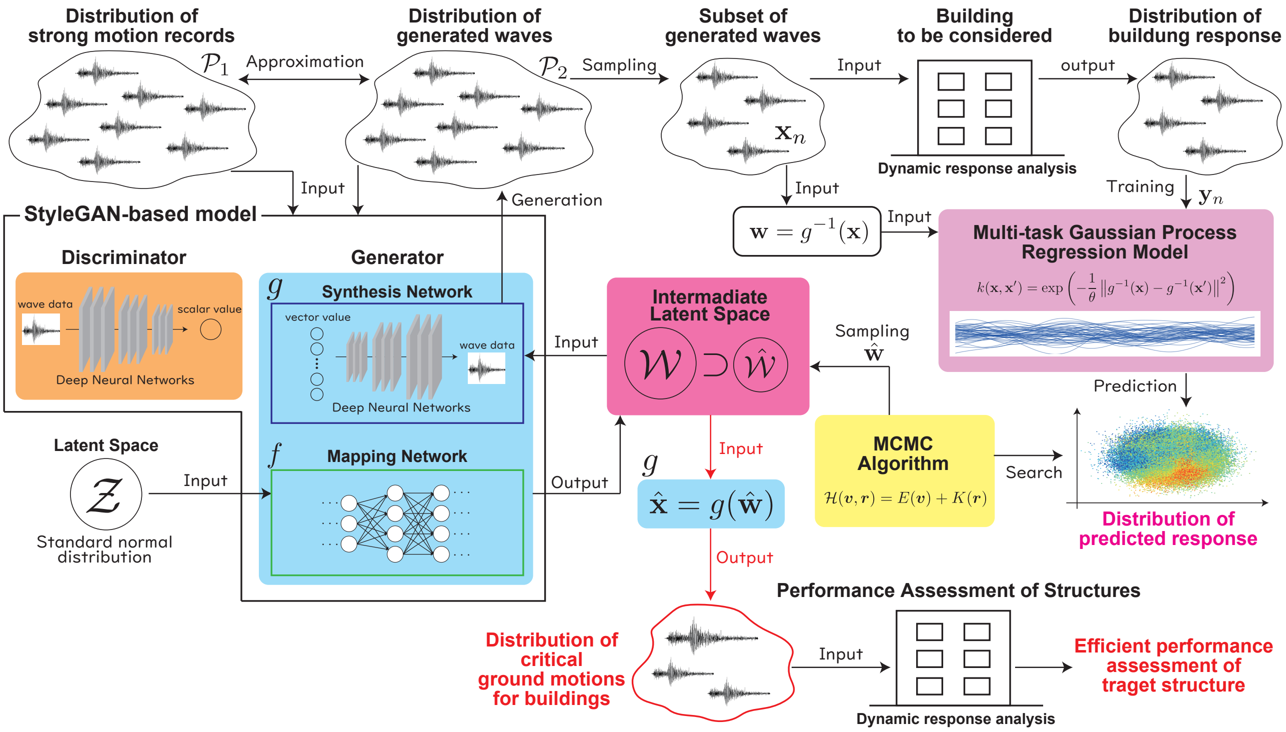

Figure 1: Overview of proposed framework.

application of several importance sampling methods to the evaluation of structural reliability. However, few studies have applied these methods to the sampling of ground motion time history data itself.

In this study, we have developed a surrogate model for dynamic response analysis that utilizes a deep generative model and Gaussian process regression. In the proposed surrogate model, input ground motion time histories are mapped by deep neural networks (DNN) into a low-dimensional feature space, called the latent space, and the mapped data is then used as an input to a Gaussian process regression model. The Gaussian process regression model then enables the surrogate model to indirectly consider ground motion time histories as input with relatively low computational cost. We also propose an efficient sampling method for sampling ground motion time histories that are capable of causing significant damage to structure using Markov chain Monte Carlo (MCMC) method based on the surrogate model predictions.

2. Overview of the proposed framework

Fig. 1 shows overview of our proposed framework. First, a dataset $\mathscr{P}_1$ for training deep generative model is compiled by collecting strong-motion observed records (Upper right of Fig. 1). $\mathscr{P}_1$ is compiled to represent the seismic hazard at the site of interest. For demonstration purpose, we constructed $\mathscr{P}_1$ consisting of one-component horizontal acceleration for crustal earthquakes in Japan, targeting K-NET (National Research Institute for Earth Science and Disaster Resilience (2023)) observation records with $M_W > 5$ and a fault distance of less than 100km (Itoi et al. (2014)). Then, a deep generation model called styleGAN2 (Karras et al. (2020)), a type of Generative Adversarial Network (GAN) (Goodfellow et al. (2014)), is used to construct a generative model capable of generating acceleration time histories. GANs can generate countless new data following a distribution that approximates the distribution of training dataset $\mathscr{P}_1$. StyleGAN2 is largely composed of two DNNs, a generator and a discriminator, and the generator is further divided into mapping network f and synthesis network g. The mapping network takes the latent variable $\mathbf{z} \in \mathscr{Z}$ as input and outputs the intermediate latent variable $\mathbf{w} \in \mathscr{W}$.

$$\mathbf{w} = f(\mathbf{z}), \quad \mathbf{z} \sim \mathscr{N}(\mathbf{0}, \mathbf{I}) \tag{1}$$

where $\mathbf{I}$ is a unit matrix, and $\mathbf{z}$ and $\mathbf{w}$ are both 512-dimensional vectors in this study. The synthesis network then considers $\mathbf{w}$ and certain constant values as input and generates ground motion time history $\mathbf{x}$.

$$\mathbf{x} = g(\mathbf{w}) \tag{2}$$

where $\mathbf{x}$ is acceleration time history data of 4096-dimensional vectors (duration 40.96 seconds). Generating a large number of ground motion data with the trained model yields a new dataset $\mathscr{P}_2$, whose probability distribution approximates the probability distribution of $\mathscr{P}_1$. The key difference between $\mathscr{P}_1$ and $\mathscr{P}_2$ is that each ground motion in $\mathscr{P}_2$ is associated with a lower-dimensional representation $\mathbf{w}$ via DNN of synthesis network g.

A surrogate model of the dynamic response analysis is then constructed using multi-task Gaussian process (MTGP) regression. A small number of seismic data are randomly sampled from $\mathscr{P}_2$, and a dynamic response analysis is performed for the building model under consideration. Then, a MTGP regression model is created with $\mathbf{w} = g^{-1}(\mathbf{x})$ as input variables, and indices of building response $\mathbf{y} \in \mathbb{R}^M$ calculated from the response analysis results as output variables. By making predictions with the trained regression model, response $\mathbf{y}'$ for a new input ground motion $\mathbf{x}'$ is predicted in a Bayesian framework.

Finally, samples of $\hat{\mathbf{w}}$ that result in greater response predictions $\mathbf{y}$ can be obtained using a MCMC method, and transformed by the synthesis network g into critical ground motions $\hat{\mathbf{x}}$ that can cause significant damages to the target building structure. These critical ground motions serve an efficient seismic performance assessment of structures.

3. METHODOLOGY

3.1. Gaussian process regression

Gaussian process is a type of stochastic process and is interpreted as an infinite-dimensional normal distribution over functions (Rasmussen and Williams (2006)). Let $\mathbf{x}$ be the input ground motion for the response analysis and y be the index of the corresponding response analysis result (e.g., peak response acceleration). The following training dataset $\mathscr{D}$ is created using N input data $\mathbf{x}_n$ and the corresponding response y_n:

$$\mathscr{D} = \{(\mathbf{x}_i, y_i) \mid i = 1, \cdots, N\} \tag{3}$$

Suppose, that the relationship between $\mathbf{x}_n$ and response y_n can be expressed using the function f as:

$$y_n = f(\mathbf{x}_n) + \varepsilon_n, \qquad \varepsilon_n \sim \mathscr{N}(0, \beta^{-1}) \tag{4}$$

where ε_n is a observation noise and β is a precision parameter. When the function f is generated following a Gaussian process with mean 0 and a covariance function $k(\mathbf{x}, \mathbf{x}')$ as $f \sim \mathrm{GP}(0, k(\mathbf{x}, \mathbf{x}'))$, the joint distribution of the vector $\mathbf{y} = (y_1, \cdots, y_N)^{\mathrm{T}}$ is the normal distribution as (Bishop (2006)):

$$p(\mathbf{y} \mid \mathbf{X}, \beta^{-1}) = \mathscr{N}(\mathbf{y} \mid \mathbf{0}, \mathbf{C}) \tag{5}$$

where $\mathbf{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_N\}$, and the covariance matrix $\mathbf{C}$ is obtained as follows:

$$\mathbf{C} = \mathbf{K} + \frac{1}{\beta}\mathbf{I} \tag{6}$$

$$\mathbf{K} = \begin{pmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_1) & \cdots & k(\mathbf{x}_N, \mathbf{x}_N) \end{pmatrix} \tag{7}$$

where $\mathbf{I}$ is a $N \times N$ unit matrix, and $\mathbf{K}$ is a $N \times N$ matrix whose (i, j) components are covariance function $k(\mathbf{x}_i, \mathbf{x}_j)$, called the Gram matrix. The covariance function $k(\mathbf{x}_i, \mathbf{x}_j)$ evaluates the similarity of the two inputs $\mathbf{x}_i$ and $\mathbf{x}_j$, and is called the kernel function. Gaussian processes are equivalent to Bayesian linear models, but they have the advantage of not requiring explicit setting of basis functions.

Next, consider predicting the response y_* when new input ground motion data $\mathbf{x}_*$ is obtained. The joint distribution of $\mathbf{y}$ and y_* is given as follows:

$$\begin{pmatrix} \mathbf{y} \\ y_* \end{pmatrix} \sim \mathscr{N}\left(\begin{pmatrix} \mathbf{0} \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbf{C} & \mathbf{k}_* \\ \mathbf{k}_*^{\mathrm{T}} & c_{**} \end{pmatrix} \right) \tag{8}$$

where $\mathbf{k}_*$ is the N-dimensional vector whose n-th component is $k(\mathbf{x}_n, \mathbf{x}_*)$, and $c_{**} = k(\mathbf{x}_*, \mathbf{x}_*) + \beta^{-1}$. The conditional distribution $p(y_* \mid \mathbf{x}_*, \mathscr{D})$ is also the normal distribution, and is obtained as:

$$p(y_* \mid \mathbf{x}_*, \mathscr{D}) = \mathscr{N}\left(\mathbf{k}_*^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{y}, c_{**} - \mathbf{k}_*^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{k}_*\right) \tag{9}$$

Thus, the predictive distribution of y_* is expressed by a kernel function $k(\cdot, \cdot)$.

3.2. Multi-task Gaussian process regression

Bonilla et al. (2007) proposed a method for extending Gaussian process regression to multi-task learning. Let $\mathbf{y} \in \mathbb{R}^M$ be the vector-valued response to a ground motion data $\mathbf{x}$, and suppose a dataset $\mathscr{D}_2$ consisting of N data is obtained.

$$\mathscr{D}_2 = \{(\mathbf{x}_i, \mathbf{y}_i) \mid i = 1, \cdots, N\} \tag{10}$$

As in section 3.1, suppose that

$$\mathbf{y}_n = \mathbf{f}(\mathbf{x}_n) + \varepsilon_n, \qquad \varepsilon_n \sim \mathscr{N}(\mathbf{0}, \Sigma) \tag{11}$$

where Σ is a $M \times M$ diagonal matrix whose i-th diagonal component is a precision parameter β_i^{-1}, and the vector-valued function $\mathbf{f}$ is assumed to follow a Gaussian process. Then, the joint probability distribution of $\mathbf{y}_2 = (y_{11}, \cdots, y_{N1}, y_{12}, \cdots, y_{N2}, \cdots, y_{1M}, \cdots, y_{NM})^{\mathrm{T}}$, where y_{nm} denotes the m-th component of the n-th response $\mathbf{y}_n$, is obtained as follows:

$$p(\mathbf{y}_2 \mid \mathbf{X}, \Sigma) = \mathscr{N}(\mathbf{y}_2 \mid \mathbf{0}, \mathbf{C}_2) \tag{12}$$

$$\mathbf{C}_2 = \mathbf{K}^f \otimes \mathbf{K} + \Sigma \otimes \mathbf{I} \tag{13}$$

where $\mathbf{K}^f$ is a $M \times M$ positive semi-definite matrix which specifies the inter-task similarities. Response $\mathbf{y}_*$ to new input ground motion $\mathbf{x}_*$ can also be predicted by the same procedure as in section 3.1. If L new input ground motions $\mathbf{X}^* = \{\mathbf{x}_1^*, \cdots, \mathbf{x}_L^*\}$ are obtained, the corresponding response $\mathbf{Y}^* = \{\mathbf{y}_1^*, \cdots, \mathbf{y}_L^*\}$ can be predicted considering the correlations of each data of $\mathbf{X}^*$. Let $\mathbf{y}^* = (y_{11}, \cdots, y_{L1}, y_{12}, \cdots, y_{L2}, \cdots, y_{1M}, \cdots, y_{LM})^{\mathrm{T}}$, then the joint distribution of $\mathbf{y}_2$ and $\mathbf{y}_*$ is given as:

$$\begin{pmatrix} \mathbf{y}_2 \\ \mathbf{y}^* \end{pmatrix} \sim \mathscr{N}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{K}^f \otimes \mathbf{K} + \Sigma \otimes \mathbf{I} & \mathbf{K}^f \otimes \mathbf{K}_* \\ \mathbf{K}^f \otimes \mathbf{K}_*^{\mathrm{T}} & \mathbf{K}^f \otimes \mathbf{K}_{**} + \Sigma \otimes \mathbf{I}_L \end{pmatrix} \right) \tag{14}$$

where $\mathbf{I}_L$ is the $L \times L$ unit matrix, and

$$(\mathbf{K}_*)_{nl} = k(\mathbf{x}_n, \mathbf{x}_l^*) \qquad l = 1, \cdots, L \tag{15}$$

$$(\mathbf{K}_{**})_{ll'} = k(\mathbf{x}_l^*, \mathbf{x}_{l'}^*) \qquad l' = 1, \cdots, L \tag{16}$$

Then the conditional distribution is obtained as follows:

$$p(\mathbf{y}_* \mid \mathbf{X}_*, \mathscr{D}_2, \Sigma) = \mathscr{N}(\mathbf{m}_*(\mathbf{X}_*), \mathbf{S}_*(\mathbf{X}_*)) \tag{17}$$

where

$$\mathbf{m}_*(\mathbf{X}_*) = \left(\mathbf{K}^f \otimes \mathbf{K}_*^{\mathrm{T}}\right) \mathbf{C}_2^{-1} \mathbf{y}_2 \tag{18}$$

$$\mathbf{S}_*(\mathbf{X}_*) = \mathbf{K}^f \otimes \mathbf{K}'_{**} + \Sigma \otimes \mathbf{I}_L - \left(\mathbf{K}^f \otimes \mathbf{K}_*^{\mathrm{T}}\right) \mathbf{C}_2^{-1} \left(\mathbf{K}^f \otimes \mathbf{K}_*\right) \tag{19}$$

3.3. Kernel function design

3.3.1. Basic concept

As the prediction of MTGP regression depends on the kernel function, appropriate kernel function design becomes an important step. According to Bishop (2006), the kernel function $k(\mathbf{x}, \mathbf{x}')$ must be an appropriate measure of the similarity between $\mathbf{x}$ and $\mathbf{x}'$. However, the acceleration time histories used in response analysis are typically high-dimensional, which make it difficult to construct a function that directly measures their similarity.

Thus, we propose a kernel function design that focuses on the latent space of the deep generative model. Jaakkola and Haussler (1998) proposed a kernel function, named the Fisher kernel, defined using a generative probability model $p(\mathbf{x})$. With reference to the construction method of the Fisher kernel, we propose the following kernel to measure the similarity of ground motion time histories:

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1}{\theta} \left\| g^{-1}(\mathbf{x}) - g^{-1}(\mathbf{x}') \right\|^2\right) \tag{20}$$

where θ is the hyperparameter of the kernel.

3.3.2. Effectiveness of our kernel function

Our proposed kernel determines that when the Euclidean distance between intermediate latent variables $\mathbf{w}$ and $\mathbf{w}'$ is short, the corresponding generated ground motions $\mathbf{x}$ and $\mathbf{x}'$ are similar. This determination is based on the following regularization error introduced into the loss function when training the generator (Karras et al. (2020)):

$$\mathscr{L}_r = \mathbb{E}_{\mathbf{w}, \gamma \sim \mathscr{N}(\mathbf{0}, \mathbf{I})} \left(\left\| \mathbf{J}_{\mathbf{w}}^{\mathrm{T}} \gamma \right\| - a \right)^2 \tag{21}$$

where γ is a random variable that follows a standard normal distribution, $\mathbf{J}_{\mathbf{w}}$ is the Jacobian matrix: $\mathbf{J}_{\mathbf{w}} = \partial g(\mathbf{w}) / \partial \mathbf{w}$, and a is a constant. Minimizing the value of Eq. (21) would encourage that a fixed-size step in the intermediate latent space $\mathscr{W}$ results in a non-zero, fixed-magnitude change in the generated data $\mathbf{x}$ (Karras et al. (2020)). We confirmed that the regularization error in Eq. (21) is sufficiently small when training the model. The other requirement for Eq. (20) to be a valid kernel is that the Gram matrix $\mathbf{K}$ should be a positive semi-definite matrix (Shawe-Taylor and Cristianini (2004)), and this requirement is confirmed during the analysis.

3.4. *Sampling method by MCMC*

Hamiltonian Monte Carlo (HMC) (Duane et al. (1987)) is used as the MCMC algorithm. It samples intermediate latent variables $\mathbf{w}$ of critical ground motions yielding greater response $y \in \mathbb{R}$. For computational efficiency, HMC samples a lower-dimensional representation of $\mathbf{w}$, denoted as $\mathbf{v} \in \mathbb{R}^Q, (Q < 512)$, which is obtained by principal component analysis (PCA) as follows:

$$\mathbf{w}_j \simeq \sum_{q=1}^{Q} \left(\mathbf{v}_j\right)_q \mathbf{u}_q \tag{22}$$

where $\mathbf{u}_q$ is the q-th principal component of the dataset $\{\mathbf{w}_n\}$. The target probability density of $\mathbf{v}$ for the HMC sampling is set such that the larger values of y have larger probability density as follows:

$$\mathscr{L}_y(\mathbf{v}) = \frac{1}{Z} \exp\left(-\frac{\lambda}{\zeta(\mathbf{v})}\right) \tag{23}$$

$$y = \zeta(\mathbf{v}) \tag{24}$$

where Z is a normalization constant, and λ is a hyperparameter. For a sampled $\hat{\mathbf{v}}$, the corresponding y can be predicted using Eq. (17), and the mean of the predicted distribution is used for $\zeta(\hat{\mathbf{v}})$. The Hamiltonian H used in HMC is set as follows:

$$H(\mathbf{v},\mathbf{r}) = E(\mathbf{v}) + \frac{1}{2}\mathbf{r}^{\mathrm{T}}\mathbf{r}, \qquad \mathbf{r} \sim \mathcal{N}(\mathbf{0},\mathbf{I}) \tag{25}$$

$$E(\mathbf{v}) = -\log \mathscr{L}_y(\mathbf{v}) \tag{26}$$

Numerical integration of the Hamiltonian equation is performed by leapfrog discretization. As only the kernel function contained in $\mathbf{K}_*$ depends on $\mathbf{v}$, the gradient of $E(\mathbf{v})$ can be calculated as:

$$\frac{\partial E(\mathbf{v})}{\partial v_q} = -\frac{\lambda}{[\zeta(\mathbf{v})]^2} \frac{\partial \zeta(\mathbf{v})}{\partial v_q} \tag{27}$$

$$\frac{\partial k(\mathbf{w}_n, \hat{\mathbf{w}})}{\partial v_q} = \frac{2}{\theta} k(\mathbf{w}_n, \hat{\mathbf{w}}) \left(\mathbf{w}_n^{\mathrm{T}}\mathbf{u}_q - \sum_{\phi=1}^{Q} v_\phi \mathbf{u}_\phi^{\mathrm{T}}\mathbf{u}_q\right) \tag{28}$$

where $\hat{\mathbf{w}}$ is the intermediate latent variable corresponding to the sampled $\hat{\mathbf{v}}$.

4. ANALYSIS RESULTS

4.1. *Analysis conditions*

A styleGAN-based model was trained on the previously mentioned dataset of crustal earthquakes, and 100,000 different ground motions were generated by the trained model. From those data, 1000 ground motions were randomly sampled for $\mathscr{D}_2$ ($N = 1000$), and the other 99,000 were used as test data. To conduct a fundamental study, the building model used in our experiments was a linear single degree of freedom system with a natural period $T = 0.2$s and a damping factor $h = 3\%$. The indices of response predicted by the MTGP regression model were the peak absolute acceleration response y_1 and the maximum relative displacement y_2 ($M = 2$).

4.2. *Estimation of the hyperparameters*

In the regression model of Eq. (17), $\mathbf{K}^f$, θ, and Σ are hyperparameters and should be set to appropriate values. First, as simulated response without noise is used in this study, the observation noise is assumed zero and $\Sigma = \mathbf{O}$. Then, according to Bonilla et al. (2007), we obtained θ that maximizes the following value of $\mathscr{L}_\theta$ by the gradient descent.

$$\mathscr{L}_\theta = -M \log|\mathbf{K}| - N \log\left|\mathbf{Y}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{Y}\right| \tag{29}$$

where $\mathbf{Y}$ is a $N \times M$ matrix whose n-th row is $\mathbf{y}_n^{\mathrm{T}}$. The optimal θ is obtained as 71.7 in our experiment. After finding the optimal θ, $\mathbf{K}^f$ was obtained as follows:

$$\mathbf{K}^f = \frac{1}{N}\mathbf{Y}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{Y} \tag{30}$$

4.3. *Regression analysis results*

The examples of predictive distribution of $\mathbf{y}_*$ for $\mathbf{x}_*$ in test data are shown in Fig. 2 in comparison with the actual response analysis results. The difference between predicted mean and the actual response are mostly less than standard deviation, indicating that the predictions are valid. The green line is the ellipse representing the covariance matrix between absolute acceleration and relative displacement, although the minor axis is almost zero because of strong correlation. This is because of the linearity of response analysis model.

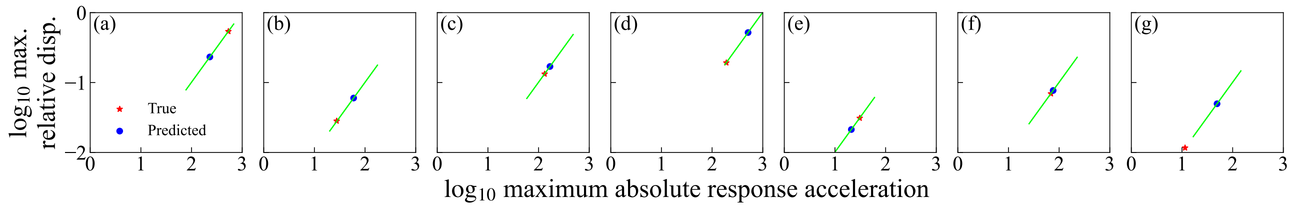

Figure 2: Prediction results by MTGP model. The red stars are the response analysis results, the blue dots are the prediction results, and the green line is an ellipse representing the 1σ region obtained from the covariance matrix. Both axes show values with common logarithm, and the original unit is $\mathrm{cm/s^2}$.

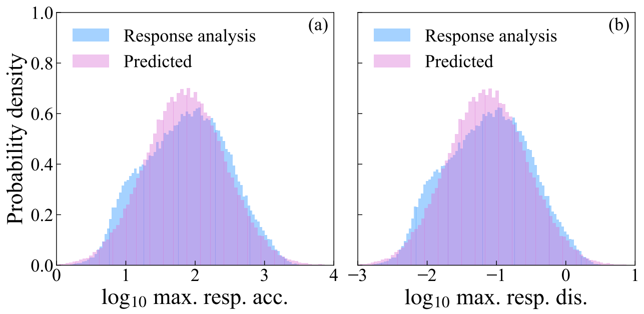

Figure 3: Comparison of the distribution of predicted results by MTGP regression model (red) and response analysis results (blue): (a) is for the peak absolute acceleration response, and (b) is for the maximum relative displacement.

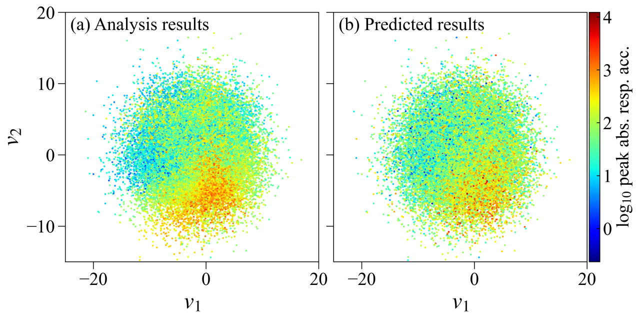

Figure 4: Distribution of the peak absolute acceleration response values in the v_1-v_2 plane: (a) is the response analysis results, and (b) is the predicted results by constructed surrogate model.

The estimated response values $\hat{\mathbf{y}}_i^*$ were then sampled from the predicted normal distribution for each point of test data. For validation, response analyses were also performed on the test data to calculate the true value of $\mathbf{y}_i^*$. Fig. 3 compares the distribution of $\mathbf{y}_i^*$ and $\hat{\mathbf{y}}_i^*$. For both absolute acceleration and relative displacement, there is good agreement between the overall distribution shape of $\mathbf{y}_i^*$ and $\hat{\mathbf{y}}_i^*$, including the tails of the distributions, suggesting that reasonable estimation is possible even in regions with a small number of data.

4.4. Sampling of critical ground motions

The styleGAN-based model and surrogate model constructed previously are used to sample ground motion time histories that are capable of causing significant damage to structures. The intermediate latent variable $\mathbf{w}$ was transformed by PCA into $\mathbf{v} \in \mathbb{R}^Q$, and the $\hat{\mathbf{v}}$ is sampled by HMC such that the peak absolute acceleration response is large. To demonstrate the effectiveness of dimension reduction by PCA, a comparison between the distribution of the response analysis results and that of the prediction results in the v_1-v_2 plane is shown in Fig. 4. Both distributions are similar, and there is a region where the response is large in the lower part ($v_2 < 0$) of each plot. Thus, efficiently obtaining ground motion data that will result in a larger response by sampling $\hat{\mathbf{v}}$ based on the surrogate model predictions is possible.

As values that enable stable sampling and sufficient accuracy in recovering $\hat{\mathbf{w}}$, we set $Q = 100$ and $\lambda = 0.8$, and $\hat{\mathbf{v}}$ was sampled at 11000 points by HMC. The burn-in interval was set to 2000, and 9000 samples of $\hat{\mathbf{w}}$ were obtained using equation (22), and then input to a styleGAN-based model to generate acceleration time history $\hat{\mathbf{x}} = g(\hat{\mathbf{w}})$. Examples of time history waveforms and Fourier spectra of the sampled ground motions are shown in Fig. 5. Fig. 6 also shows the waveforms of the absolute acceleration response obtained from the response analysis with the four ground motions shown in

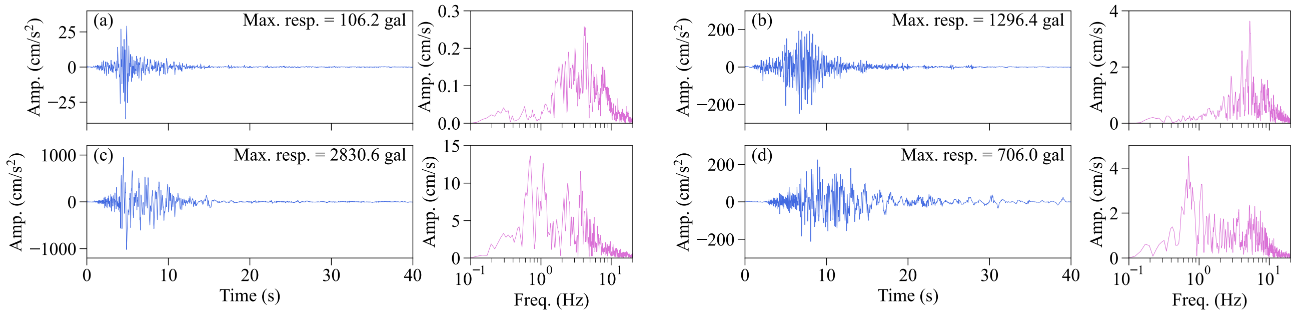

Figure 5: Examples of sampled ground motion time history data and its Fourier spectrum. Fourier spectra were obtained by FFT after removing frequency components below 0.1 Hz with a high-pass filter. In the upper right corner of each waveform figure, the peak absolute acceleration response values are shown when a response analysis is performed using that time history data.

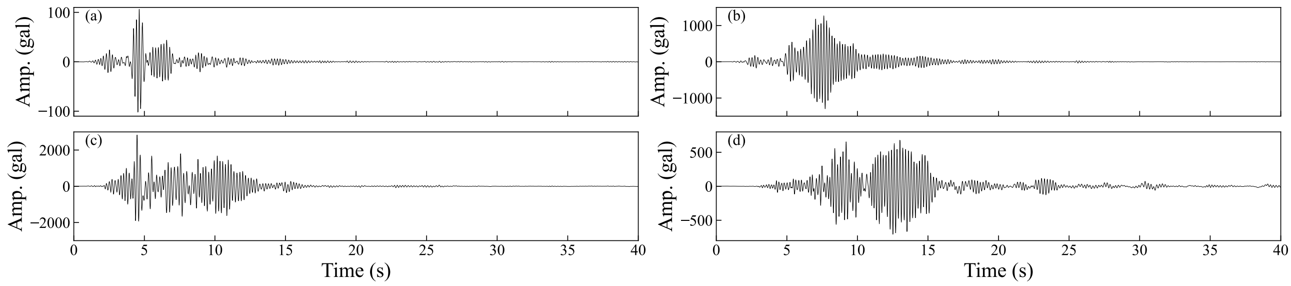

Figure 6: Absolute acceleration response waveforms obtained from the response analysis inputting the ground motion data in Fig. 5, respectively. The letters in the upper left of the figure represent the corresponding input and response, respectively

Fig. 5. Various ground motion time histories are sampled. Observing the waveforms in Fig. 5 (a) and (b), the peak ground acceleration (PGA) values are not large, while the spectrum shows a dominance around 5 Hz, which is natural frequency of the building model. Thus, the response is amplified significantly compared to the PGA of the input. Conversely, the Fourier spectrum of Fig. 5 (c) does not show a clear predominance around 5 Hz, but the response was large due to the large PGA of the input. Then, focusing on the relationship between the input and response, in Fig. 6 (a)-(c), the amplitude of the response is maximum near the time when the input amplitude is maximum, but in Fig. 6 (d), the peak response occurs around 13s, which does not correspond to the time of PGA. This may be caused by subsequent vibrations around 13s that are consistent with the natural frequency of the building model. As described above, the ground motion time histories can be sampled in a manner that the response of the building model becomes large due to various factors.

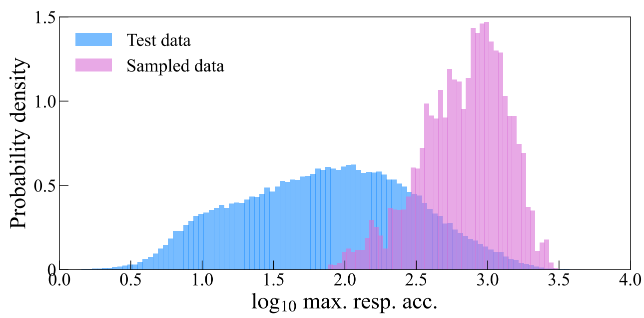

Figure 7: Comparison of response analysis results for ground motions in test data and sampled ground motions.

Finally, response analysis is performed on all sampled critical ground motions. Fig. 7 shows the distribution of peak absolute acceleration responses to the sampled ground motions in comparison with responses to all ground motions in test data. The sampled ground motion data is concentrated in the region of larger response, confirming the effectiveness of the proposed method.

5. CONCLUSIONS

In this study, we constructed a surrogate model for dynamic response analysis of structures using a deep generative model and multi-task Gaussian process regression. Utilizing the surrogate model, we also proposed a efficient sampling method for sampling ground motion time histories that can cause significant damage to structures. The novelty and superiority of these methods is that they utilize and manage time history data of earthquake ground motions, and reduce biases caused by oversimplification. The proposed method allows efficient assessment of the seismic performance of structures with a small number of response analyses. The used styleGAN-based model is a probabilistic model, and the Gaussian process regression can be interpreted in the framework of Bayesian probability. It is expected that performance assessment of structures is possible while probabilistically evaluating time history data of earthquake ground motions. However, a linear building model was used in this study to perform the fundamental study. Studying the applicability of our framework is important in nonlinear dynamic response analysis.

6. ACKNOWLEDGEMENT

This study was supported by JSPS KAKENHI Grant Number JP22J23006 and funded by MEXT No. JPMXD0222682433.

7. REFERENCES

Applied Technology Council (2018). *Seismic Performance Assessment of Buildings (FEMA P-58).*

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York :Springer.

Bonilla, E. V., Chai, K., and Williams, C. (2007). “Multi-task gaussian process prediction.” *Advances in Neural Information Processing Systems*, Vol. 20.

Bucher, C. and Bourgund, U. (1990). “A fast and efficient response surface approach for structural reliability problems.” *Structural Safety*, 7(1), 57–66.

Duane, S., Kennedy, A., Pendleton, B. J., and Roweth, D. (1987). “Hybrid monte carlo.” *Physics Letters B*, 195(2), 216–222.

Engelund, S. and Rackwitz, R. (1993). “A benchmark study on importance sampling techniques in structural reliability.” *Structural Safety*, 12(4), 255–276.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). “Generative adversarial networks.

Itoi, T., Stafford, P. J., Kaneko, M., and Sato, I. (2014). “Stochastic model of ground motion time history for crustal earthquake.” *14th The Japan Earthquake Engineering Symposium*, 2031–2040.

Jaakkola, T. and Haussler, D. (1998). “Exploiting generative models in discriminative classifiers.” *Advances in Neural Information Processing Systems*, Vol. 11.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T. (2020). “Analyzing and improving the image quality of stylegan.” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June).

National Research Institute for Earth Science and Disaster Resilience (2023). “NIED K-NET. National Research Institute for Earth Science and Disaster Resilience, doi:10.17598/NIED.0004.

Nishida, A., Igarashi, S., Sakamoto, S., Uchiyama, Y., Yamamoto, Y., Muramatsu, K., and Takada, T. (2015). “Hazard-consistent ground motions generated with a stochastic fault-rupture model.” *Nuclear Engineering and Design*, 295, 875–886.

Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. The MIT Press.

Shawe-Taylor, J. and Cristianini, N. (2004). *Kernel methods for pattern analysis*. Cambridge university press.