

Article

Sequence Deep Learning for Seismic Ground Response Modeling: 1D-CNN, LSTM, and Transformer Approach

Yongjin Choi ¹, Huyen-Tram Nguyen ², Taek Hee Han ³, Youngjin Choi ³ and Jaehun Ahn ^{2,*}

¹ School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA; ychoi402@gatech.edu

² Department of Civil and Environmental Engineering, Pusan National University, Busan 46241, Republic of Korea; nhtram153@pusan.ac.kr

³ Ocean Space Development & Energy Research Department, Korea Institute of Ocean Science and Technology, Busan 49111, Republic of Korea; taekheehan@kiost.ac.kr (T.H.H.); yjchoi@kiost.ac.kr (Y.C.)

* Correspondence: jahn@pusan.ac.kr

Abstract: Accurate seismic ground response analysis is crucial for the design and safety of civil infrastructure and establishing effective mitigation measures against seismic risks and hazards. This is a complex process due to the nonlinear soil properties and complicated underground geometries. As a simplified approach, the one-dimensional wave propagation model, which assumes that seismic waves travel vertically through a horizontally layered medium, is widely adopted for its reasonable performance in many practical applications. This study explores the potential of sequence deep learning models, specifically 1D convolutional neural networks (1D-CNNs), long short-term memory (LSTM) networks, and transformers, as an alternative for seismic ground response modeling. Utilizing ground motion data from the Kiban Kyoshin Network (KiK-net), we train these models to predict ground surface acceleration response spectra based on bedrock motions. The performance of the data-driven models is compared with the conventional equivalent-linear analysis model, SHAKE2000. The results demonstrate that the deep learning models outperform the physics-based model across various sites, with the transformer model exhibiting the smallest average prediction error due to its ability to capture long-range dependencies. The 1D-CNN model also shows a promising performance, albeit with occasional higher errors than the other models. All the data-driven models exhibit efficient computation times of less than 0.4 s for estimation. These findings highlight the potential of sequence deep learning approaches for seismic ground response modeling.

Keywords: earthquake; seismic ground response modeling; convolutional neural networks (CNNs); long short-term memory (LSTM) networks; transformer



Citation: Choi, Y.; Nguyen, H.-T.; Han, T.H.; Choi, Y.; Ahn, J. Sequence Deep Learning for Seismic Ground Response Modeling: 1D-CNN, LSTM, and Transformer Approach. *Appl. Sci.* **2024**, *14*, 6658. <https://doi.org/10.3390/app14156658>

Academic Editors: Nicholas Vassiliou Sarlis, Jiwei Zhang, Hui Cao and Song Zhang

Received: 25 June 2024

Revised: 22 July 2024

Accepted: 27 July 2024

Published: 30 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Accurate estimation of seismic ground response is crucial for the design and safety of civil infrastructure, as it is the foundation to determine the impact of earthquake-generated ground motion on structures. Accurately modeling this response is vital for establishing effective mitigation measures against seismic risks and hazards for structures built on the ground [1,2].

The mechanics of seismic wave propagation during an earthquake involve intricate physical processes. When a fault occurs, it releases energy in the form of seismic waves. These waves propagate from the source through the soil media to the site. During this propagation, some wave components are filtered, while others are amplified depending on the site conditions, a phenomenon known as the site effect [3]. Seismic ground response analysis evaluates ground surface motions based on bedrock motion, accounting for the site effect. This process is inherently complex and includes uncertainties due to the heterogeneous nature of soil layers, varying soil properties, and underground geometries [4].

As a simplified approach, the one-dimensional wave propagation model [5–8] is widely adopted for site response analysis. This model assumes that seismic waves travel vertically through a horizontally layered medium and has proven reasonable for many practical applications as a physics-based model. Hashash et al. [9] apply a one-dimensional wave propagation model to analyze the seismic response in the Mississippi embayment. Zheng et al. [10] use this model for site response analysis in the New Madrid seismic zone. Zalachoris et al. [8] evaluate this model using borehole arrays. It is also widely used as the benchmark model to calibrate the nonlinear soil models for response analysis [9,11,12]. To capture the full complexity of wave propagation, full 3D numerical models can be employed, but describing the 3D underground geometry is challenging work, and entails significant computational costs [13–15].

Recent studies have demonstrated the potential of data-driven models using machine learning (ML) as successful surrogates for physics-based models. These models learn complex physical phenomena from data, achieving outstanding performance in various applications, such as global weather forecasting [16,17], the dynamics of granular flow and rigid body interaction, and fluid flow simulation [18–20]. By leveraging machine learning architectures tailored to focus on learning specific physical phenomena, these models often surpass conventional physics-based models in terms of accuracy or computational efficiency.

In terms of the application of data-driven approaches for ground motion prediction, Fayaz et al. [21] introduce an autoencoder-based system for the real-time estimation of ground motion acceleration. The framework efficiently processes initial p-wave data to predict spectral acceleration with high accuracy, implying its validity for on-site earthquake early warning systems. Akhani et al. [22] propose a hybrid computational intelligence approach that integrates genetic algorithms, neural networks, and regression analysis to predict spectral acceleration. Their model outperforms the accuracy of traditional ground motion models such as Campbell et al. and Abrahamson et al. [23,24]. Hu et al. [25] utilize PCA and genetic algorithms to simulate ground motion, effectively matching the amplitude, spectrum, and duration characteristics for realistic seismic input generation.

For seismic response analysis, where data consist of motion sequences or time series, sequence deep learning architectures can be promising options for learning the underlying sequential behavior of the motions. Popular architectures for sequence deep learning include 1D convolutional neural networks (1D-CNNs) [26], which capture local patterns in the sequence, and long short-term memory (LSTM) [27] networks, which can model long-term dependencies. For more advanced models, the transformer network shows a state-of-the-art performance [28] for sequential data. These models are designed to effectively capture the sequential or temporal relationships between the sequence of input features and the resulting responses.

The application of sequence deep learning models has been successful throughout various civil engineering fields. Choi et al. [29] used LSTM for predicting the pore pressure response in liquefiable sands under cyclic loading. Zhang et al. [30] and Zhang et al. [31] used LSTM to learn the stress–strain behavior of soils. In seismic response modeling, Hong et al. [32] proposed a CNN-based model to estimate the seismic ground response based on the earthquake records. The model performance was compared with the conventional one-dimensional physics-based model. Similarly, Li et al. [33] used a CNN and LSTM for seismic response modeling, and they compared the performance with the FEM model. Liao et al. [34] improved the LSTM model by using attention mechanisms [28] and applied their model to the seismic response prediction of bridges. The improved model showed a promising performance improvement compared to conventional LSTM. Zhang et al. [35] used the transformer, a state-of-the-art sequence model, for the structural seismic response for both linear and nonlinear systems and showed improved accuracy compared to LSTM. The above studies highlight that the data-driven approach can successfully model the seismic response for unseen conditions and potentially surpass the physics-based models under their research conditions.

Although previous studies proposed CNN- and LSTM-based models for ground seismic response modeling, the transformer-based model, which is the most advanced sequence learning architecture, has not been studied for this purpose. Instead, research involving transformers has primarily focused on structural response modeling. Additionally, the comprehensive comparison between the performance of the sequence learning models for seismic response is still limited.

To address these gaps, we compare the CNN, LSTM, and transformer approaches for seismic ground response modeling. Based on these sequence learning architectures, we aim to (1) develop site-specific machine learning models for seismic ground response modeling and (2) explore the validity of sequence deep learning architectures (CNN, LSTM, transformer) for learning the site response. We prepare the ground motion measurement data from the Kiban Kyoshin Network (KiK-net). The models are trained on ground motion measurement data for individual sites to consider the site-specific conditions. We then compare the performance of these data-driven models with the conventional physics-based model, SHAKE2000 [4], as a baseline to evaluate their performance and validity for seismic ground response prediction. The results provide new insights into the potential of advanced sequence deep learning models for application in seismic ground response.

2. Earthquake Data

We utilized the ground motion data from Hong et al. [32], which employ the strong-motion seismograph network Kiban Kyoshin Network (KiK-net) [36] in Japan. This network, established by the National Research Institute for Earth Science and Disaster Prevention (NIED), comprises seismometers installed in boreholes and on the surface at various sites (marked with black hollow circles in Figure 1) across the country to record earthquake motions. Each KiK-net site provides information about site-specific data, acceleration history measurements, and ground properties essential for seismic ground response analysis.

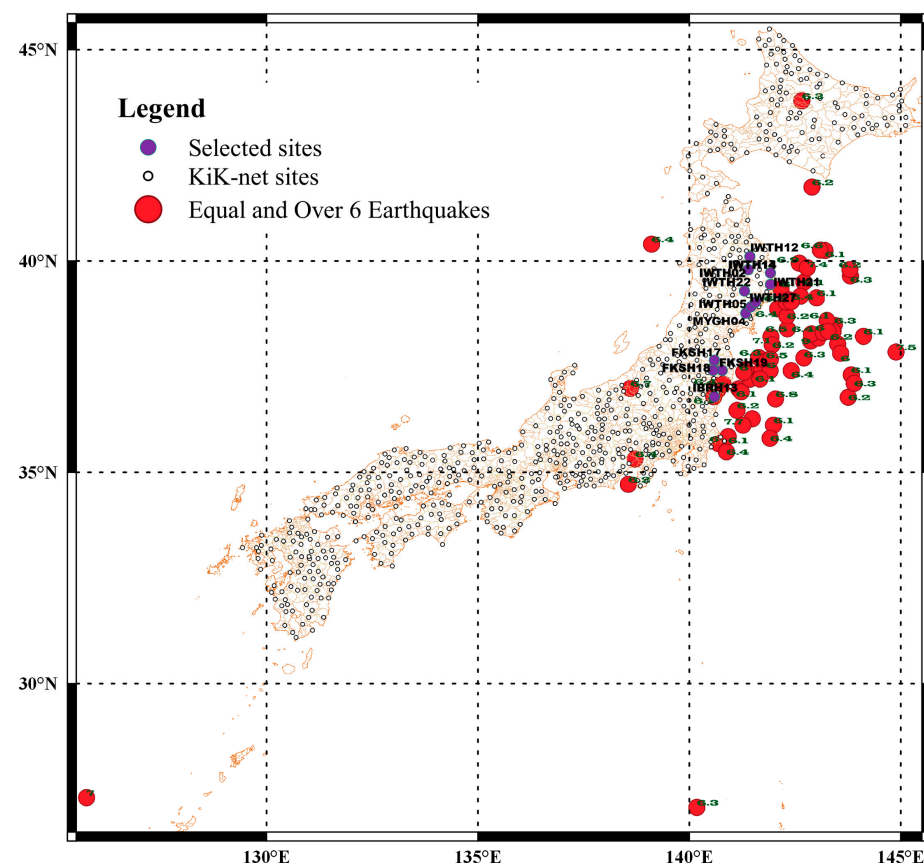


Figure 1. Selected earthquake sites for the analysis (reproduced from Hong et al. [32]).

Hong et al. [32] focus on ground motion data from 12 selected sites (marked with purple-colored circles in Figure 1). Each site has 100 earthquake events recorded after 2008, with moment magnitudes greater than 6 and peak ground accelerations (PGAs) exceeding 0.01 g. These thresholds are to ensure more complete waveforms and a broader range of recorded ground motions as well as to focus on impactful seismic activities.

The shear wave velocity and density profiles for these sites are presented in Figure 2. Table 1 summarizes the average shear wave velocity ($V_{s,30}$), average natural period (T_g), and the corresponding National Earthquake Hazard Reduction Program (NEHRP) site classification based on $V_{s,30}$ and T_g for each site. For a more comprehensive description including the ground layer profiles obtained from borehole data, please refer to Hong et al. [32].

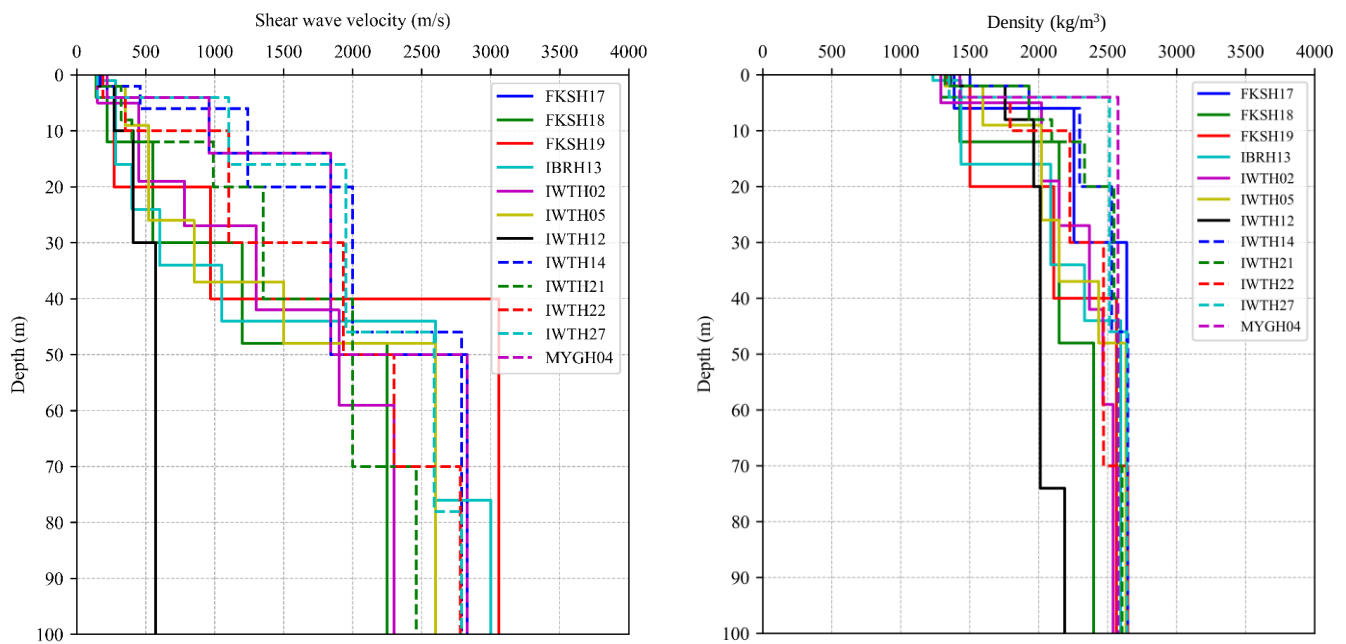


Figure 2. The shear wave velocity (reproduced from Hong et al. [32]) and the density profile for the selected sites.

Table 1. The information of the selected sites (reproduced from Hong et al. [32]).

Site ID	Site Name	$V_{s,30}$ (m/s)	T_g (s)	NEHRP Site Classification	Description
FKSH17	Kawamata	544	0.22	C	Dense soil, soft rock
FKSH18	Miharu	307.2	0.39	D	Stiff soil
FKSH19	Miyakoji	338.1	0.35	D	Stiff soil
IBRH13	Takahagi	335.4	0.36	D	Stiff soil
IWTH02	Tamayama	816.3	0.15	B	Rock
IWTH05	Fujisawa	442.1	0.27	C	Dense soil, soft rock
IWTH12	Kunohe	367.9	0.33	C	Dense soil, soft rock
IWTH14	Taro	816.3	0.15	B	Rock
IWTH21	Yamada	521.1	0.23	C	Dense soil, soft rock
IWTH22	Towa	532.1	0.23	C	Dense soil, soft rock
IWTH27	Rikuzentakata	670.3	0.18	C	Dense soil, soft rock
MYGH04	Towa	849.8	0.14	B	Rock

The earthquake datasets include acceleration time history measurements at the bedrock (100 m below the ground surface) and the ground surface. Using these time history data, we generate acceleration, velocity, and displacement response spectra with a five percent damping ratio. The deep learning models in this study take the response spectra (acceleration, velocity, and displacement) at the bedrock as the input and predict the surface

acceleration response spectrum. The following section provides a detailed explanation of the models.

3. Prediction Models

We propose three kinds of data-driven seismic ground response prediction models that are based on the following deep learning architectures: (1) 1D convolutional neural networks (1D-CNNs), (2) long short-term memory (LSTM), and (3) transformer. These deep learning architectures are specialized to learn the sequential dependency of the input sequence to the output with their own unique architecture. We design the models to estimate the spectral seismic response of the ground surface based on the input seismic motion measured at 100 m below the surface. The performance of the deep learning models is later compared with the physics-based model, SHAKE2000, in Section 5, as it is usually used as a benchmark for response analysis models and their calibration [9,11,12]. In this section, we explain the overview of the deep learning architecture for each model for learning the sequential relationship, in addition to the physics-based model.

3.1. Physics-Based Model

SHAKE2000 [6,37], a widely adopted software for site-specific seismic response analysis [5,38–40], is used as the physics-based model for predicting the ground response in this study. SHAKE2000 employs a 1D wave propagation model, assuming that seismic waves travel vertically through horizontally layered soil deposits. The model inputs include a soil profile, which encompasses the thickness of layers and soil properties such as density and dynamic properties, and a bedrock motion, which can be acceleration, velocity, or displacement time histories, serving as the seismic excitation at the base of the 1D soil column.

SHAKE2000 utilizes an equivalent-linear analysis model [5,41] to account for nonlinear soil behavior under seismic loading. The bedrock time history motion ($a_{in}(t)$ in Equation (1)) is first transformed from the time domain to the frequency domain using a Fast Fourier Transform (FFT in Equation (1)). The model then calculates the wave propagation through the soil layers using transfer functions ($f(\omega)$ in Equations (1) and (2)) based on the layers' dynamic properties. In Equation (2), $u_{out}(\omega)$ is the ground surface displacement and $u_{in}(\omega)$ is the bedrock displacement. The equivalent-linear procedure iteratively adjusts the shear modulus and damping ratio of each layer according to the induced shear strains based on shear modulus reduction and damping curves. This process repeats until the iteration of the shear modulus and damping ratio converges. Upon convergence, the model calculates the final response in the frequency domain and converts it back to the time domain using an inverse FFT (IFFT in Equation (1)), returning the ground surface time history motion ($a_{out}(t)$ in Equation (1)).

$$a_{out}(t) = IFFT[f(\omega) \times FFT[a_{in}(t)]] \quad (1)$$

$$f(\omega) = \frac{u_{out}(\omega)}{u_{in}(\omega)} \quad (2)$$

In this study, in order to determine the dynamic soil properties (i.e., shear modulus reduction and damping curves) for the site, we classify the site layer types as rock, gravel, and sand based on the layer material information provided by the KiK-Net borehole data. Detailed layer information for KiK-Net sites can be found in Hong et al. [32]. Then, the mean values of the dynamic properties from the literature [42] are taken based on the classified layer types. Figure 3 shows the shear modulus reduction and damping curves used in this study. Another important input data for SHAKE2000 are the shear wave velocity and density profile. We use the data provided by KiK-Net as described in Figure 2. Based on the selected properties, the equivalent-linear analysis is conducted to estimate the ground surface response spectra.

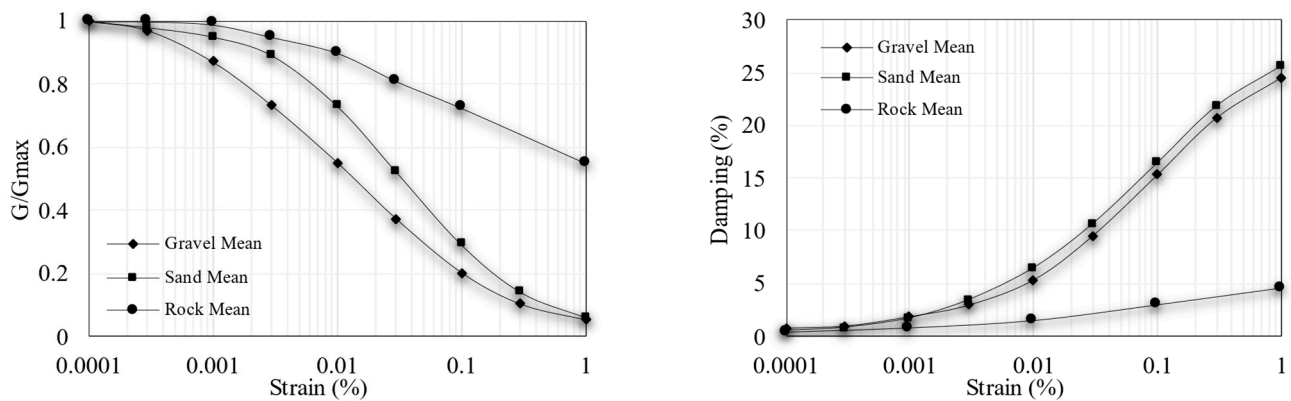


Figure 3. Normalized shear modulus reduction curves and damping ratio curves (reproduced from Hong et al. [32]).

3.2. One-Dimensional Convolutional Neural Network (CNN)-Based Model

3.2.1. Overview

Convolutional neural networks (CNNs) have been widely used in image processing. CNNs learn the local spatial pattern in input data through convolution kernels, which sweep across the 2D data to extract meaningful adjacent features [43]. While commonly associated with 2D data, CNNs can also be applied to 1D sequential data [26]. One-dimensional CNNs (1D-CNNs) are designed for tasks involving time series and signal processing. In 1D CNNs, the convolutional kernel operates along a single dimension to capture local dependencies and patterns within input sequences.

The convolutional kernel (see Figure 4) slides over the input sequence, performing element-wise multiplications to capture relationships between adjacent elements. As the network deepens, the convolutional layers learn more complex and abstract representations of the input sequences. The hierarchical structure of 1D CNNs allows them to capture sequential dependencies, and understand underlying patterns and relationships. Pooling layers downsample the feature maps, reducing dimensions while preserving key features, thus focusing on the most informative aspects of the input sequences. More technical details can be found in Kiranyaz et al. and Alzubaidi et al. [26,43].

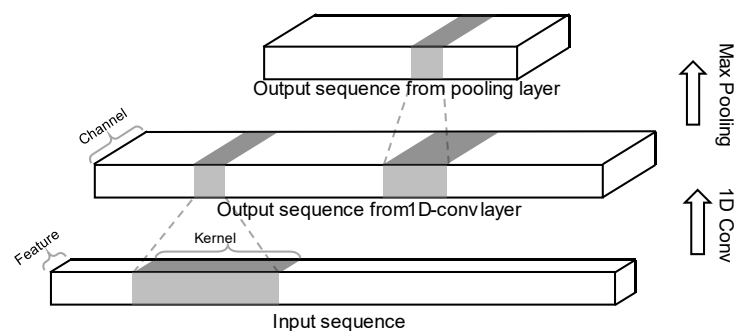


Figure 4. Convolutional neural networks (CNNs) with pooling layer.

3.2.2. Model Architecture

The ability of 1D CNNs to learn sequential dependencies makes them suitable for seismic response modeling. One-dimensional CNNs can learn the relationships between the input earthquake motion and the corresponding ground response. We design the model to take the bedrock spectral motion as a sequential input, and let it extract relevant features and patterns contributing to the seismic ground surface spectral response. Through stacks of CNN and pooling layers, the model learns the nonlinear and complex relationships between the input earthquake motion and the resulting ground surface response hierarchically, enabling the accurate prediction of the seismic response.

Figure 5 shows our CNN-based model for seismic ground response estimation. The model maps the input sequence, which is the acceleration, velocity, and displacement spectra, to the output sequence, which is the ground surface acceleration spectrum. The model consists of three consecutive sets of 1D-CNN and max-pooling layers. Each convolution layer is configured with the kernel size of 24, 12, and 6, and output channel size of 16, 32, and 16, respectively. The pooling size is all set to 2, which downsamples the sequence by a factor of two. After the last CNN layer, a dense layer is placed to map the output of the CNN layers back to the original sequence length. This structure follows the typical sequence learning of 1D-CNNs, which is also used by Li et al. [33].

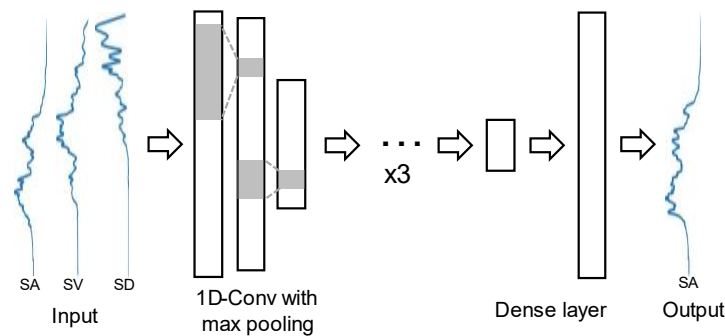


Figure 5. One-dimensional CNN-based model for seismic ground response modeling.

3.3. Long Short-Term Memory (LSTM) Networks-Based Model

3.3.1. Overview

Long short-term memory (LSTM) networks [27] are a type of recurrent neural network (RNN) specifically designed to effectively handle sequential data and capture long-term dependencies. Unlike traditional RNNs, which struggle with vanishing or exploding gradients when dealing with long sequences [44], LSTMs introduce a memory cell and gating mechanisms to selectively remember and forget information over the sequence. This allows LSTMs to maintain long-term memory and learn complex patterns in sequential data. LSTMs have proven to be successful in various sequence learning tasks in geotechnical and structural areas [30,31,34,45] where capturing long-range dependencies is needed.

LSTMs learn sequential dependencies by utilizing three gating mechanisms in the memory cell: input gate i_t^Θ , forget gate f_t^Θ , and output gate o_t^Θ (Figure 6), which include learnable parameters. These gates selectively regulate the previous cell information (C_{t-1} and h_{t-1}), new information (x_t), and output of the current cell C_t and h_t . Here, C denotes the cell state, and h denotes the hidden state. The input gate i_t^Θ controls the flow of input information (x_t and h_{t-1}) into the memory cell, deciding which information should be stored. The forget gate f_t^Θ determines what information should be discarded from the previous cell state C_{t-1} and the new information x_t , allowing the network to selectively forget irrelevant or outdated information. The output gate o_t^Θ regulates the flow of information in the memory cell (x_t and h_{t-1}) to its output, referring to the cell state C_t . For more details, refer to Choi et al. [29].

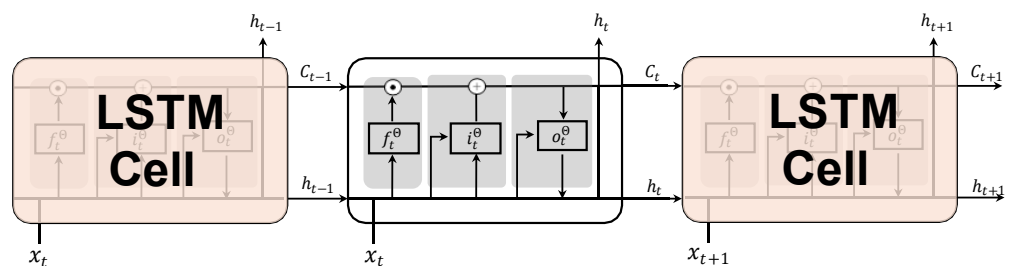


Figure 6. Long short-term memory (LSTM) cell structure.

Compared to 1D-CNNs, LSTMs excel at capturing long-term dependencies within the data. While 1D-CNNs are effective at learning local patterns and short-term dependencies through convolutional kernels, they struggle with long-range correlations due to their localized receptive fields. In contrast, LSTMs' memory cells and gating mechanisms allow them to retain and utilize information over extended sequences, making them more suitable for modeling temporal information change like seismic response data.

3.3.2. Model Architecture

LSTMs can be effectively applied to model the seismic response. We configure the model to take the bedrock spectral motion as a sequential input, and learn to predict the ground surface spectral response, by learning the relevant features and patterns throughout the natural periods contributing to the response. This learning process is facilitated by the selective learning ability of the LSTM's gated structure over the sequence.

Figure 7 shows our LSTM-based model for seismic ground response estimation. The model has three LSTM layers with a hidden size of 32. This structure follows the typical sequence learning architecture of the previous research [31,34,45]. Note that we set the LSTM to be bidirectional to improve the learning capacity since the sequential dependency of the spectral motion might not be one directional, but more likely to be related to back and forth. After the last LSTM layer, the output sequence is flattened and passes through the dense layer to return the output response spectrum.

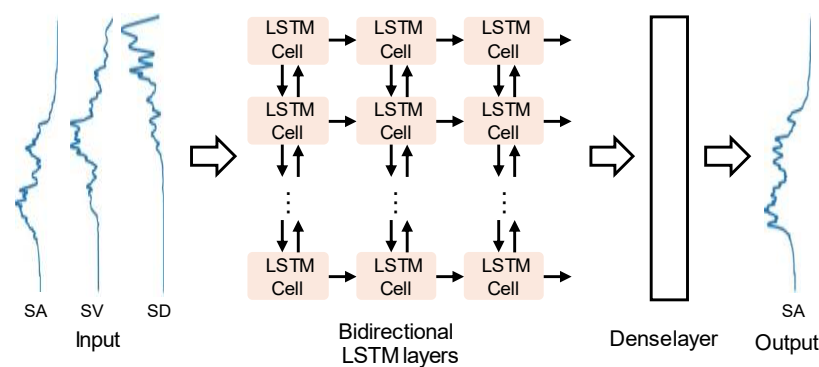


Figure 7. LSTM-based model for seismic ground response estimation.

3.4. Transformer-Based Model

3.4.1. Overview

Transformers [28] have significantly advanced the field of deep learning, particularly in sequential data processing based on a self-attention mechanism. Unlike RNN-based models, including LSTM, which may struggle to maintain information over long sequences due to the vanishing gradient problem [44], transformers employ a self-attention mechanism [28] that enables processing entire sequences of data in parallel. This capability allows them to consider the dependencies and relationships across the elements in sequence more efficiently, irrespective of their proximity.

The self-attention mechanism (Figure 8) operates by comparing each element in the sequence ($x_{t_0}, x_{t_1}, \dots, x_{t_k}$ in Figure 8) to every other element with a set of learnable parameters. This pair-wise operation returns the attention scores, which determine the relative importance between the elements. The attention scores are then used to weight the contributions of the other elements ($x_{t_1}, \dots, x_{t_k}$) to the target element (x_{t_0}). The summation of the weighted sequence forms a new context-aware representation of the target element, x_{t_0} . This process allows the model to capture long-range dependencies within the sequence by focusing on relevant information from other elements with adaptive weighting based on learned importance in parallel. The entire operation entails the above process for all elements ($x_{t_0}, x_{t_1}, \dots, x_{t_k}$).

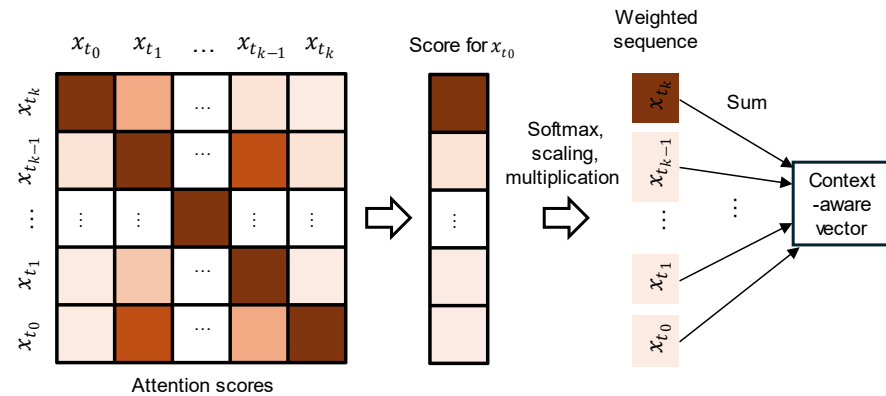


Figure 8. Self-attention mechanisms (revised from Chollet [46]).

By leveraging self-attention, transformers offer an efficient way to integrate contextual information, allowing the model to emphasize the most relevant parts of the data sequence for a deeper understanding of its sequential relationships.

The transformer architecture consists of an encoder–decoder structure (Figure 9). The encoder (left side of Figure 9) processes the input sequence through multiple layers, each applying a self-attention mechanism followed by dense layers. The output from the encoder provides the contextual representation of the input sequence. The decoder (right side of Figure 9), also composed of multiple layers, generates the output sequence by utilizing two types of attention: self-attention to process the output sequence generated so far and encoder–decoder attention to integrate information from the encoded input sequence. Each layer in the decoder also includes a feedforward neural network. The layer normalization [47] and residual connections [48] are applied to both the encoder and decoder to enhance and stabilize the learning process.

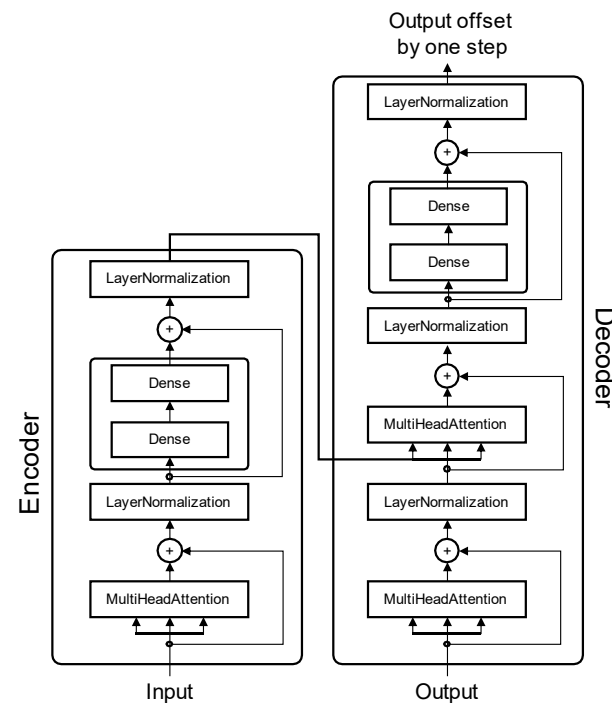


Figure 9. Transformer encoder–decoder architecture (revised from [28]).

Moreover, multi-head attention is introduced. It extends the self-attention mechanism by using multiple attention heads. Each head learns different parts of the input, and their outputs are concatenated to form the final representation. This allows the model to capture

various aspects of the data in parallel, leading to a richer and more nuanced representation of the sequence. For full technical details, refer to Vaswani et al. [28].

Gao et al. [49] utilize a transformer for structural health monitoring, which integrates classification, localization, and segmentation tasks. The transformer network enhances the accuracy of processing images for detecting and assessing structural damages owing to the self-attention mechanism that learns and exploits interdependencies among multiple attributes and tasks in image processing. Shan et al. [50] use a transformer to measure the vibration profiles of high-rise buildings from unmanned aerial vehicle (UAV)-recorded videos. The use of transformer networks helps mitigate the drifting issues induced by UAV movements, ensuring accurate full-field deformation measurements.

3.4.2. Model Architecture

Figure 10 shows our transformer-based model for seismic ground response estimation. The model first embeds the input features into a 16-dimensional embedding space using a dense layer. To incorporate the information about the natural period, we add positional encodings to the embedding space based on the natural period values, which helps the model recognize the input features associated with the corresponding natural periods. The processed embedding is then fed into a transformer module with the following configuration: 3 encoder layers and 3 decoder layers, each with 4 attention heads and a feedforward network dimension of 128. Finally, the output from the transformer is passed through a dense layer to generate the final predictions. This architecture follows the basic transformer architecture that is used to learn sequence data [46].

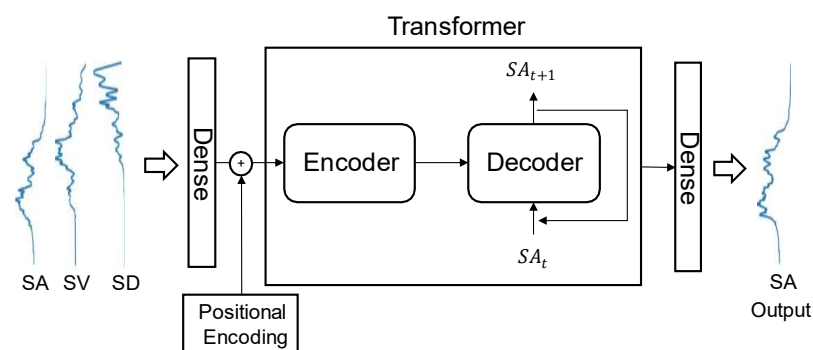


Figure 10. Transformer-based model for seismic ground response estimation.

4. Training

Each model contains the learnable parameters in their own components. The CNN model (Figure 5) contains the learnable parameters in the convolution kernel in the three convolution layers and the last dense layer. The LSTM model (Figure 7) contains the learnable parameters in the LSTM cells in the three LSTM layers and the last dense layer. The transformer model (Figure 10) contains the learnable parameters in the first dense embedding layer with the positional encoder, the transformer that includes encoder and decoder layers, and the last dense layer. We train these parameters to minimize the mean squared loss (MSE) between the measured response spectrum at the ground surface and the predicted spectrum (Equation (3)). The training is conducted for each of the 12 sites explained in Table 1 to consider the site-specific ground characteristics. Each site contains 100 earthquake motions recorded from 50 distinct seismic events in two different directions (east–west and north–south directions), resulting in a total data pool of 1200 earthquake motions. For each site, we randomly divided the 100 earthquake motions into three datasets: 80 for training, 10 for validation, and 10 for testing. During the training process, we randomly sampled from the validation data to assess the model’s performance.

$$loss_{\Theta} = \frac{1}{m} \left(S_a^{pred} - S_a^{true} \right)^T \left(S_a^{pred} - S_a^{true} \right) \quad (3)$$

where Θ is the learnable parameter set, S_a^{pred} is the predicted ground surface response spectrum, S_a^{true} is the measured ground surface response spectrum, and m is the number of data points.

5. Results and Discussion

5.1. Prediction Performance

5.1.1. Response Spectra

We first evaluate the prediction performance of the response spectra. Figures 11–22 show the prediction results for the sites explained in Table 1. Note that the weighted moving average is applied to the prediction for reducing the prediction noise as performed by Hong et al. [32]. The subfigures “a” to “f” in each figure show the six representative response spectra prediction results among 10 testing datasets for the corresponding site. The data-driven models—CNN-, LSTM-, and transformer-based models—generally show good prediction performances. The physics-based model, SHAKE, tends to overestimate the responses for most cases.

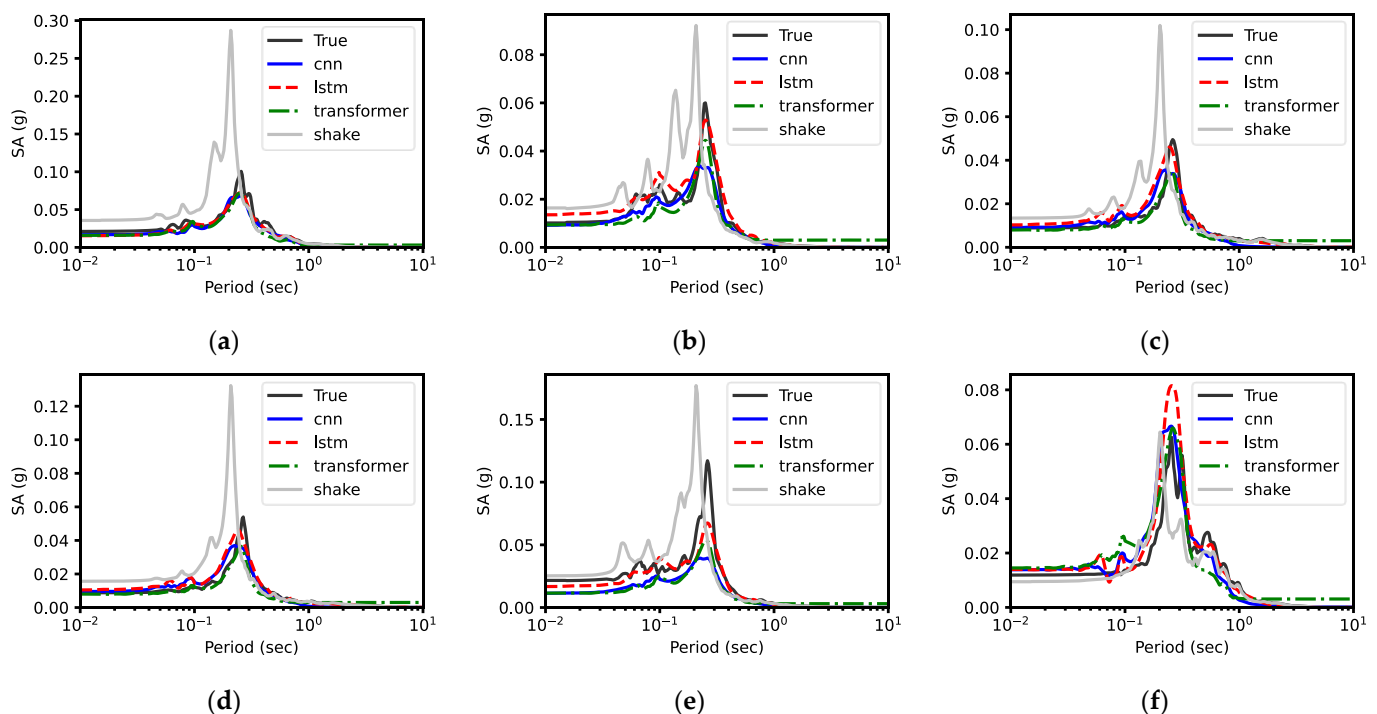


Figure 11. The six samples of acceleration response spectra prediction results compared to the measurements (true) for site FKSH17. Subfigures (a–f) correspond to the six samples for the current site.

The data-driven models predict the overall spectral trends well and capture the critical natural period where the response shows the peak amplitude and its amplitude. For example, in Figure 11a (site FKSH18), the predicted spectra align well with the true spectral trend. Further, the true peak amplitude and the predicted peak amplitudes show a good agreement at around the critical natural period of 0.96 s. These successful predictions are observed throughout the majority of the test datasets. In contrast, the physics-based model tends to overestimate the amplitudes, particularly for the sites FKSH17 (Figure 11), IWTH05 (Figure 16), IWTH21 (Figure 19), and IWTH22 (Figure 20).

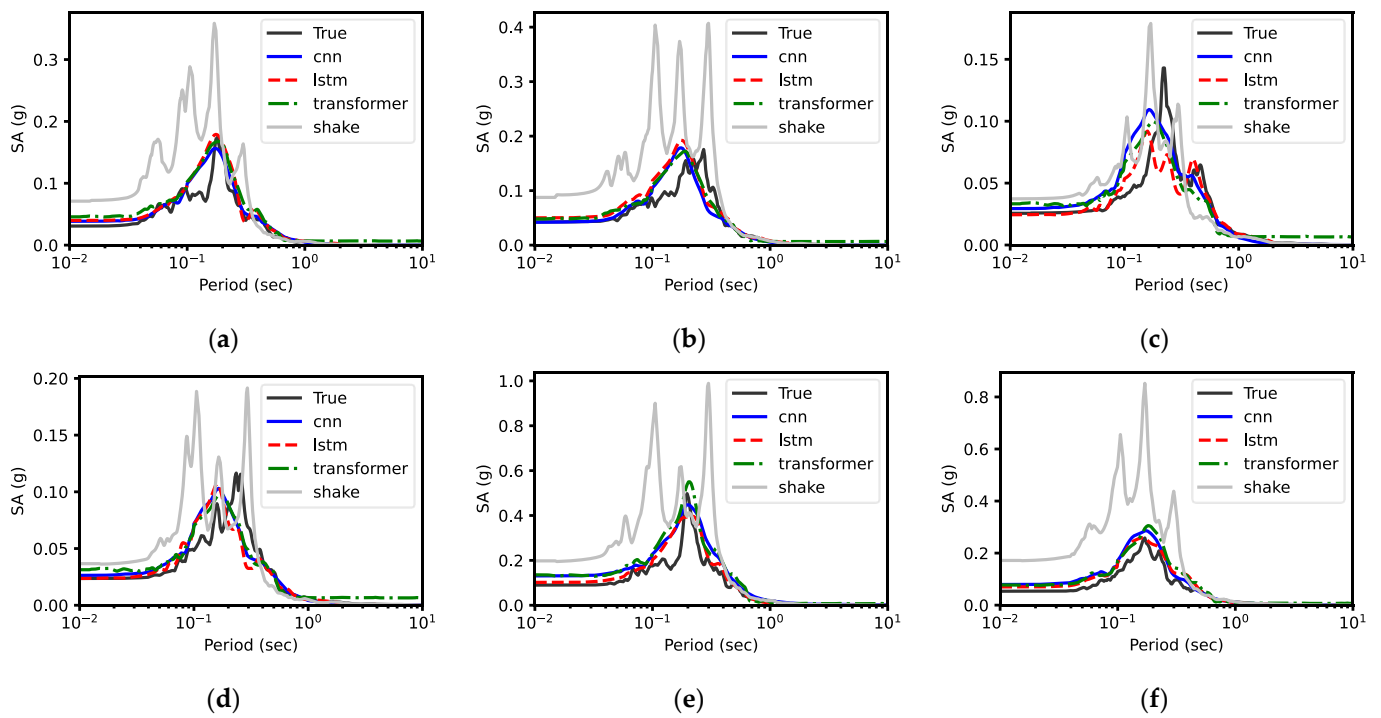


Figure 12. The six samples of acceleration response spectra prediction results compared to the measurements (true) for site FKSH18. Subfigures (a–f) correspond to the six samples for the current site.

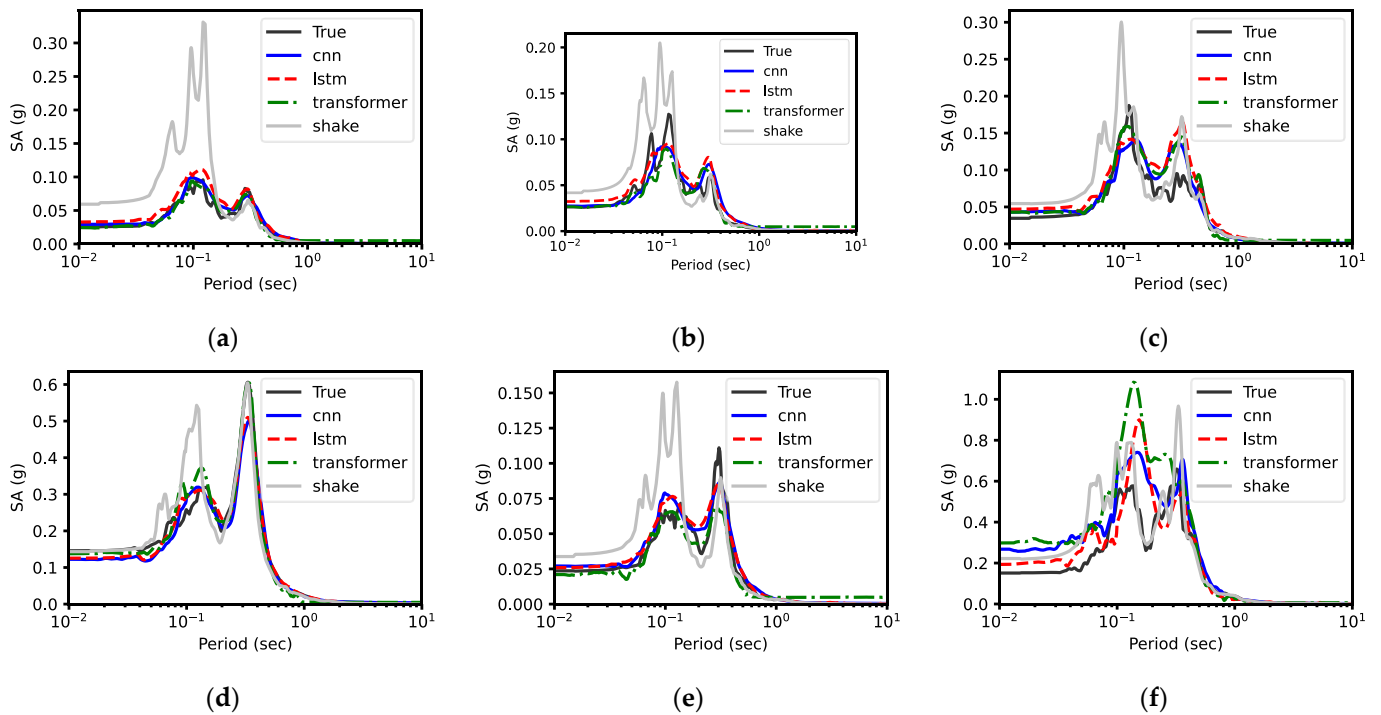


Figure 13. The six samples of acceleration response spectra prediction results compared to the measurements (true) for site FKSH19. Subfigures (a–f) correspond to the six samples for the current site.

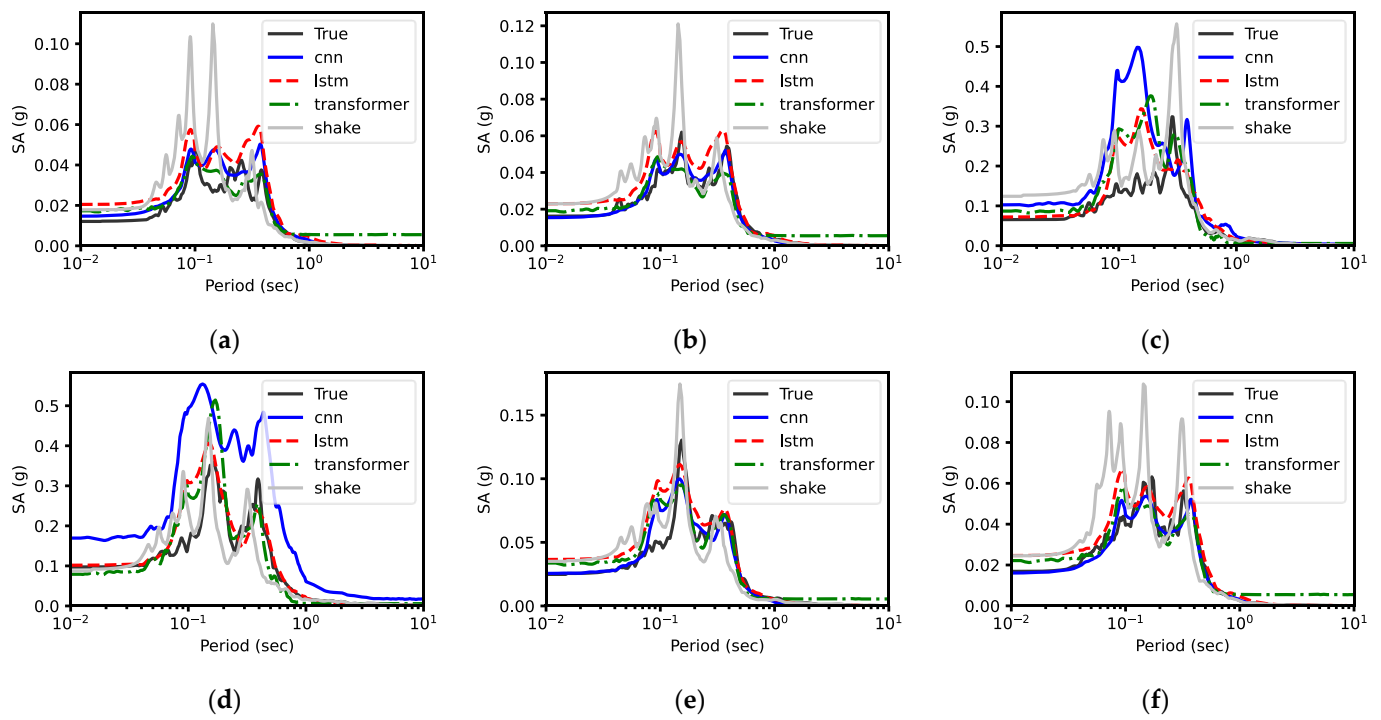


Figure 14. The six samples of acceleration response spectra prediction results compared to the measurements (true) for site IBRH13. Subfigures (a–f) correspond to the six samples for the current site.

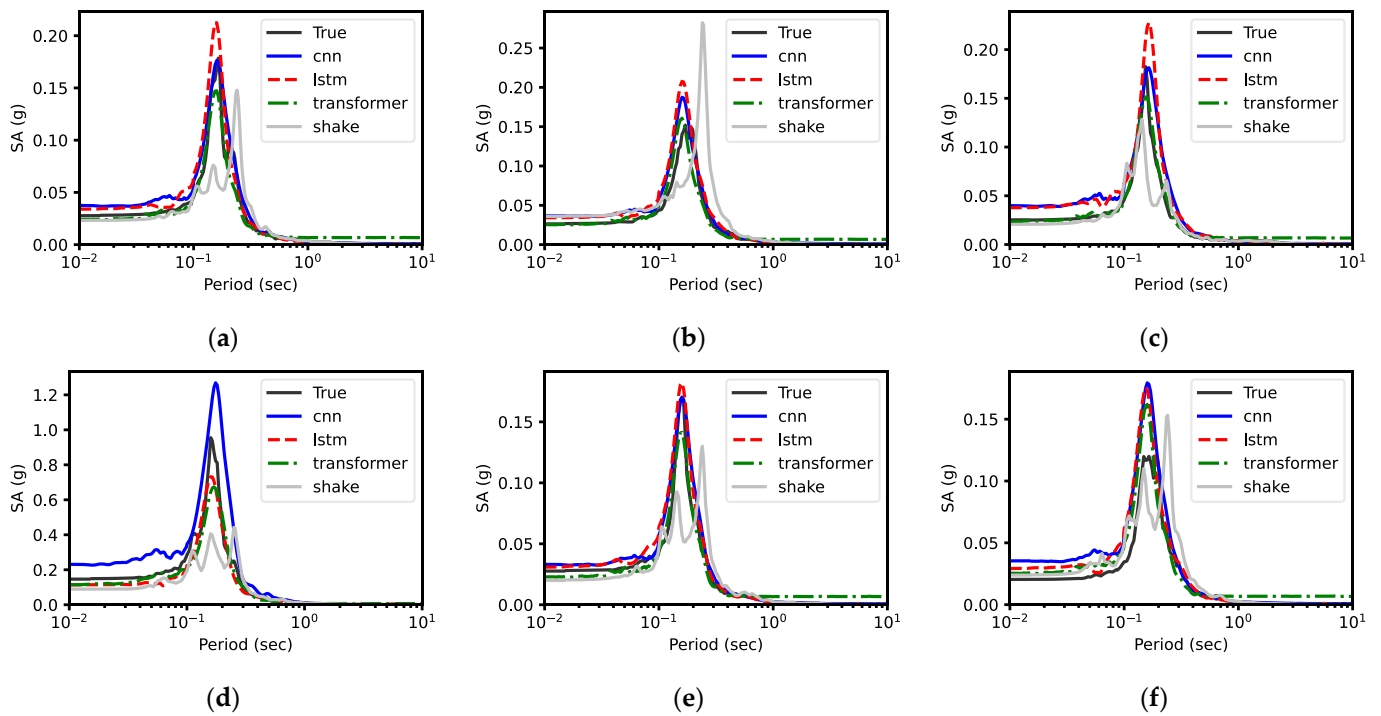


Figure 15. The six samples of acceleration response spectra prediction results compared to the measurements (true) for site IWTH02. Subfigures (a–f) correspond to the six samples for the current site.

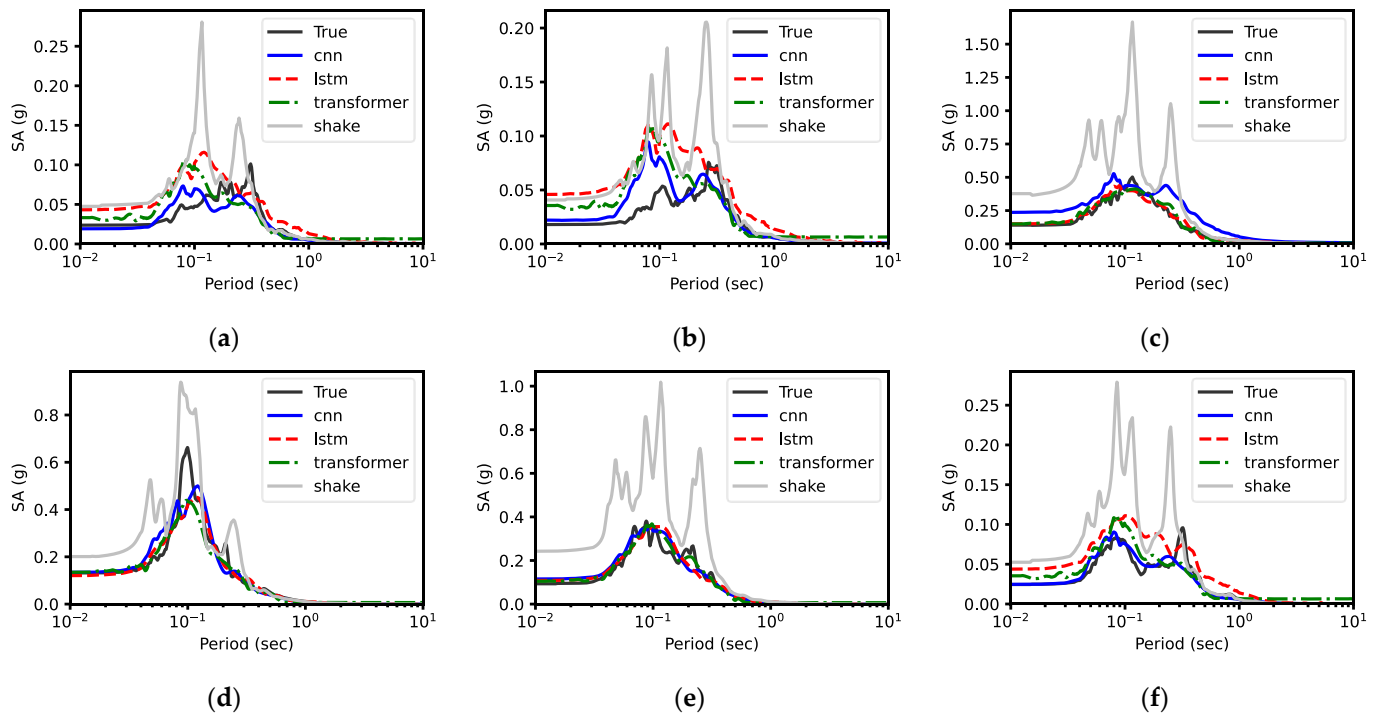


Figure 16. The six samples of acceleration response spectra prediction results compared to the measurements (true) for site IWTH05. Subfigures (a–f) correspond to the six samples for the current site.

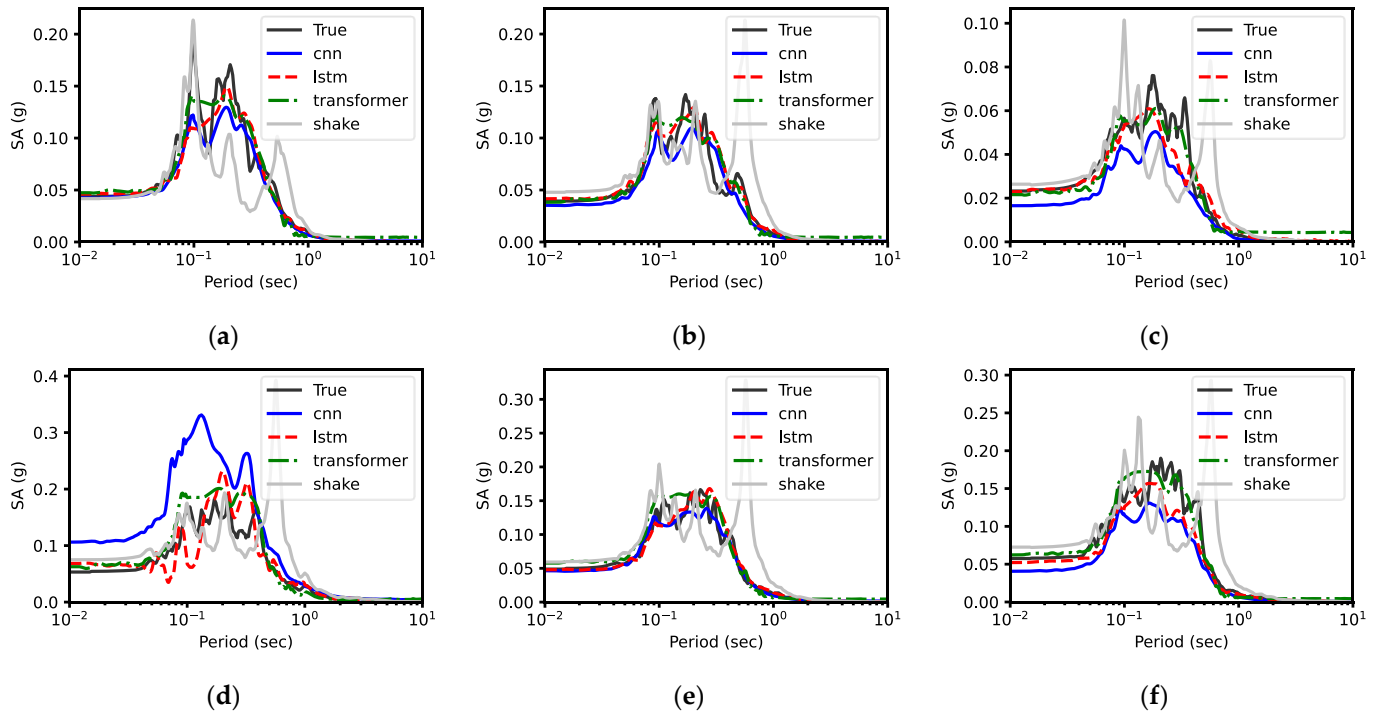


Figure 17. The six samples of acceleration response spectra prediction results compared to the measurements (true) for site IWTH12. Subfigures (a–f) correspond to the six samples for the current site.

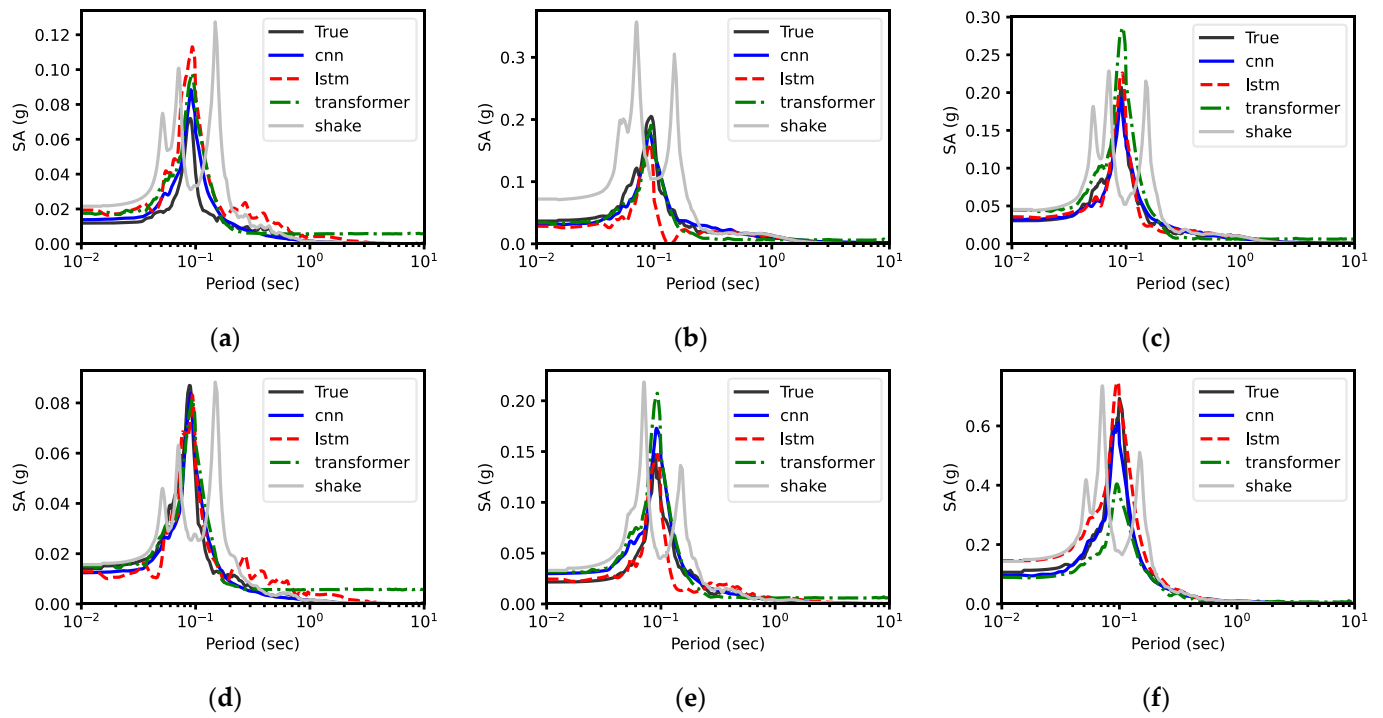


Figure 18. The six samples of acceleration response spectra prediction results compared to the measurements (true) for site IWTH14. Subfigures (a–f) correspond to the six samples for the current site.

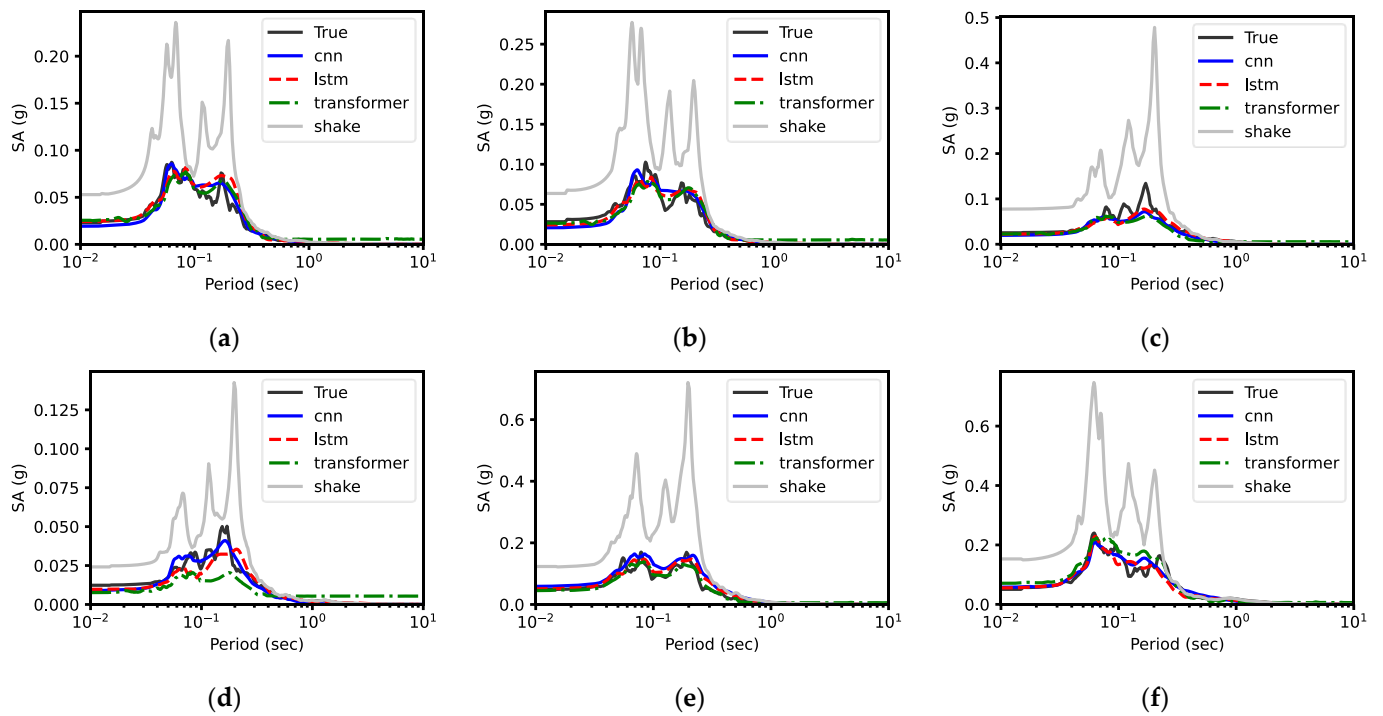


Figure 19. The six samples of acceleration response spectra prediction results compared to the measurements (true) for site IWTH21. Subfigures (a–f) correspond to the six samples for the current site.

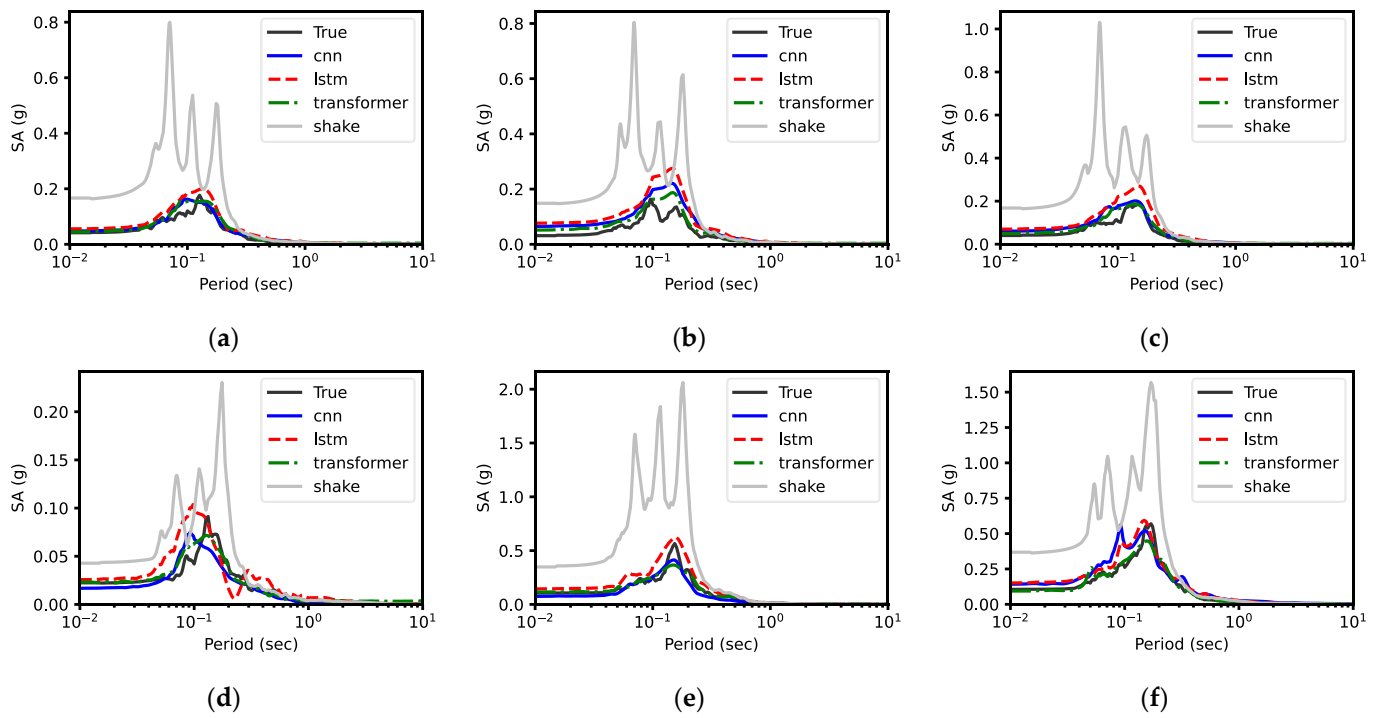


Figure 20. The six samples of acceleration response spectra prediction results compared to the measurements (true) for site IWTH22. Subfigures (a–f) correspond to the six samples for the current site.

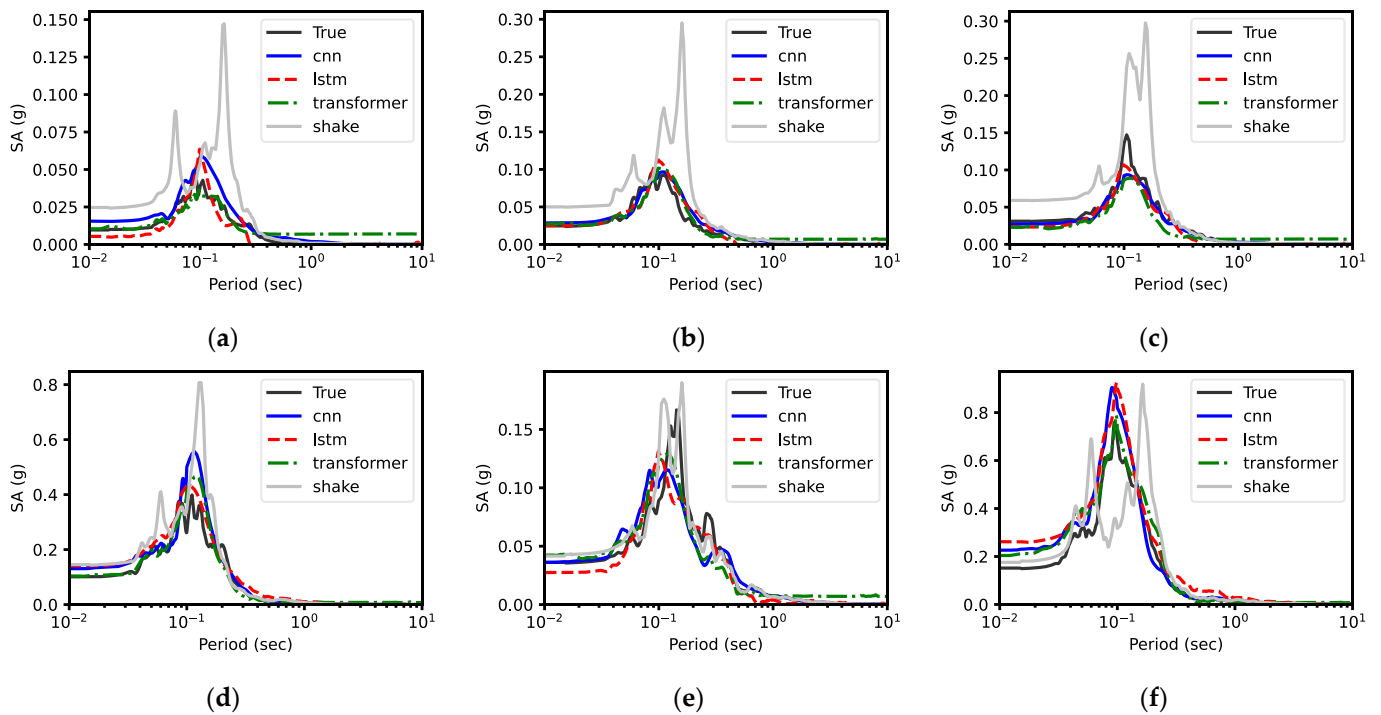


Figure 21. The six samples of acceleration response spectra prediction results compared to the measurements (true) for site IWTH27. Subfigures (a–f) correspond to the six samples for the current site.

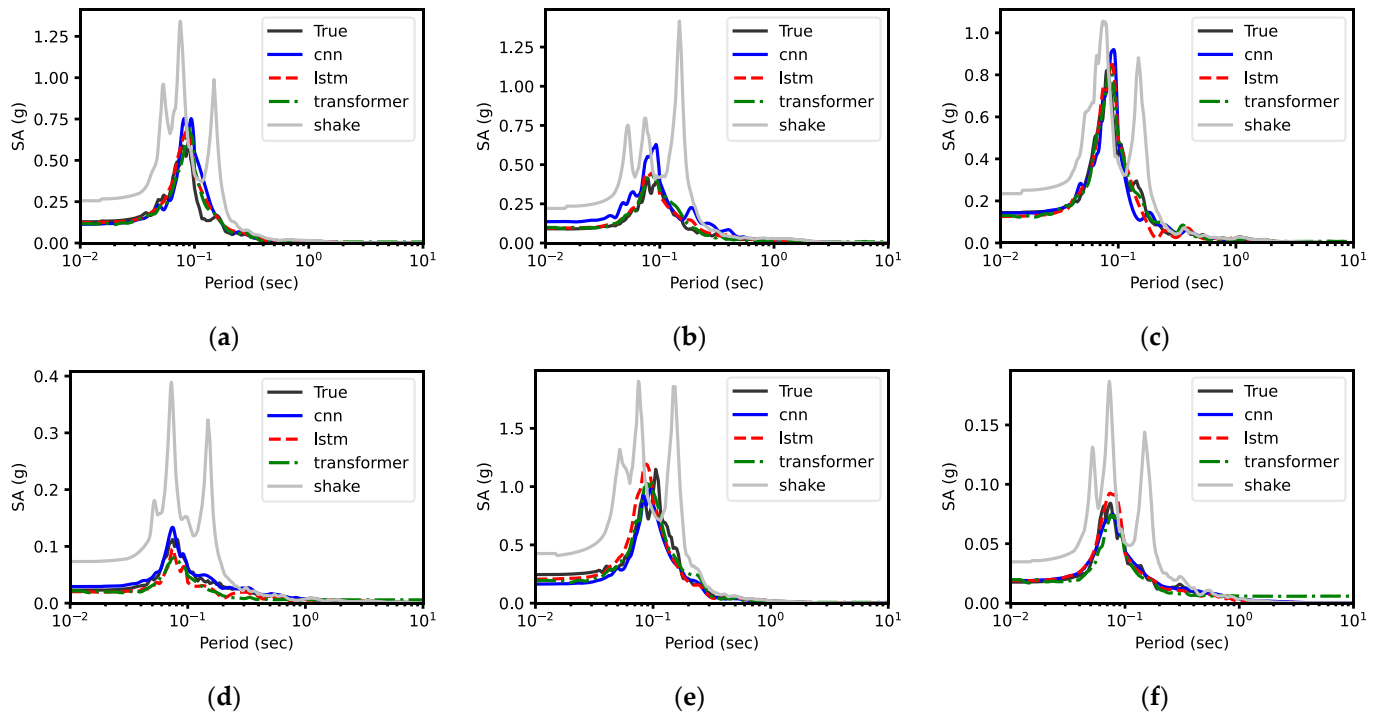


Figure 22. The six samples of acceleration response spectra prediction results compared to the measurements (true) for site MYGH04. Subfigures (a–f) correspond to the six samples for the current site.

In a few instances, the data-driven models deviate from the true spectrum, such as Figure 13f for the site FKSH19, Figure 14c for the site IBRH13, and Figure 16a,b for the site IWTH05. In these cases, the models fail to capture the overall trends of the response spectra and the peak amplitudes at the corresponding natural periods. For example, in Figure 14c, the models tend to show the peak amplitude at around 0.1–0.2 s, while the true peak amplitude occurs at 0.3 s. Additionally, the CNN model exhibits larger overestimations compared to the other models.

Among the models, the CNN occasionally makes a large overestimation and fails to capture the overall spectral trend (e.g., Figure 14c,d for site IBRH13 and Figure 17d for site IWTH12). These excessive overestimations and mispredictions are less frequently observed in the LSTM and transformer models. This can be attributed to the limited ability of the CNN to learn feature dependencies over long sequences, as it primarily focuses on adjacent features. In contrast, the LSTM and transformer models are inherently designed to consider the long sequences and their feature dependencies.

5.1.2. Prediction Error

We evaluate the root mean squared error ($RMSE_j$) with respect to periods over all 10 test ground motion datasets for the site j with Equation (4).

$$RMSE_j = \sqrt{\frac{\sum_{i=1}^l L^i}{l}} \quad (4)$$

$$L^i = \left(\mathbf{S}_a^{i,pred} - \mathbf{S}_a^{i,true} \right)^2 \quad (5)$$

Here, L^i is the squared error for the test ground motion dataset $i \in [1, 2, \dots, 10]$ between $\mathbf{S}_a^{i,pred}$ and $\mathbf{S}_a^{i,true}$. $\mathbf{S}_a^{i,pred}$ represents the predicted ground response spectrum and $\mathbf{S}_a^{i,true}$ represents the true ground response spectrum, for the test dataset i . l is the number of test ground motions for a site, which is 10.

Figure 23 shows the errors for all 12 sites. In general, the data-driven models show lower errors than the conventional physics-based model. Particularly, the physics-based model shows larger errors between 0.01 and 0.4 s for the sites FKSH17 (Figure 23a), FKSH18 (Figure 23b), IWTH05 (Figure 23f), IWTH21 (Figure 23i), IWTH22 (Figure 23j), and MYGH04 (Figure 23l). The error tends to be smaller for the sites IWTH12 (Figure 23g), IWTH14 (Figure 23h), and IWTH27 (Figure 23k), but still large in the natural period ranges from 0.5 to 1.0 s, 0.04 to 0.2 s, and 0.15 to 0.25 s, respectively.

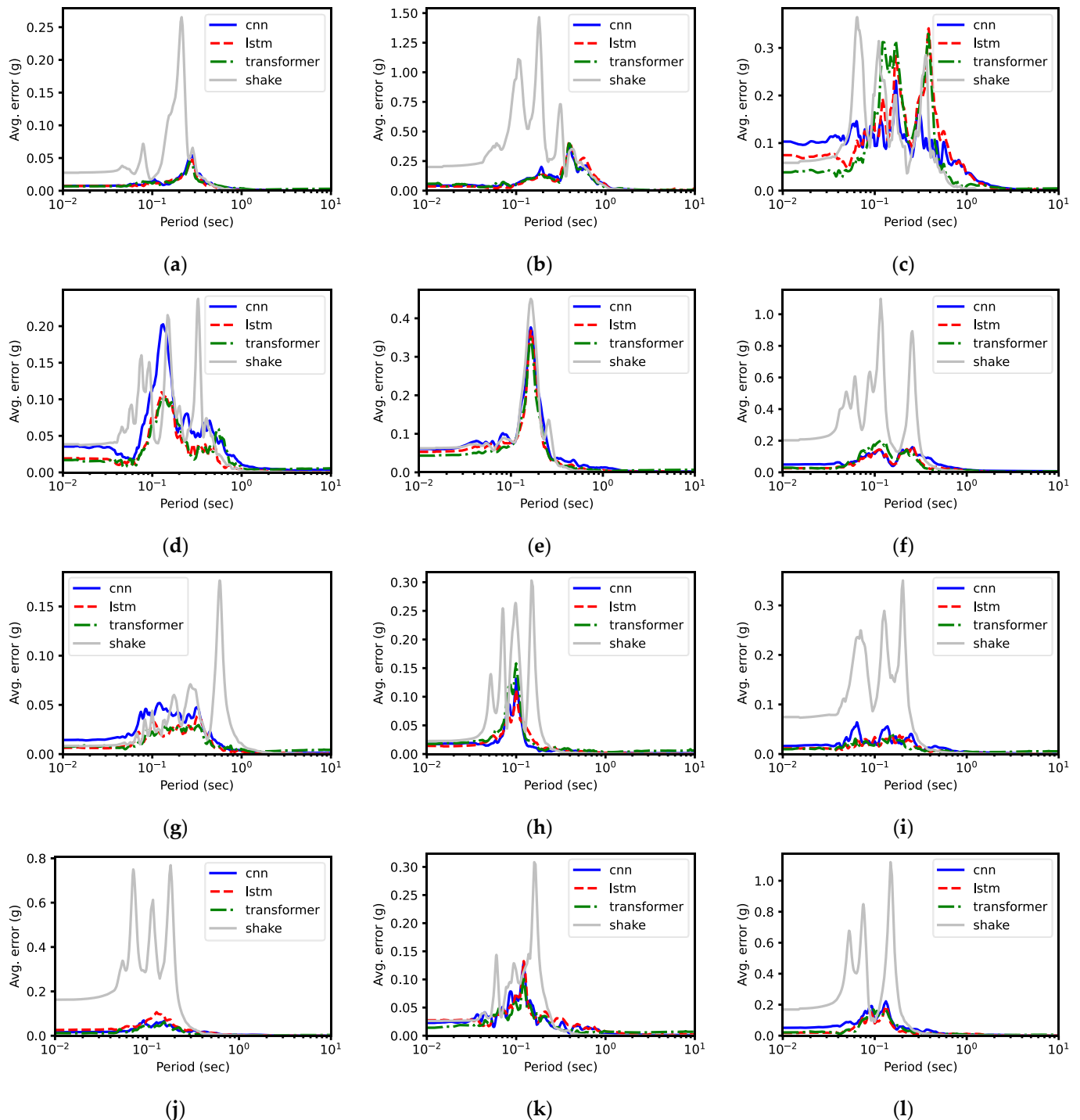


Figure 23. Acceleration response spectra prediction errors with respect to periods for all sites: (a) FKSH17, (b) FKSH18, (c) FKSH19, (d) IBRH13, (e) IWTH02, (f) IWTH05, (g) IWTH12, (h) IWTH14, (i) IWTH21, (j) IWTH22, (k) IWTH27, (l) MYGH04.

There are no distinct error trend differences between the data-driven models and physical model for the sites FKSH19 (Figure 23c), IBRH13 (Figure 23d), and IWTH02 (Figure 23e). The sites FKSH19 and IBRH13 show mixed error trends across the natural periods between the models, although the LSTM and transformer models slightly outperform the CNN and physics-based models. For the site IWTH02, all models exhibit similar prediction error trends over the natural periods.

Among the data-driven models, the CNN model shows larger errors between the natural periods of 0.01 and 1.0 s compared to the LSTM and transformer for the sites FKSH19 (Figure 23c), IBRH13 (Figure 23d), IWTH12 (Figure 23g), and MYGH04 (Figure 23l). For the sites IBRH13 (Figure 23d), IWTH21 (Figure 23g), and IWTH21 (Figure 23i), some error surges are observed between the period 1.0 and 1.0 s. The higher error of the CNN model can be attributed to its weaker ability to consider the sequential feature dependencies, as mentioned earlier. For the other sites, the models show similar error trends. For the site FKSH19 (Figure 23c), the error shows a mixed trend for all models.

We compare the model performances for all sites using the global error (see Equation (6)), as presented in Figure 24. It represents the average RMSE error over all the sites shown in Figure 23.

$$\text{Global error} = \frac{1}{m} \sum_{j=1}^m \text{RMSE}_j \quad (6)$$

Here, m is the number of sites ($=12$), j is the index of the sites, and RMSE_j is the root mean squared error computed in Equation (4).

The proposed deep learning models generally outperform the physics-based model, particularly for the natural periods ranging from 0.01 to 0.4. Among the deep learning models, the CNN model tends to show a higher error, particularly in the lower frequency range between 0.01 and 0.1 than the other models. The LSTM and transformer models exhibit similar error trends over the periods.

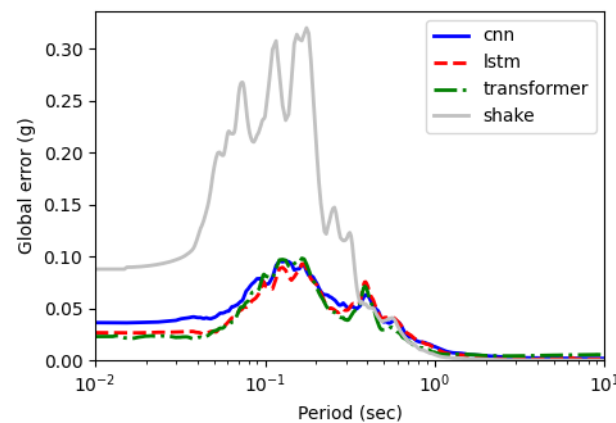


Figure 24. Global average prediction errors of acceleration response spectra for all sites.

To provide a quantitative error summary for each model, we use the average mean squared error (MSE_{avg}) to show the quantitative error summary for each model (Equation (7)).

$$MSE_{avg} = \frac{1}{l} \sum_{i=1}^l MSE^i \quad (7)$$

where, $MSE^i = \frac{1}{m} \sum_{k=1}^m L^i[k]$

Here, $L^i[k]$ is the k -th data point in the sequence of error L^i computed using Equation (5). m is the number of data points in the response spectrum, which is 500. l is the number of test ground motions ($=10$) as explained in Equation (4). The results are presented in Table 2.

Table 2. Average mean squared error (MSE_{avg}) for the models.

Site	CNN	LSTM	Transformer	SHAKE
FKSH17	0.000129	0.000092	0.000101	0.001697
FKSH18	0.000391	0.000177	0.000156	0.009639
FKSH19	0.009381	0.012412	0.009521	0.120766
IBRH13	0.007119	0.009947	0.008425	0.013708
IWTH02	0.002300	0.000726	0.000830	0.004351
IWTH05	0.004891	0.003666	0.003003	0.005632
IWTH12	0.003959	0.002878	0.004252	0.076996
IWTH14	0.000458	0.000163	0.000139	0.001461
IWTH21	0.000830	0.000689	0.001646	0.006386
IWTH22	0.000529	0.000830	0.000315	0.048460
IWTH27	0.000824	0.000750	0.000536	0.002810
MYGH04	0.004510	0.002340	0.002776	0.064900
Avg.	0.002943	0.002889	0.002642	0.029734

The data-driven models outperform the physics-based model for most of the sites by an order of magnitude smaller errors. However, this difference is less pronounced for the sites IBRH13 and IWTH05 than the other sites. For the sites IWTH02 and IWTH14, the CNN model exhibits noticeably larger errors than the LSTM and transformer models. Considering the overall average errors summarized in the last row of Table 2, the transformer model shows the smallest error compared to the other data-driven models. This can be attributed to its ability to capture longer sequential feature dependencies by leveraging attention mechanisms. We anticipate that the transformer model will exhibit increased performance gains compared to the CNN and LSTM models as the sequence length of the data grows.

5.2. Computational Performance

We compare the inference time of the CNN, LSTM, and transformer models, as shown in Table 3. Each model contains 442 K, 16,060 K, and 44 K parameters. The computations are performed on an NVIDIA Quadro RTX 5000 GPU. All models demonstrate fast prediction times, with inference completed in less than a second. The LSTM model exhibits a slightly longer computation time compared to the other models due to its sequential processing nature and the need to maintain and update the cell states at each time step. However, despite this minor difference, all the deep learning models show minimal computational costs.

Table 3. Computation time for the data-driven model.

CNN Model (s)	LSTM Model (s)	Transformer Model (s)
0.3830	0.4087	0.3894

6. Conclusions

In this study, we explore the potential of sequence deep learning models—specifically 1D-CNN, LSTM, and transformer architectures—for seismic ground response modeling. Using ground motion data from the KiK-net network, we train these models to predict the ground surface acceleration response spectrum from bedrock motions and compare their performance to the traditional physics-based model, SHAKE2000.

The data-driven models successfully predict the overall trends of the response spectra and the peak amplitude at critical periods, showing consistently lower prediction errors compared to the physics-based model across various sites. Among the models, the transformer model exhibits the smallest average prediction error owing to its ability to capture long-range dependencies in the seismic data with its self-attention mechanism. The LSTM models also demonstrated good performance due to their ability to manage sequential dependencies. The 1D-CNN, while generally effective, occasionally shows higher errors,

probably due to the limited ability to handle the long-range sequential dependencies. All the data-driven models exhibit efficient computation times of less than 0.4 s for estimation. The results provide new insights into the potential of sequence deep learning models for improving seismic ground response estimation accuracy.

While our results highlight the potential of data-driven approaches for seismic response modeling, the current models are trained for each site to consider the site-specific condition. In other words, if one model that is trained on a specific site is applied to another site for prediction, the model is not likely to consider the site characteristics properly. Our future research will focus on developing an integrated model that can consider the site characteristics across different sites in the prediction.

Author Contributions: Conceptualization, Y.C. (Yongjin Choi) and J.A.; formal analysis, Y.C. (Yongjin Choi), H.-T.N., T.H.H., Y.C. (Youngjin Choi), and J.A.; investigation, Y.C. (Yongjin Choi), H.-T.N. and T.H.H.; methodology, Y.C. (Yongjin Choi) and Y.C. (Youngjin Choi); project administration, J.A.; validation, Y.C. (Yongjin Choi) and J.A.; writing—original draft, Y.C. (Yongjin Choi); writing—review and editing, Y.C. (Youngjin Choi). All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Korea Institute of Marine Science & Technology Promotion (KIMST) funded by the Ministry of Oceans and Fisheries (20220364), and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2022R1I1A3069043).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Shegay, A.V.; Miura, K.; Fujita, K.; Tabata, Y.; Maeda, M.; Seki, M. Evaluation of seismic residual capacity ratio for reinforced concrete structures. *Resilient Cities Struct.* **2023**, *2*, 28–45. [\[CrossRef\]](#)
2. Yan, Z.; Ramhormozian, S.; Clifton, G.C.; Zhang, R.; Xiang, P.; Jia, L.-J.; MacRae, G.A.; Zhao, X. Numerical studies on the seismic response of a three-storey low-damage steel framed structure incorporating seismic friction connections. *Resilient Cities Struct.* **2023**, *2*, 91–102. [\[CrossRef\]](#)
3. Rathje, E.M.; Kottke, A.R.; Trent, W.L. Influence of Input Motion and Site Property Variabilities on Seismic Site Response Analysis. *J. Geotech. Geoenviron. Eng.* **2010**, *136*, 607–619. [\[CrossRef\]](#)
4. Barani, S.; De Ferrari, R.; Ferretti, G. Influence of soil modeling uncertainties on site response. *Earthq. Spectra* **2013**, *29*, 705–732. [\[CrossRef\]](#)
5. Kaklamanos, J.; Baise, L.G.; Thompson, E.M.; Dorfmann, L. Comparison of 1D linear, equivalent-linear, and nonlinear site response models at six KiK-net validation sites. *Soil Dyn. Earthq. Eng.* **2015**, *69*, 207–219. [\[CrossRef\]](#)
6. Ordonez, G.A. *SHAKE2000: A Computer Program for the 1D Analysis of Geotechnical Earthquake Engineering Problems*; Geomotions, LLC: Lacey, WA, USA, 2000.
7. Astroza, R.; Pastén, C.; Ochoa-Cornejo, F. Site response analysis using one-dimensional equivalent-linear method and Bayesian filtering. *Comput. Geotech.* **2017**, *89*, 43–54. [\[CrossRef\]](#)
8. Zalachoris, G.; Rathje, E.M. Evaluation of one-dimensional site response techniques using borehole arrays. *J. Geotech. Geoenviron. Eng.* **2015**, *141*, 04015053. [\[CrossRef\]](#)
9. Hashash, Y.M.A.; Park, D. Non-linear one-dimensional seismic ground motion propagation in the Mississippi embayment. *Eng. Geol.* **2001**, *62*, 185–206. [\[CrossRef\]](#)
10. Zheng, W.; Luna, R. *Nonlinear Site Response Analysis in the New Madrid Seismic Zone*; University of Missouri: Rolla, MO, USA, 2004.
11. Kwok, A.O.; Stewart, J.P.; Hashash, Y.M.; Matasovic, N.; Pyke, R.; Wang, Z.; Yang, Z. Use of exact solutions of wave propagation problems to guide implementation of nonlinear seismic ground response analysis procedures. *J. Geotech. Geoenviron. Eng.* **2007**, *133*, 1385–1398. [\[CrossRef\]](#)
12. Park, D.; Hashash, Y.M. Evaluation of seismic site factors in the Mississippi Embayment. I. Estimation of dynamic properties. *Soil Dyn. Earthq. Eng.* **2005**, *25*, 133–144. [\[CrossRef\]](#)
13. Huang, J.; McCallen, D. Applicability of 1D site response analysis to shallow sedimentary basins: A critical evaluation through physics-based 3D ground motion simulations. *Earthq. Eng. Struct. Dyn.* **2024**, *53*, 2876–2907. [\[CrossRef\]](#)

14. Özcebe, A.; Smerzini, C.; Paolucci, R.; Pourshayegan, H.; Plata, R.R.; Lai, C.; Zuccolo, E.; Bozzoni, F.; Villani, M. On the comparison of 3D, 2D, and 1D numerical approaches to predict seismic site amplification: The case of Norcia basin during the M6. 5 2016 October 30 earthquake. In *Earthquake Geotechnical Engineering for Protection and Development of Environment and Constructions*; CRC Press: Boca Raton, FL, USA, 2019; pp. 4251–4258.
15. Zhang, W.; Dong, Y.; Crempien, J.G.F.; Arduino, P.; Kurtulus, A.; Taciroglu, E. A comparison of ground motions predicted through one-dimensional site response analyses and three-dimensional wave propagation simulations at regional scales. *Earthq. Spectra* **2024**, *40*, 1215–1234. [[CrossRef](#)]
16. Lam, R.; Sanchez-Gonzalez, A.; Willson, M.; Wirnsberger, P.; Fortunato, M.; Alet, F.; Ravuri, S.; Ewalds, T.; Eaton-Rosen, Z.; Hu, W. Learning skillful medium-range global weather forecasting. *Science* **2023**, *382*, 1416–1421. [[CrossRef](#)] [[PubMed](#)]
17. Salman, A.G.; Kanigoro, B.; Heryadi, Y. Weather forecasting using deep learning techniques. In Proceedings of the 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Depok, Indonesia, 10–11 October 2015; pp. 281–285.
18. Choi, Y.; Kumar, K. Graph Neural Network-based surrogate model for granular flows. *Comput. Geotech.* **2024**, *166*, 106015. [[CrossRef](#)]
19. Choi, Y.; Kumar, K. Inverse analysis of granular flows using differentiable graph neural network simulator. *arXiv* **2024**, arXiv:2401.13695. [[CrossRef](#)]
20. Pfaff, T.; Fortunato, M.; Sanchez-Gonzalez, A.; Battaglia, P.W. Learning mesh-based simulation with graph networks. *arXiv* **2020**, arXiv:2010.03409.
21. Fayaz, J.; Galasso, C. A deep neural network framework for real-time on-site estimation of acceleration response spectra of seismic ground motions. *Comput.-Aided Civil. Infrastruct. Eng.* **2022**, *38*, 87–103. [[CrossRef](#)]
22. Akhani, M.; Kashani, A.R.; Mousavi, M.; Gandomi, A.H. A hybrid computational intelligence approach to predict spectral acceleration. *Measurement* **2019**, *138*, 578–589. [[CrossRef](#)]
23. Campbell, K.W.; Bozorgnia, Y. NGA ground motion model for the geometric mean horizontal component of PGA, PGV, PGD and 5% damped linear elastic response spectra for periods ranging from 0.01 to 10 s. *Earthq. Spectra* **2008**, *24*, 139–171. [[CrossRef](#)]
24. Abrahamson, N.A.; Silva, W.J. Empirical response spectral attenuation relations for shallow crustal earthquakes. *Seismol. Res. Lett.* **1997**, *68*, 94–127. [[CrossRef](#)]
25. Hu, J.; Ding, Y.; Lin, S.; Zhang, H.; Jin, C. A Machine-Learning-Based Software for the Simulation of Regional Characteristic Ground Motion. *Appl. Sci.* **2023**, *13*, 8232. [[CrossRef](#)]
26. Kiranyaz, S.; Avci, O.; Abdeljaber, O.; Ince, T.; Gabbouj, M.; Inman, D.J. 1D convolutional neural networks and applications: A survey. *Mech. Syst. Signal Process.* **2021**, *151*, 107398. [[CrossRef](#)]
27. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
28. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 6000–6010.
29. Choi, Y.; Kumar, K. A machine learning approach to predicting pore pressure response in liquefiable sands under cyclic loading. In Proceedings of the Geo-Congress 2023, Los Angeles, CA, USA, 26–29 March 2023; pp. 202–210.
30. Zhang, P.; Yin, Z.Y.; Jin, Y.F.; Ye, G.L. An AI-based model for describing cyclic characteristics of granular materials. *Int. J. Numer. Anal. Methods Geomech.* **2020**, *44*, 1315–1335. [[CrossRef](#)]
31. Zhang, N.; Shen, S.-L.; Zhou, A.; Jin, Y.-F. Application of LSTM approach for modelling stress–strain behaviour of soil. *Appl. Soft Comput.* **2021**, *100*, 106959. [[CrossRef](#)]
32. Hong, S.; Nguyen, H.-T.; Jung, J.; Ahn, J. Seismic Ground Response Estimation Based on Convolutional Neural Networks (CNN). *Appl. Sci.* **2021**, *11*, 760. [[CrossRef](#)]
33. Li, L.; Jin, F.; Huang, D.; Wang, G. Soil seismic response modeling of KiK-net downhole array sites with CNN and LSTM networks. *Eng. Appl. Artif. Intell.* **2023**, *121*, 105990. [[CrossRef](#)]
34. Liao, Y.; Lin, R.; Zhang, R.; Wu, G. Attention-based LSTM (AttLSTM) neural network for Seismic Response Modeling of Bridges. *Comput. Struct.* **2023**, *275*, 106915. [[CrossRef](#)]
35. Zhang, Q.; Guo, M.; Zhao, L.; Li, Y.; Zhang, X.; Han, M. Transformer-based structural seismic response prediction. In *Structures*; Elsevier: Amsterdam, The Netherlands, 2024; Volume 61. [[CrossRef](#)]
36. Aoi, S.; Kunugi, T.; Fujiwara, H. Strong-motion seismograph network operated by NIED: K-NET and KiK-net. *J. Jpn. Assoc. Earthq. Eng.* **2004**, *4*, 65–74. [[CrossRef](#)]
37. Schnabel, P.B. *SHAKE, a Computer Program for Earthquake Response Analysis of Horizontally Layered Sites*; Report No. EERC 72-12; University of California: Berkeley, CA, USA, 1972.
38. Fei Li, X. Comparative Analysis of Two Seismic Response Analysis Programs in the Actual Soft Field. *Int. J. Eng.* **2020**, *33*, 784–790.
39. Hoult, R.D.; Lumantarna, E.; Goldsworthy, H.M. Ground motion modelling and response spectra for Australian earthquakes. In Proceedings of the Australian Earthquake Engineering Society 2013 Conference, Hobart, TAS, Australia, 15–17 November 2013.
40. Lasley, S.; Green, R.; Rodriguez-Marek, A. Comparison of equivalent-linear site response analysis software. In Proceedings of the 10th US National Conference on Earthquake Engineering, Anchorage, AK, USA, 21–25 July 2014.
41. Idriss, I.M.; Seed, H.B. Seismic Response of Horizontal Soil Layers. *J. Soil Mech. Found. Div.* **1968**, *94*, 1003–1031. [[CrossRef](#)]

42. Seed, H.B.; Wong, R.T.; Idriss, I.; Tokimatsu, K. Moduli and damping factors for dynamic analyses of cohesionless soils. *J. Geotech. Eng.* **1986**, *112*, 1016–1032. [[CrossRef](#)]
43. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, *8*, 1–74.
44. Bengio, Y.; Simard, P.; Frasconi, P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Netw.* **1994**, *5*, 157–166. [[CrossRef](#)]
45. Zhang, R.; Chen, Z.; Chen, S.; Zheng, J.; Büyüköztürk, O.; Sun, H. Deep long short-term memory networks for nonlinear structural seismic response prediction. *Comput. Struct.* **2019**, *220*, 55–68. [[CrossRef](#)]
46. Chollet, F. *Deep Learning with Python*; Simon and Schuster: New York, NY, USA, 2021.
47. Ba, J.L.; Kiros, J.R.; Hinton, G.E. Layer normalization. *arXiv* **2016**, arXiv:1607.06450.
48. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
49. Gao, Y.; Yang, J.; Qian, H.; Mosalam, K.M. Multiattribute multitask transformer framework for vision-based structural health monitoring. *Comput.-Aided Civ. Infrastruct. Eng.* **2023**, *38*, 2358–2377. [[CrossRef](#)]
50. Shan, J.; Huang, P.; Loong, C.N.; Liu, M. Rapid full-field deformation measurements of tall buildings using UAV videos and deep learning. *Eng. Struct.* **2024**, *305*, 117741. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.