Broadband Ground Motion Synthesis by Diffusion Model with Minimal Condition

Jaeheun Jung [*1] **Jaehyuk Lee** [*1] **Changhae Jung** [1] **Hanyoung Kim** [1] **Bosung Jung** [1] **Donghun Lee** [1]

Abstract

Shock waves caused by earthquakes can be devastating. Generating realistic earthquake-caused ground motion waveforms help reducing losses in lives and properties, yet generative models for the task tend to generate subpar waveforms. We present High-fidelity Earthquake Groundmotion Generation System (HEGGS) and demonstrate its superior performance using earthquakes from North American, East Asian, and European regions. HEGGS exploits the intrinsic characteristics of earthquake dataset and learns the waveforms using an end-to-end differentiable generator containing conditional latent diffusion model and hi-fidelity waveform construction model. We show the learning efficiency of HEGGS by training it on a single GPU machine and validate its performance using earthquake databases from North America, East Asia, and Europe, using diverse criteria from waveform generation tasks and seismology. Once trained, HEGGS can generate three dimensional E-N-Z seismic waveforms with accurate P/S phase arrivals, envelope correlation, signal-to-noise ratio, GMPE analysis, frequency content analysis, and section plot analysis.

1. Introduction

Broadband ground motion caused by seismic waves is crucial in the study of earthquakes and geology since it includes important features related to subsurface structures of the solid Earth. At the same time, it is a great challenge from signal processing perspective, as observed ground motion waveform signals cover a wide frequency band and are caused by rare and unevenly distributed earthquake events.

As the size of systematically recorded seismic waveforms grew, various improvements in seismological applications were made by analyzing historically observed seismic waveforms. For example, the accuracy of earthquake analysis was improved, early warning systems for earthquake-prone areas were polished, and earthquake-resistant architectural designs became more robust. Recently deep learning found successful applications in seismology (Mousavi & Beroza, 2022), such as seismic signal denoising (Zhu et al., 2019), fault recognition (An et al., 2021), and earthquake event detection (Mousavi et al., 2020; Saad et al., 2023).

However, the field still faces a shortage of data, particularly for large-scale earthquakes as they are rare (Shi et al., 2024; Katsanos et al., 2010). Recently, deep-learning based synthesis of seismic waveforms has emerged as a potential solution, mostly employing GAN-based generative models conditioned with various geological and seismological information (Wang et al., 2021; Florez et al., 2022; Li et al., 2024; Chen et al., 2024). However, the synthesized waveforms from these models often lack seismological realism, such as phase arrival times and amplitude of ground motions.

We see this problem as a mixed artifact of conditioned generation model architecture and unreliable conditioning information. Hence, we propose adapting diffusion model architecture (Sohl-Dickstein et al., 2015; Ho et al., 2020), which have shown superior realism and stability in image generation, to seismic waveform data in order to present a novel generation system for seismologically realistic ground motions with bare minimum set of conditioning information.

Our Contribution

- We design a novel seismic waveform generation model, HEGGS, that requires a bare-minimum set of conditional information on the earthquake and the observation point.
- We demonstrate constructing datasets for HEGGS using openly available seismic datasets, such that observed waveforms are paired with the source earthquake and time-aligned to the earthquake origin time.
- We design an end-to-end training method for the model and demonstrate its effectiveness by learning to gen-

[*]Equal contribution [1]Department of Mathematics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, Republic of Korea. Correspondence to: Donghun Lee <holy@korea.ac.kr>.

Proceedings of the 42^{nd} International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

erate high-fidelity earthquake ground motions from earthquakes in North America, East Asia, and Europe.

- We validate the superior fidelity of generated samples from HEGGS against benchmark models in various perspectives including seismology-inspired metrics such as GMPE analysis and phase arrival prediction.

2. Key Idea

Our goal is to robustly synthesize the broadband ground motion data with high level of seismological realism when compared to actual seismograph-recorded waveforms caused by earthquake events. We cast this seismic waveform synthesis problem into conditional waveform generation task to learn from seismic waveform databases, with minimal dependency on conditional information.

We use the following as the minimal conditional info:

1. s_{lat}, s_{lon} : latitude and longitude of the station to observe the waveform data.
2. e_{lat}, e_{lon} : latitude and longitude of the hypocenter.
3. e_{dep} : depth of the hypocenter, in kilometers.
4. M_L : local magnitude of the earthquake.

This set of information is usually considered insufficient for ground motion synthesis. For example, seismological properties of the source earthquake such as focal mechanism or local geological properties of observation point such as V_{S30} are often required in prior works. Instead of additionally demanding the often expensive-to-obtain info, we desire a generative model that learns directly from seismic waveforms with *minimal conditional info* as metadata.

When an earthquake happens, its shock waves propagate and are recorded by nearby seismographs as three-dimensional seismic waveforms named seismograms. Naturally, these seismograms are correlated by the common information about the source earthquake as well as the information specific to each of the seismographs such as geological conditions around the observation location. Regional seismic waveform datasets contain multiple waveform observations that can be paired to a source earthquake event, as illustrated in the left panel of Figure 1. The observations paired to the same earthquake would share properties of the same source earthquake, *while containing different information pertaining to their respective observation location*. This is how we liberate ourselves from asking additional conditional info.

We exploit this intrinsic pair-ability of the seismic waveform datasets, and construct paired waveform-metadata datasets from three earthquake databases from different continents: INSTANCE (Michelini et al., 2021) from Europe, KMA (Han et al., 2023) from East Asia, and SCEDC (SCEDC, 2013) from North America. Raw waveforms and corresponding metadata including locations, earthquake ID, magnitude and earthquake occurrence time are collected, and the processing of each datasets are detailed in Appendix A.

3. Method

Inspired by a conditional music generation method (Ghosal et al., 2023), our method first creates spectrograms with a diffusion model and then convert spectrograms into waveforms. Our generative model fully exploits the pair-ability of seismic waveform datasets shown in Section 2 to train both the diffusion process for spectrogram generation and the high-fidelity decoder for waveform generation. We name the method HEGGS, an acronym for High-fidelity Earthquake Groundmotion Generation System.

3.1. Pair-Exploiting Diffusion Model

For each earthquake event, we sample a pair of waveforms (W^{src}, W^{tgt}) from dataset and convert it to spectrograms (X^{src}, X^{tgt}) and construct conditional vector of target station $\vec{c}_{tgt}$ by preprocessing.

Let $q(x_{1:T}; X)$ be the forward process of the diffusion model, and consider two trajectories $q(x_t^{src}|X^{src})$ and $q(x_t^{tgt}|X^{tgt})$. Recall that X^{src} and X^{tgt} shares the property of earthquake, we may assume that from X^{src} and $\vec{c}_{tgt}$ we can gather enough features of earthquake to generate X^{tgt}. In this approach, we may consider the transform map $\eta(x_t^{src}, \vec{c}_{tgt}, t)$ for $t > 0$ which maps the latent of input X^{src} to the latent of target X^{tgt} as a random variable, with following assumption:

$$\eta(x_t^{src}, \vec{c}_{tgt}, t) \sim q(x_t^{tgt}|X^{tgt}). \tag{1}$$

Referring (Salimans & Ho, 2022), the loss function $\mathcal{L}_{DM}$ of diffusion model in $\mathbf{x}$-space (sample space) is:

$$\mathcal{L}_{DM} = \mathbb{E}_{(X^{tgt}, \vec{c}_{tgt}), \epsilon, t} \|X^{tgt} - \mathbf{x}_\theta(x_t^{tgt}, \vec{c}_{tgt}, t)\|^2. \tag{2}$$

while the SNR weight is simplified.

Considering the Equation (1), we rewrite the loss function as

$$\mathcal{L}'_{DM} = \mathbb{E}_{(X^{src}, X^{tgt}, \vec{c}_{tgt}), \epsilon, t} \|X^{tgt} - \mathbf{m}_\theta(x_t^{src}, \vec{c}_{tgt}, t)\|^2 \tag{3}$$

where

$$\mathbf{m}_\theta(x, \vec{c}, t) = \mathbf{x}_\theta(\eta(x, \vec{c}, t), \vec{c}, t). \tag{4}$$

Hence, we predict $\mathbf{m}_\theta$ by neural network, which is a composition of latent transform function and denoising model.

For the sampling of the reverse process, we exploit the same procedure of the denoising process of diffusion, as

$$x_{t-1}^{tgt} = \tilde{\mu}_t(x_t^{tgt}, \mathbf{m}_\theta(x_t^{tgt}, \vec{c}_{tgt}, t)) + \sigma_t \mathbf{z}, \mathbf{z} \sim N(0, I) \tag{5}$$

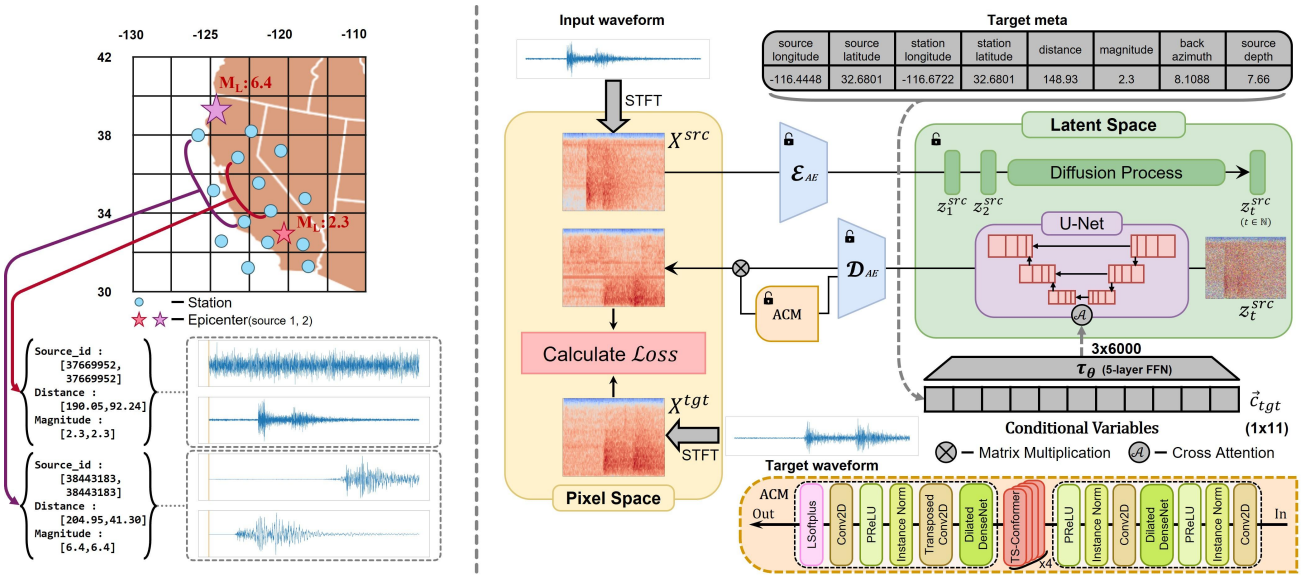


Figure 1. Left : Visualization of SCEDC data using paired waveforms. It shows that earthquake events can be detected at greater distances depending on magnitude. Right : Diagram of the waveform generative model architecture of HEGGS and its training loss.

where $\tilde{\mu}_t(x_t, x_0)$ is mean vector of $q(x_{t-1}|x_t, x_0)$, introduced in Eq. (7) of (Ho et al., 2020).

This is equivalent to conventional denoising process, as

$$\eta(x_t^{tgt}, \vec{c}_{tgt}, t) \overset{d}{=} x_t^{tgt} \tag{6}$$

by assumption and thus

$$\mathbf{m}_\theta(x_t^{tgt}, \vec{c}_{tgt}, t) = \mathbf{x}_\theta(x_t^{tgt}, \vec{c}_{tgt}, t). \tag{7}$$

Therefore, pair-exploiting training process of HEGGS allows the diffusion model to generate X^{tgt} from the Gaussian noise $x_T^{tgt} \sim \mathcal{N}(0, I)$ following conventional reverse process with $\mathbf{m}_\theta$.

3.2. End-to-end Model Training

From the idea of LDM (Rombach et al., 2022), we consider the autoencoder comprised of a downsampling module $\mathcal{E}_{AE}$ and an upsampling module $\mathcal{D}_{AE}$, and construct diffusion model on latent space with smaller dimension. If there were a suitable pretrained autoencoder, the LDM loss would be

$$\mathcal{L}'_{LDM} = \mathbb{E}_{(Z^{src}, Z^{tgt}, \vec{c}_{tgt}), \epsilon, t} \| Z^{tgt} - \mathbf{m}_\theta(z_t^{src}, \vec{c}_{tgt}, t) \|^2 \tag{8}$$

where $Z = \mathcal{E}_{AE}(X)$, z_t^{src} is latent of diffusion process of Z^{src}.

There is no suitable encoder-decoder model for seismic waveforms, so we modify Equation (8) into an end-to-end loss function as shown below:

$$\mathcal{L}_{ours} := \mathbb{E}_{(X^{src}, X^{tgt}, \vec{c}_{tgt}), \epsilon, t} \| X^{tgt} - \mathcal{D}_{AE}(\mathbf{m}_\theta(z_t^{src}, \vec{c}_{tgt}, t)) \|^2 \tag{9}$$

where $z_t^{src} = \sqrt{\overline{\alpha}_t}\mathcal{E}_{AE}(X^{src}) + \sqrt{1 - \overline{\alpha}_t}\epsilon$.

Using $\mathcal{L}_{ours}$ as the loss function, we train the waveform generation model end-to-end, covering the encoder, the diffusion module, and the decoder with ACM (Amplitude Correction Module), as shown in the right panel of Figure 1. For the detailed implementation in the diffusion module, we used a U-Net backbone for $\mathbf{m}_\theta$, brought $\mathcal{E}_{AE}$ and $\mathcal{D}_{AE}$. More details on the specifications of HEGGS and its training recipe can be found in Appendix B.

After training diffusion model with HEGGS, we generate waveform with conventional reverse process by setting z_T^{tgt} by Gaussian noise or Z^{src}. The details with pseudocode of training and generation, can be found in Appendix J.

4. Empirical Verification

We showcase the performance of HEGGS in three cases:

1. generate waveforms of existing earthquakes at existing observation stations $\vec{c}_{tgt}$, using an observed earthquake information W^{src} as input waveform.
2. generate at arbitrary locations $\vec{c}'_{tgt}$, using an observed earthquake information W^{src} as input waveform.
3. generate waveforms of fictitious earthquake information $\vec{c}'_{tgt}$ (also, without W^{src}).

The first case is designed to verify the fidelity of HEGGS, by comparing the generated samples to ground truth waveforms. We present quantitative results in Section 4.1 and qualitative analyses with visualizations in Sections 4.2.1 to 4.2.3. The results from the second case are presented throughout Sections 4.1 and 4.2, and in Section 4.2.4 we show the results from the third case.

4.1. Quantitative Evaluation

To assess and compare the effectiveness of models synthesizing seismic waves, we conducted a comprehensive quantitative analysis focusing on key parameters including P-wave and S-wave arrival times, GMPE analysis, and similarity measures such as envelope correlation, spectrogram image similarity, as well as signal-to-noise ratio (SNR), and peak signal-to-noise ratio (PSNR). Specifically, we compare generated synthetic waveform W^{pred} from $\vec{c}_{tgt}$ and compare it with corresponding ground truth waveform W^{tgt} to compute each metric.

For comparison, we also trained the following benchmark models on SCEDC: Seismogen (Wang et al., 2021), Conseisgen (Li et al., 2024), BBGAN (Florez et al., 2022) and LDM (Rombach et al., 2022). Since the input shape of waveform W^{tgt} or spectrogram X^{tgt} is different from each of the models, we give reasonable modifications to them for the training and evaluation. The detailed changelogs are listed in Appendix E.

4.1.1. Phase Arrival Times

The arrival times of P-wave and S-wave are the most basic, but important properties of seismic waveforms, as determining P-wave and S-wave from the seismogram is the first step of earthquake analysis. Since we want to assess the fidelity of the generated waveforms, we fine-tune the SeisBench (Woollam et al., 2022) implementation of EQTransformer (EQT) (Mousavi et al., 2020) on each dataset to use it in labeling phase arrivals times and compare them with the ground truth arrival times. The detailed training recipe we used for EQT is given in Appendix D. We present the performance measure of the finetuned EQT in the test dataset in Table 1. Note that the waveforms are inherently normalized as preprocessing step of EQT prediction.

Table 1. Performance of EQT picker on each test dataset. Samples with errors less than 0.5s are considered to be positive for F1 score computation.

Dataset	P_MAE(s)	S_MAE(s)	P_F1	S_F1
SCEDC	0.1116	0.2189	0.9728	0.9384
KMA	0.0993	0.1362	0.9635	0.9624
INSTANCE	0.1738	0.3151	0.9797	0.9099

Using the finetuned EQT as the phase picker, we label generated synthetic waveforms at a random station for every earthquake event from the test dataset. The mean absolute error (MAE) metrics of the P wave and the S wave phase arrival times from the ground truth labels are reported in the first two columns of Table 2. The resulting phase arrival times of the P wave and the S wave are considered to be close to the ground-truth arrival times on all datasets, as the MAE values are measured to be significantly small while other benchmark models failed to generate earthquake event signals on correct arrival time. Notably, generating with input waveform W^{src} gives better results compared to generating without W^{src} (i.e. with noise), as the observation W^{src} contains earthquake-specific information.

4.1.2. Similarity Measures

We also compare the synthesized waveforms and corresponding spectrogram directly to the ground truth waveforms. We use general-purpose similarity measures: the envelope correlation, SNR and PSNR for the waveforms, and MSE for the spectrograms.

Envelope correlation was calculated to measure the similarity between the envelopes of synthesized and observed seismic waves, providing insights into the overall waveform fidelity. We applied Savitzky-Golay Filtering (Savitzky & Golay, 1964) technique with polyorder 3 before calculating the envelope correlation. In implementation, we exploit (Beyreuther et al., 2010) implementation of cross correlation, which includes the waveform normalization. Furthermore, SNR and PSNR metrics were employed to evaluate the quality of the synthesized seismic waves in terms of signal clarity and fidelity to the original data. Additionally, we compare the synthesized spectrogram X^{syn} and spectrogram of observed seismic signals X^{tgt} to quantify their similarity using image similarity. We normalized both spectrograms to compare

The results are summarized in Table 2. The proposed method outperforms the benchmark models on almost all similarity metrics, which may imply that the generated X^{pred} and W^{pred} are more realistic in most cases. These quantitative analyses provide a comprehensive assessment of how similar HEGGS-generated waveforms are to the actual observed seismic ground motion.

4.1.3. GMPE Analysis

Ground Motion Prediction Equation (GMPE) is a widely used mathematical modeling in seismology that predicts the ground shaking intensity caused by earthquakes, and it is crucial for seismic hazard assessment and earthquake-resistant structural engineering. The GMPE model relates earthquake parameters, like local magnitude M_L and hypocentral distance R_{hypo}, to ground motion characteris-

Table 2. Results of quantitative analysis. Models were evaluated with W^{src} when it is trained with paired data, otherwise without W^{src}, except (*): evaluated without W^{src}, while the model was trained with paired data.

Dataset	Model	Input	Waveform					Spectrogram
			P_MAE (s)	S_MAE (s)	*env.corr*	SNR	PSNR	MSE
SCEDC	SeismoGen (Wang et al., 2021)	w/o W^{src}	1.9558	3.6246	0.4895	-8.6166	23.5431	1.4124
		w/ W^{src}	1.8426	3.3325	0.5454	-8.6282	23.6354	0.8063
	ConSeisGen (Li et al., 2024)	w/o W^{src}	3.9724	6.8992	0.3246	-8.6216	23.6416	0.7461
		w/ W^{src}	3.9102	6.8055	0.2980	-8.5341	23.5329	0.9356
	BBGAN (Florez et al., 2022)	w/o W^{src}	6.4210	10.416	0.1950	-3.0093	23.7598	1.6150
		w/ W^{src}	diverged					
	LDM (Rombach et al., 2022)	w/o W^{src}	1.1142	1.7294	0.6932	-3.0202	**24.7573**	0.2838
		w/ W^{src}	0.5633	0.7808	0.7726	-3.0015	19.6269	0.2426
	HEGGS (ours)	(*)w/o W^{src}	0.5025	0.8003	0.7963	-2.9891	24.6816	0.1531
		w/ W^{src}	**0.4760**	**0.5476**	**0.8187**	**-2.0051**	24.6553	**0.1512**
KMA	LDM (Rombach et al., 2022)	w/o W^{src}	1.6233	2.1125	0.7703	-3.0006	25.3883	0.3521
		w/ W^{src}	1.3521	1.6845	0.8076	-2.9989	26.3658	0.3785
	HEGGS (ours)	(*)w/o W^{src}	0.2988	0.5551	0.8769	-3.0034	26.2769	0.1867
		w/ W^{src}	**0.2763**	**0.4644**	**0.8785**	**-2.9768**	**26.8730**	**0.1682**
INSTANCE	LDM (Rombach et al., 2022)	w/o W^{src}	0.8417	0.7847	0.7921	-3.0062	22.0767	0.2927
		w/ W^{src}	0.8187	0.7875	0.7898	**-2.9904**	22.0956	0.2841
	HEGGS (ours)	(*)w/o W^{src}	0.5192	0.6804	0.8299	-3.0004	22.3690	0.1376
		w/ W^{src}	**0.5085**	**0.6378**	**0.8301**	-2.9976	**22.6870**	**0.1308**

tics, such as Peak Ground Acceleration (PGA). Since M_L is obtained from the peak amplitude of the waveforms, the GMPE analysis shows how the scale of the generated waveforms from HEGGS matches the real observations.

Computing PGA (Emolo et al., 2015; Boore et al., 2014; Lanzano et al., 2019a;b) is closely related to local magnitude M_L, which would be calculated by distinct formula (Han et al., 2023; Uhrhammer et al., 2011; Di Bona, 2016) for each region. We follow the conventions in geoscience research to decide PGA computation formula, which is detailed in Appendix F.

GMPE analysis result, shown in Figure 2, reveals that the synthetic seismic waveforms generated by HEGGS closely resemble how the observed ground truths in distribution. Also, the similarity in PGA values indicates that the magnitude of synthesized waveforms is similar to the ground truth.

4.2. Qualitative Evaluation

We perform qualitative analyses to evaluate the seismological fidelity and reliability of HEGGS-generated waveforms.

4.2.1. Waveform Analysis

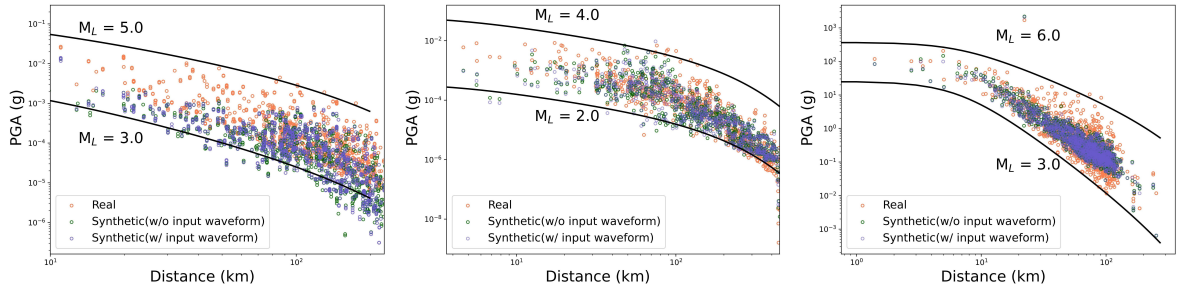
Synthetic waveforms can be visually inspected alongside real seismic waveforms for similarities in terms of waveform morphology, including amplitude, shape, and duration of seismic signals.

Figure 3 compares a set of representative synthesized waveforms and real waveforms from three datasets. Notably, both synthesized waveform and real waveform depict similar patterns of seismic activity, including distinct seismic phases and their corresponding arrivals. This alignment underscores the effectiveness of our approach in accurately replicating the seismic signal's morphology and temporal evolution. More waveform examples can be found in Appendix H.

4.2.2. Spectrogram Comparison

We also show the output spectrogram of HEGGS, compared to the spectrogram of the ground truth waveform to examine their time-frequency characteristics. This provides insights into the similarities of temporal distribution of energy across different frequency bands.

In Figure 4 we show the comparison of spectrograms, where each corresponds to the waveforms in Figure 3. Upon com-



(a) SCEDC (b) KMA (c) INSTANCE

Figure 2. Result of GMPE analysis in PGA values with respect to the distance. The points in the figures represent the PGA values calculated from randomly selected waveforms from the test set containing earthquakes filtered by the magnitude range indicated by the black solid lines, and synthetic waveforms using the corresponding metadata. The subfigures correspond to the earthquake source: SCEDC (North America), KMA (East Asia), and INSTANCE (Europe).

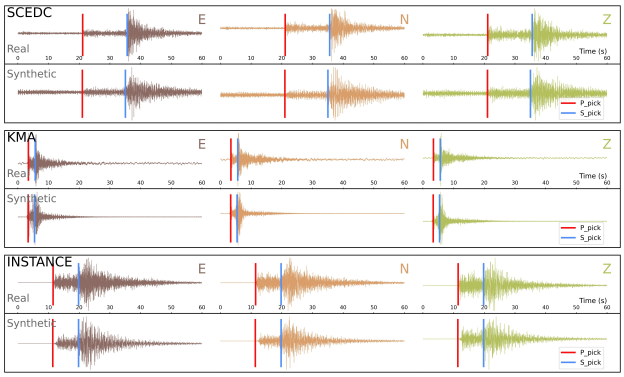
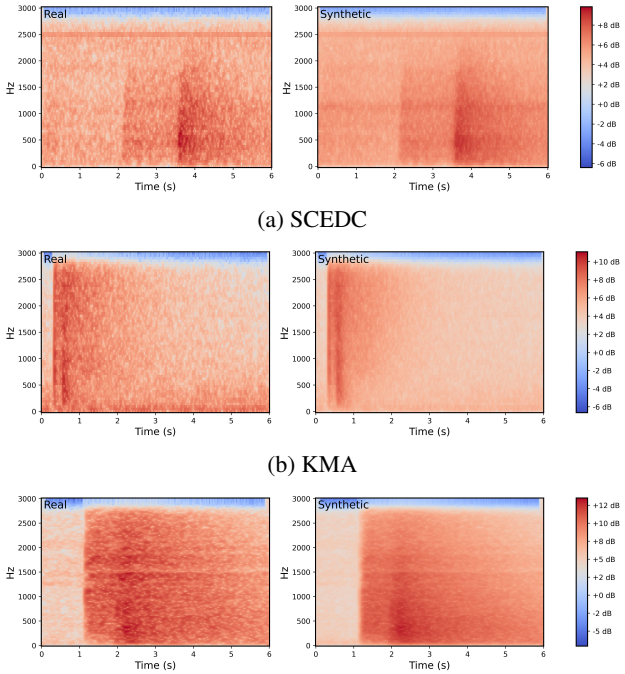


Figure 3. Comparison of 3-component real and synthetic waveforms from earthquake datasets SCEDC (North America), KMA (East Asia), and INSTANCE (Europe). For each panel, top shows a real waveform and the bottom shows a synthetic waveform generated with the same metadata. The phase arrivals marked as red (P) and blue (S) lines are detected by EQT.



(a) SCEDC

(b) KMA

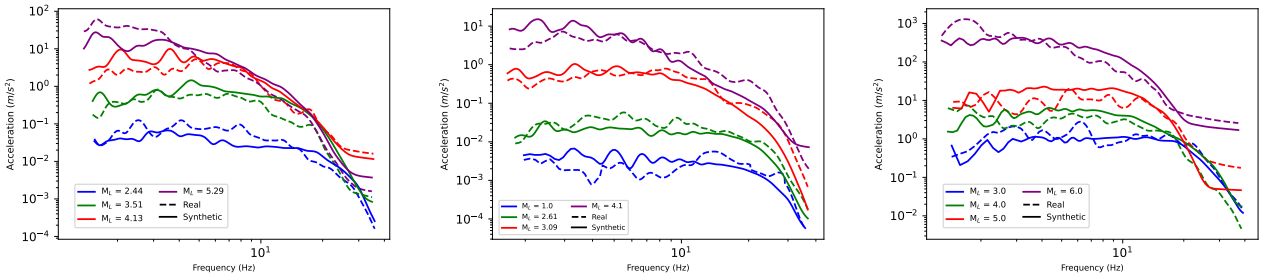
(c) INSTANCE

Figure 4. Comparison of real and synthetic spectrograms.

paring the synthesized spectrogram with the real spectrogram, several key observations come to light. Both spectrograms exhibit remarkable similarities in terms of phase arrival times and frequency band distribution, indicative of the efficacy of our synthesis approach in capturing essential seismic signal characteristics. However, it is discernible that the synthesized spectrogram exhibits a slightly lower resolution compared to the real spectrogram, with some details appearing less defined. This reduction in resolution is particularly evident in the depiction of fine-scale frequency variations and subtle signal features. Despite this difference, the overall agreement between the synthesized and real spectrograms underscores the fidelity of our synthesis method in reproducing the fundamental characteristics of seismic signals. More spectrogram examples are shown in Appendix H.

4.2.3. Frequency Content Analysis

We also analyze how the energy released during an earthquake is retained in different frequencies. This analysis is closely related to the concept of corner frequency (Boore, 1983) and the seismic moment (M_0). The corner frequency is generally associated with the earthquake's magnitude. Specifically, the corner frequency identifies the point at which high-frequency energy begins to decline sharply, indicating that larger earthquakes generally have lower corner frequencies. The M_0 represents the total energy released by the earthquake, which corresponds to an increase in amplitude on the spectrum as the earthquake's magnitude increases. By comparing synthetic and observed seismic



(a) SCEDC (b) KMA (c) INSTANCE

Figure 5. (a)-(c): Frequency contents of synthetic waveform compared to the real waveform

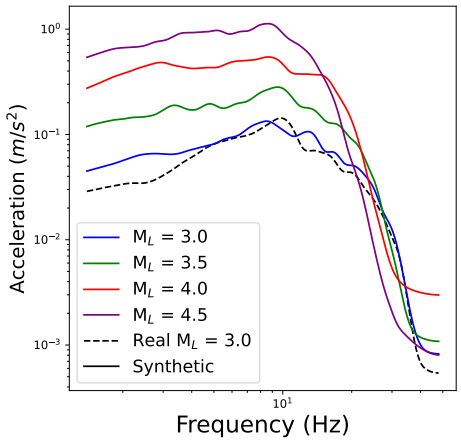


Figure 6. Magnitude Manipulation

signals, we aim to evaluate the similarity between the two characteristics of corner frequency and M_0 across different magnitudes.

We apply both Fast Fourier Transform (FFT) and Konno-Ohmachi-smoothing (Konno & Ohmachi, 1998) technique to enhance our comparison. Also, we apply Wood-Anderson simulations (Havskov & Ottemöller, 2010) to compare results from distinct stations. The resulting spectra are shown in Figures 5a to 5c. We observe significant differences in corner frequency and M_0 across different earthquake magnitudes. We also observe the trend of reduced corner frequency reduces and increased M_0 as the magnitude grows.

4.2.4. Magnitude Manipulation

Synthesis of waveforms from fictitious earthquake is difficult challenging problem in DL-seismology area, especially with large magnitude, since the seismological features of large earthquake is hard to capture and large magnitude earthquake data is very rare, which requires the extrapolation ability of the model. We select c_{tgt} from the test dataset, change M_L, and generate waveform with modified c'_{tgt}, without W^{src} and analyze the frequency contents in Figure 6. The result is quite promising: the corner frequency gets smaller and M_0 gets larger properly when the magnitude grows, which is consistent to the theory as explained in (Geller, 1976).

4.2.5. Waveform Analysis on Synthetic Stations

By arranging virtual observation stations in a linear manner, spatial variations of seismic waves could be observed, facilitating an understanding of seismic event characteristics. The synthesized seismic waves reflected seismic activity at the virtual observation stations, enabling exploration of subsurface structures and seismic wave propagation characteristics.

The sections in Figure 7 represented the positions of observation stations horizontally and represented the temporal and frequency characteristics of seismic activity vertically. Through such visualization, comparisons between synthesized and observed seismic waves could be conducted, assessing the fidelity of the synthesized seismic waves in reflecting seismic events. Results from section plots clearly visualized spatial and temporal variations of seismic activity, serving as crucial criteria for evaluating the extent to which HEGGS accurately reproduces actual observed results. More section plot examples of seismic events are provided in Appendix I.

5. Ablation Studies

Compared to the conventional latent diffusion model, we introduced two major components, the efficient learning framework and amplitude correction module, to generate high-quality seismic waveforms. In this section, we present the results of the ablation study to evaluate the role of each component, on SCEDC dataset.

To assess the effectiveness of this approach, we conduct training of the LDM (Rombach et al., 2022) with two distinct training schemes: the original and modified one trained by Equation (8) with paired data. Unfortunately, conventional LDM training on our dataset was diverged. Hence we tried to train LDM to generate normalized waveform, which

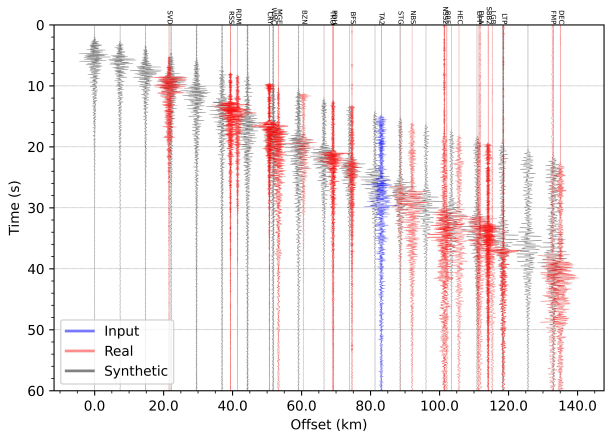


Figure 7. Section plots comparing synthetic and real waveforms. The vertical capital letters displayed at the top of the plot are the station ID that observed each real waveform.

Table 3. Results of ablation study. * represents the generation of normalized waveform. $env.corr$ refers to the envelope correlation between synthesized waveform and real waveform.

Model	P_MAE (s)	S_MAE (s)	$env.corr$
LDM*	1.1142	1.7294	0.6932
+paired data*	0.5633	0.7808	0.7726
+end-to-end train	0.8014	1.5367	0.6239
+ACM (HEGGS)	0.4760	0.5476	0.8187
LDM+ACM	1.1131	1.6372	0.6981
+paired data	0.7748	0.9402	0.7965

is a relaxed version of our task.

After that, we tried to generate unnormalized waveforms, by changing the training framework. Preserving the model architecture, we trained the model which has same architecture, but by end-to-end training Equation (9). The only difference between this model and ours is the amplitude correction module ACM.

The results can be found in Table 3. On first two rows, learning with paired data were very effective to increase the quality of waveform, especially as the phase arrival times were twice as accurate. Comparing 2nd and 3rd rows, the overall scores seem to be worse, but the model in 2nd row often generates unrealistic waveforms in qualitative analysis results in Appendix G. Also, note that the result of 3rd row is the result of unnormalized waveform generation while 2nd row generates normalized waveform. Even the difficulty of generation problems were increased, the paired training shows better results, compared to the baseline model LDM. This may indicate the failure of VAE pretraining that pretrained VAE could not capture the amplitude as important feature. The amplitude correction module ACM helps to improve the quality of seismic waveform synthesis, as shown in the 3rd and 4th rows of Table 3.

Thanks to reviewers, we found that the ACM has ability to allow LDM trainable with unnormalized waveforms, as shown in 5th row of Table 3. The last row of Table 3 shows the remarkable improvement induced by pair-exploiting strategy, but still not better than HEGGS with end-to-end training.

6. Discussion

The HEGGS training method, which takes advantage of the seismic dataset characteristic by training the model with paired data, allows for two modes of generation criteria: with and without W^{src}. Although HEGGS is trained using W^{src}, the fidelity of generation without W^{src} is promising, and much better than the benchmark models. Generation without W^{src} allows us to synthesize waveforms of non-existent earthquakes and simulate the ground motion with different magnitude or location, which is big challenge in seismology. As shown in Section 4.2.4, HEGGS shows the theoretically-expected trend of corner frequency and M_0, but may not be perfect since we only used minimal condition about location and magnitude. We expect larger success with additional geological features, which we did not included in minimal condition, for this non-existent earthquake synthesis challenge.

The seismic synthesis studies are inevitably build on regional dataset, since each observatories are operated independently by each government and thus the waveform formats are unaligned. Another challenging problem arises here to build global model by training multiple models on individual dataset, with consideration of robust consistency, especially on border. We expect HEGGS would be the effective starting point of this research direction.

Compared to other methods, HEGGS shows superior fidelity with minimal conditions, especially for the P/S phase arrival times. We expect HEGGS would applicable to downstream tasks which is sensitive to the phase arrival times,such as early warning systems, earthquake modeling and disaster, hence we are planning to develop algorithms for those downstream tasks using HEGGS as a near-future research.

7. Conclusion

In this paper, we propose HEGGS, an efficient training framework for seismic waveform synthesis utilizing a diffusion model and a minimal set of conditions. Our approach generates seismic waveforms using only readily accessible information, such as location and magnitude, thereby avoiding the need for extra conditions.

To empirically validate the proposed method, we constructed a seismic dataset from the SCEDC, INSTANCE and KMA dataset by collecting simultaneously paired observations aligned with the earthquake's origin time. We demonstrate that HEGGS produces more realistic waveforms than existing benchmark models by applying seismic domain-specific metrics, such as envelope correlation and P/S phase arrival times, for expert-level comparison and applications.

Impact statement

Our work enables high-fidelity seismic waveform synthesis, enhancing earthquake modeling, early warning systems, and disaster preparedness while promoting AI use in geophysical research.

Acknowledgements

The authors would like to thank Prof. Seongryong Kim and Dr. Jongwon Han of Seismology Lab, Korea University for their professional opinions and feedback on the seismological aspects of this work. The authors are supported in part by the Korea Meteorological Administration Research and Development Program under Grant KMI2021-01112 (RS-2021-KM211112).

References

An, Y., Guo, J., Ye, Q., Childs, C., Walsh, J., and Dong, R. Deep convolutional neural network for automatic fault recognition from 3d seismic datasets. *Computers & Geosciences*, 153:104776, 2021. ISSN 0098-3004. doi: https://doi.org/10.1016/j.cageo.2021.104776. URL https://www.sciencedirect.com/science/article/pii/S0098300421000807.

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., and Wassermann, J. Obspy: A python toolbox for seismology. *Seismological Research Letters*, 81(3):530–533, 2010.

Boore, D. M. Stochastic simulation of high-frequency ground motions based on seismological models of the radiated spectra. *Bulletin of the Seismological Society of America*, 73(6A):1865–1894, 1983.

Boore, D. M., Stewart, J. P., Seyhan, E., and Atkinson, G. M. Nga-west2 equations for predicting pga, pgv, and 5% damped psa for shallow crustal earthquakes. *Earthquake Spectra*, 30(3):1057–1085, 2014.

Chen, G., Li, J., and Guo, H. Deep generative model conditioned by phase picks for synthesizing labeled seismic waveforms with limited data. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–15, 2024. doi: 10.1109/TGRS.2024.3384768.

Di Bona, M. A local magnitude scale for crustal earthquakes in italy. *Bulletin of the Seismological Society of America*, 106(1):242–258, 01 2016. ISSN 0037-1106. doi: 10.1785/0120150155. URL https://doi.org/10.1785/0120150155.

Emolo, A., Sharma, N., Festa, G., Zollo, A., Convertito, V., Park, J.-H., Chi, H.-C., and Lim, I.-S. Ground-motion prediction equations for south korea peninsula. *Bulletin of the Seismological Society of America*, 105(5):2625–2640, 2015.

Esser, P., Rombach, R., and Ommer, B. Taming transformers for high-resolution image synthesis, 2020.

Florez, M. A., Caporale, M., Buabthong, P., Ross, Z. E., Asimaki, D., and Meier, M. Data-Driven Synthesis of Broadband Earthquake Ground Motions Using Artificial Intelligence. *Bulletin of the Seismological Society of America*, 112(4):1979–1996, 04 2022. ISSN 0037-1106. doi: 10.1785/0120210264. URL https://doi.org/10.1785/0120210264.

Geller, R. J. Scaling relations for earthquake source parameters and magnitudes. *Bulletin of the Seismological Society of America*, 66(5):1501–1523, 10 1976. ISSN 0037-1106. doi: 10.1785/BSSA0660051501. URL https://doi.org/10.1785/BSSA0660051501.

Ghosal, D., Majumder, N., Mehrish, A., and Poria, S. Text-to-audio generation using instruction guided latent diffusion model. In *Proceedings of the 31st ACM International Conference on Multimedia*, pp. 3590–3598, 2023.

Han, J., Joo Seo, K., Kim, S., Sheen, D., Lee, D., and Byun, A. Research Catalog of Inland Seismicity in the Southern Korean Peninsula from 2012 to 2021 Using Deep Learning Techniques. *Seismological Research Letters*, 95(2A):952–968, 12 2023. ISSN 0895-0695. doi: 10.1785/0220230246. URL https://doi.org/10.1785/0220230246.

Havskov, J. and Ottemöller, L. *Magnitude*, pp. 151–191. Springer Netherlands, Dordrecht, 2010. ISBN 978-90-481-8697-6. doi: 10.1007/978-90-481-8697-6_6. URL https://doi.org/10.1007/978-90-481-8697-6_6.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

Katsanos, E. I., Sextos, A. G., and Manolis, G. D. Selection of earthquake ground motion records: A state-of-the-art review from a structural engineering perspective. *Soil dynamics and earthquake engineering*, 30(4):157–169, 2010.

Konno, K. and Ohmachi, T. Ground-motion characteristics estimated from spectral ratio between horizontal and vertical components of microtremor. *Bulletin of the Seismological Society of America*, 88(1):228–241, 1998.

Lanzano, G., Luzi, L., Pacor, F., Felicetta, C., Puglia, R., Sgobba, S., and D'Amico, M. A revised ground-motion prediction model for shallow crustal earthquakes in italy. *Bulletin of the Seismological Society of America*, 109(2): 525–540, 02 2019a. ISSN 0037-1106. doi: 10.1785/0120180210. URL https://doi.org/10.1785/0120180210.

Lanzano, G., Luzi, L., Pacor, F., Puglia, R., Felicetta, C., D'Amico, M., and Sgobba, S. Update of the ground motion prediction equations for italy. In *Proceedings of the 7th International Conference on Earthquake Geotechnical Engineering (7ICEGE)*, Rome, Italy, 2019b. International Society for Soil Mechanics and Geotechnical Engineering (ISSMGE). URL https://www.earth-prints.org/bitstream/2122/12973/1/7ICEGE_ITA18_final.pdf.

Li, Y., Yoon, D., Ku, B., and Ko, H. Conseisgen: Controllable synthetic seismic waveform generation. *IEEE Geoscience and Remote Sensing Letters*, 21:1–5, 2024. doi: 10.1109/LGRS.2023.3338652.

Lu, Y.-X., Ai, Y., and Ling, Z.-H. Mp-senet: A speech enhancement model with parallel denoising of magnitude and phase spectra. *arXiv preprint arXiv:2305.13686*, 2023.

McPhillips, D. F., Herrick, J. A., Ahdi, S., Yong, A. K., and Haefner, S. Updated compilation of vs30 data for the united states, 2020. URL https://www.sciencebase.gov/catalog/item/5f44290282ce4c3d1222da63. Accessed on Month Day, Year.

Michelini, A., Cianetti, S., Gaviano, S., Giunchi, C., Jozinović, D., and Lauciani, V. Instance – the italian seismic dataset for machine learning. *Earth System Science Data*, 13(12):5509–5544, 2021. doi: 10.5194/essd-13-5509-2021. URL https://essd.copernicus.org/articles/13/5509/2021/.

Mohinder S. Grewal, Lawrence R. Weill, A. P. A. *Appendix C: Coordinate Transformations*, pp. 456–501. John Wiley & Sons, Ltd, Hoboken, NJ, 2007. ISBN 9780470099728. doi: https://doi.org/10.1002/9780470099728.app3. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470099728.app3.

Mousavi, S. M. and Beroza, G. C. Deep-learning seismology. *Science*, 377(6607):eabm4470, 2022.

Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., and Beroza, G. C. Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature communications*, 11(1): 3952, 2020.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

Saad, O. M., Chen, Y., Siervo, D., Zhang, F., Savvaidis, A., Huang, G.-c. D., Igonin, N., Fomel, S., and Chen, Y. Eqcct: A production-ready earthquake detection and phase picking method using the compact convolutional transformer. *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

Salimans, T. and Ho, J. Progressive distillation for fast sampling of diffusion models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=TIdIXIpzhoI.

Savitzky, A. and Golay, M. J. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8):1627–1639, 1964.

SCEDC. Southern California Earthquake Center, 2013. URL https://dx.doi.org/10.7909/C3WD3xH1.

Sheen, D., Kang, T., and Rhie, J. A Local Magnitude Scale for South Korea. *Bulletin of the Seismological Society of America*, 108(5A):2748–2755, 07 2018. ISSN 0037-1106. doi: 10.1785/0120180112. URL https://doi.org/10.1785/0120180112.

Shi, Y., Lavrentiadis, G., Asimaki, D., Ross, Z. E., and Azizzadenesheli, K. Broadband Ground-Motion Synthesis via Generative Adversarial Neural Operators: Development and Validation. *Bulletin of the Seismological Society of America*, 114(4):2151–2171, 03 2024. ISSN 0037-1106. doi: 10.1785/0120230207. URL https://doi.org/10.1785/0120230207.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., and Ganguli, S. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pp. 2256–2265. PMLR, 2015.

Uhrhammer, R. A., Hellweg, M., Hutton, K., Lombard, P., Walters, A. W., Hauksson, E., and Oppenheimer, D. California Integrated Seismic Network (CISN) Local Magnitude Determination in California and Vicinity. *Bulletin of the Seismological Society of America*, 101(6):2685–2693, 12 2011. ISSN 0037-1106. doi: 10.1785/0120100106. URL https://doi.org/10.1785/0120100106.

Wang, F. pynga. https://github.com/fengw/pynga, 2012.

Wang, T., Trugman, D., and Lin, Y. Seismogen: Seismic waveform synthesis using gan with application to seismic data augmentation. *Journal of Geophysical Research: Solid Earth*, 126(4):e2020JB020077, 2021.

Woollam, J., Münchmeyer, J., Tilmann, F., Rietbrock, A., Lange, D., Bornstein, T., Diehl, T., Giunchi, C., Haslinger, F., Jozinović, D., Michelini, A., Saul, J., and Soto, H. SeisBench—A Toolbox for Machine Learning in Seismology. *Seismological Research Letters*, 93(3):1695–1709, 03 2022. ISSN 0895-0695. doi: 10.1785/0220210324. URL https://doi.org/10.1785/0220210324.

Zhu, W., Mousavi, S. M., and Beroza, G. C. Seismic signal denoising and decomposition using deep neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 57(11):9476–9488, 2019. doi: 10.1109/TGRS.2019.2926772.

A. Dataset Construction

We used three datasets (SCEDC(SCEDC, 2013), KMA(Han et al., 2023) and INSTANCE(Michelini et al., 2021)) from different regions. In this section, we explain how each dataset was constructed. All datasets are collected from corresponding APIs and processed to have 60-seconds duration and applied $1 \sim 45$Hz bandpass filter.

We split each dataset into training dataset and test dataset, according to the earthquake event, to evaluate the fidelity of generated waveform for the earthquake which is unseen during the training.

Table 4. Features of each dataset

dataset	SCEDC		KMA		INSTANCE	
Features	Train	Test	Train	Test	Train	Test
#observations	71,488	17,878	237,755	58,925	72,904	19,872
#source event	2,098	525	2,052	514	2,265	593
#station	149	149	134	134	578	534
average #station per events	34.07	34.05	115.87	114.64	24.43	25.29
average magnitude	2.45	2.45	1.45	1.45	3.36	3.36
average epicentral distance	125.25	126.71	235.48	234.22	57.82	57.79
average focus depth	8.51	8.65	11.52	11.73	12.47	11.97

A.1. SCEDC

We exploit earthquake catalog of SCEDC (SCEDC, 2013) provided by SeisBench(Woollam et al., 2022). We selected waveforms with a sampling rate of 100Hz that included 60 seconds from the earthquake and applied a bandpass filter in the $1 \sim 45$Hz range to construct our data. Unfortunately, the Seisbench-provided dataset had fewer than 13 stations per earthquake events on average, therefore we utilized Obspy API(Beyreuther et al., 2010) to collect additional observations on more stations in the station list of (Uhrhammer et al., 2011) for each earthquake. Using earthquakes from the catalog during the years 2016 to 2019, we constructed a new dataset with approximately 34 stations per source. The Table 4 shows the count of datasets we used.

The V_{S30} information was sourced from (McPhillips et al., 2020) and used only during the GMPE analysis, not during the training or model inference processes. The average value was used if multiple V_{S30} values were present for a single station code. For station codes without V_{S30} data, $760m/s$ was assigned to negate the influence of V_{S30} during GMPE analysis.

A.2. KMA

KMA data source consist of continuous waveform data were employed, which are operated by KMA (Korea Meteorological Administration) and KIGAM (Korea Institute of Geoscience and Mineral Resources). We exploit the dataset appear in (Han et al., 2023) which is constructed from earthquake catalog provided by KMA, spanning from 2016-2020, and used subset consist of observations from broadband sensors. Similarly to SCEDC, the waveforms have a sampling rate of 100hz, a duration of 60 seconds, and a frequency $1 \sim 45$Hz.

A.3. INSTANCE

We used the Seisbench-provided version INSTANCE dataset and created a subset by selecting only the traces satisfying:

1. includes records for 60 seconds from the earthquake occurrence time
2. local magnitude is larger than 3.0.
3. P-arrival time is included in the metadata to ensure that the earthquake signal is observed.

For the EQT evaluation in Table 1, we excluded waveforms which include multiple event signals, which are out of our scope.

B. Implementation details

We implement the proposed model with following implementation details.

During the training, W_{src} is fixed for a specific earthquake source ID, and W_{tgt} is sampled from earthquakes with the same source ID. Among these, if metadata contained P/S phase labels, samples are randomly selected from those with labels. If P/S phase labels are absent, samples are chosen randomly without considering P/S phase labels. And also we conduct preprocessing of seismic data.

We implement using single NVIDIA-RTX A6000 with 48GB memory. For training, we set the number of epochs to 500 and the training batch size to 4. To enhance training efficiency, we apply an accumulation step 4, resulting in an effective batch size of 16. For the loss, we set the maximum diffusion steps to $T = 1000$ and SNR weight 5. We minimize the loss by AdamW optimizer with learning rate 10^{-5} and *pytorch.optim* defaults. During the training, we applied learning rate decaying technique with linear scheduler. The total duration of training is approximately 65 hours.

B.1. Neural Network Architecture

We utilize the U-Net backbone with cross-attention architecture similar to (Rombach et al., 2022; Ghosal et al., 2023), to represent $\mathbf{m}_\theta$, with modification in the domain-specific encoder τ_θ to map $\vec{c}_{tgt}$ to hidden feature $\tau_\theta(\vec{c}_{tgt})$. For the implementation, we construct τ_θ by 5-layer FFN model. The encoded conditional vector $\tau_\theta(\vec{c}_{tgt})$ will be provided as a value and key of cross attention module $Attn(Q, K, V)$ while U-Net feature is provided as query Q.

For $\mathcal{E}_{AE}$ and $\mathcal{D}_{AE}$, we take same architectures from VAE of (Esser et al., 2020) and give a modification on $\mathcal{D}_{AE}$. With the vanilla module $\mathcal{D}_{AE}$, we find that the proposed model is not effective in accurately predicting the amplitude of the output waveform. Therefore, we propose to attach an additional module ACM after $\mathcal{D}_{AE}$ to predict the amplitude correction feature and multiply it to the predicted spectrogram. In detail, we utilize the encoder, TSConformer blocks and Magnitude mask decoder module from MP-SeNet (Lu et al., 2023) and provide output of $\mathcal{D}_{AE}$ and auxiliary phase spectrogram induced by GriffinLim algorithm to correct the amplitude and enhance the quality of generation. Improving the original implementation (Lu et al., 2023) that allows only reducing the output, we add four TSConformer blocks and replace the final sigmoid activation function with Softplus function to provide the ability to increase as well.

C. Pre-processing Recipe

C.1. Conditional Vector Pre-processing

We explain the process of $\vec{c}_{tgt}$ constuction. Recall the variables that we are used to synthesize waveform are:

1. s_{lat}, s_{lon} : latitude and longitude of the station to observe the waveform data.
2. e_{lat}, e_{lon} : latitude and longitude of epicenter.
3. e_{dep} : depth of the hypocenter, unit of kilometers.
4. M_L : magnitude of the earthquake.

We preprocessed those variables to construct an 11-dimensional condition vector and later provide it to our condition encoder module τ_θ.

First of all, we encode locational information $s_{lat}, s_{lon}, e_{lat}$ and e_{lon} with the following process:

1. Normalize the values to get $s'_{lat}, s'_{lon}, e'_{lat}$ and e'_{lon} with following:

$$s'_{lat} = \frac{s_{lat} - l_{lat}}{u_{lat} - l_{lat}}, e'_{lat} = \frac{e_{lat} - l_{lat}}{u_{lat} - l_{lat}}, s'_{lon} = \frac{s_{lon} - l_{lon}}{u_{lon} - l_{lon}} \text{ and } e'_{lon} = \frac{e_{lon} - l_{lon}}{u_{lon} - l_{lon}} \tag{10}$$

where (l_{lat}, u_{lat}) and (l_{lon}, u_{lon}) represent the lower and upper bounds of latitude and longitude, respectively, for the region of interest.

In our datasets, we summarize those bounds in Table 5.

Table 5. upper and lower bounds of the region of interest

Dataset (region)	l_{lat}	u_{lat}	l_{lon}	u_{lon}
SCEDC (Southern California)	32.0	37.9	-121.0	-114.1
KR (South Korea)	33.12	38.60	124.64	131.87
INSTANCE (Italy)	35.00	48.03	5.32	20.01

2. Motivated from polar coordinate transformation(Mohinder S. Grewal, 2007), which is commonly used in GPS field, we further encode normalized coordinate to following:

$$\begin{aligned} c_{sta} &= (cos(s'_{lat})cos(s'_{lon}), sin(s'_{lat})cos(s'_{lon}), sin(s'_{lon})) \\ c_{epi} &= (cos(e'_{lat})cos(e'_{lon}), sin(e'_{lat})cos(e'_{lon}), sin(e'_{lon})) \end{aligned} \tag{11}$$

Secondly, we compute the back azimuth angle Azi and encode by

$$c_{azi} = (cos(Azi), sin(Azi)) \tag{12}$$

Lastly, we compute and normalized epicentral distance R_{epi}, focus depth d_s and magnitude M_L. Each are normalized by following formula:

	SCEDC	KMA	INSTANCE
R'_{epi}	$(R_{epi} - 125.542401)/55.810322$	$(R_{epi} - 219.91)/119.99$	$(R_{epi} - 57.8158)/31.7465$
d'_s	$(d_s - 8.564146)/4.658161$	$(d_s - 11.59)/5.40$	$(d_s - 12.3680)/13.2456$
M'_L	$(M_L - 2.0)/6.4$	$(M_L - 0.35)/5.24$	$(M_L - 3.0)/6.5$

Concatenating the processed features $c_{sta}, c_{epi}, c_{azi}, R'_{epi}, d'_s$ and M'_L, we get an 11-dimensional conditional vector $\vec{c}_{tgt}$ for our problem, the synthesis of seismic ground motion.

C.2. spectrogram construction

The generation target of out model is spectrogram, which is in time-frequency domain. We report the process of spectrogram construction as pre-processing. We employed the STFT (Short-Time Fourier Transform) with a hop length 16. Given that the spectrogram's scale is closely related to the earthquake's amplitude, we used an $nfft$ and $window\ length$ of 128 and applied a logarithmic scale transformation for better scale adjustment. Consequently, the original waveform data of size 3×6000 was reshaped into $3 \times 64 \times 376$.

D. EQT Training Details

We used EQTransformer (Mousavi et al., 2020) provided by SeisBench (Woollam et al., 2022). Starting from pre-trained model provided by SeisBench, we finetune the model with our dataset, with the same training protocol. After standardizing the waveforms, we trained the model using the Adam optimizer, with a batch size of 512 and a learning rate of 10^{-3}, for 100 epochs. Other hyperparameters of the optimizer were set to default. For hyperparameter search, the learning rate ranged from 10^{-2} to 10^{-5}, and the performance was best when it was 10^{-3}.

E. Details on Benchmark Models

E.1. SeismoGen (Wang et al., 2021)

SeismoGen is a CGAN-based model that generates waveforms conditioned on the presence of seismic events (e.g., P or S waves). The Discriminator takes both the waveform and the presence of seismic events as inputs. It then divides the waveform into high and low frequency components, analyzing each to determine if waveform is real or synthetic. SeismoGen used data from three stations in Oklahoma: V34A, V35A, and V36A, while we used data from 149 stations from SCEDC. Our synthesis approach used station and earthquake information instead of presence of seismic events. SeismoGen generated

waveforms as 40 seconds at 40Hz, but we aimed for 60 seconds at 100Hz. We used an input noise length of 1500 and added upsampling at the end of the first convolution layer. The basic training used noise as input, and for comparison with HEGGS, we also trained using waveform. When using waveforms, we modified each pipeline to utilize one ENZ channel. The hyper-parameters we used included the Generator learning rate and Discriminator learning rate are set to 10^{-4} and 10^{-6}, using the RMSprop optimizer over 3000 epochs. The λ is set to 10 when using noise and 15 when using the input waveform. We saved the best model based on envelope correlation. We experimented with learning rates ranging from 10^{-4} to 10^{-7}, using both Adam and RMSprop optimizers. The value of λ was tested at 5, 10, and 15. The best-performing combination of these parameters was selected for the final model. Additionally, the results reported in Table 2 reflect the best performance achieved across 30 iterations. Addressing the instability of the original method, we added the L1 loss Equation (13) from pix2pix(Isola et al., 2017) as an additional loss term to improve training stability.

E.2. ConSeisGen (Li et al., 2024)

ConSeisGen is an ACGAN-based model that generates waveforms conditioned on the epicentral distance. The Discriminator consists of two components: D_P, which learn determining whether the waveform is real or synthetic, and D_Q, which learn regression estimating the distance between the epicenter and the station. While ConSeisGen generated waveforms with 3 channels and a length of 4096, we aimed to generate waveforms with 3 channels and a length of 6000. We modified the first linear layer and removed upsampling in the final layer. ConSeisGen used KiK-net data, which began recording shortly before the arrival of the P-wave. However, the SCEDC data utilized in this model was recorded from the onset of the earthquake for a duration of 60 seconds. ConSeisGen generates waveforms based on the epicentral distance. However, waveforms can vary even at the same distance due to factors like magnitude and geological conditions. To generate waveforms for specific locations, we utilized minimal additional condition such as station data and source data along with the epicentral distance. The hyper-parameters we used included the Generator learning rate and Discriminator learning rate are set to 2×10^{-4} and 10^{-5}, using the Adam optimizer over 5000 epochs. Referring eq.4 of (Li et al., 2024), the loss function consists of Adversarial Loss, Regression Loss(L_{reg}), and Diversity Improvement Loss(L_{di}). The L_{reg} computes the $l1$ loss between D_Q's output and the condition vector, with the λ_{reg} set to 1. The L_{di} aims to prevent mode collapse by maximizing the distance between feature maps, with λ_{di} set to 10 when using noise and 5 when using waveforms. We experimented with learning rates ranging from 10^{-4} to 10^{-6}, using both Adam and RMSprop optimizers. The value of λ_{di} was tested at 5, 10, and 15, while λ_{reg} was fixed at 1. The best-performing combination of these parameters was selected for the final model. Additionally, the results reported in Table 2 reflect the best performance achieved across 30 iterations. Addressing the instability of the original method, we added the L1 loss Equation (13) from pix2pix(Isola et al., 2017) as an additional loss term to improve training stability.

$$L_{\mathrm{L1}}(G) = \mathbb{E}_{x,y,z}\left[\|x_{tgt} - G(z,y)\|_1\right] \tag{13}$$

E.3. BBGAN (Florez et al., 2022)

BBGAN is a conditional generative model within the Wasserstein GAN framework. The original conditions of BBGAN are V_{S30}, earthquake magnitude, and epicentral distance. We modified conditional vector to ours, add conditional vector encoder τ_θ to both generator and discriminator, modified the last upsample layer of generator to have scale factor 3 (original: 2), and lastly increased the number of hidden features of last convolution block of discriminator, corresponding to our waveform shape $(3, 6000)$. Those changes allows the model to generate $(3, 6000)$ shape waveform from the provided conditional vector. To further improve the performance, we replaced all relu activations of generator and leaky relu activations of discriminator to gelu activation. Additionally, while the original BBGAN paper utilized data from Japanese networks K-NET and KiK-net with earthquake magnitudes larger than 4.5, our approach employed data from the SCEDC (SCEDC, 2013) with earthquake magnitude larger than 2.0 for training. In the training process, we set 500 training epoch and batch size 32, and Adam optimizer with learning rate 5×10^{-7} and $\beta = (0.9, 0.999)$. Also the final loss function is composed of adversarial loss, L1 reconstruction loss, and a KL divergence term. The L1 regularization term was set to 25, and the KL regularization term was set to 0.01. For evaluation during the validation loop, envelope correlation was used as the performance metric. During the training, the linear learning rate decay technique was applied.

E.4. LDM (Rombach et al., 2022)

E.4.1. VAE (ESSER ET AL., 2020) PRETRAINING

Due to lack of pretrained weights of VAE trained on seismic spectrogram, we first need to train VAE to encode X^{tgt} and X^{src} to latent vector Z^{tgt} and Z^{src}.

Employing equation (25) of (Rombach et al., 2022), we set the loss function for VAE training is:

$$L_{total} = min_{\mathcal{E}_{AE},\mathcal{D}_{AE}}, max_{\psi}[L_{rec}(x, \mathcal{D}_{AE}((x))) - L_{adv}(\mathcal{D}_{AE}(\mathcal{E}_{AE}(x))) + logD_{\psi}(x) + \lambda_{kl}KL] \tag{14}$$

where λ_{kl} is low weighted Kullback-Libler regularization term by factor 10^{-6}.

Unfortunately, the VAE training on our spectrogram diverged, due to difficulty on magnitude processing. Therefore, we apply standardization on spectrogram to relax the problem. And the latent space size is $64 \times 16 \times 94$.

We report reconstruction performance of the Auto-encoder model using the proposed our metrics. The reconstruction performance results as follow in Table 6.

Table 6. Reconstruction result

	waveform					spectrogram
Model	P_MAE (s)	S_MAE (s)	envelope corr	SNR	PSNR	MSE
VAE	0.5155	0.7066	0.7567	-2.9984	25.1800	0.2459

E.4.2. LDM (ROMBACH ET AL., 2022)

We train LDM using the pretrained VAE Appendix E.4.1 and DDPM(Ho et al., 2020) scheduler. Additionally, the overall model architecture is adapted and modified base on the TANGO (Ghosal et al., 2023) model and code. But, while TANGO models incorporate text-encoded conditions through Large Language Model, the seismic data does not exist text conditions. Therefore, we employ our preprocessed conditions and apply our conditional vector encoder τ_θ for training. During model training, the learning target is set the samples from the DDPM scheduler. Training is conduct using two methods and training losses.

- Equation (15): not utilizing the characteristic of paired data
- Equation (16): utilizing the characteristic of paired data

We set the hyperparameters for the AdamW optimizer as follows: an initial learning rate 10^{-5} and $\beta = (0.9, 0.999)$, and a weight decay of 10^{-2} and adam epsilon 10^{-8}. Also, we apply the learning rate decaying technique with the linear scheduler. The training batch size is set to 4 with an accumulation step of 4, resulting in a total effective batch size of 16. The model is trained for 500 epochs. The results indicate that training with paired data outperforms training without paired data.

$$L'_{LDM} = \mathbb{E}_{(Z^{tgt},\vec{c}_{tgt}),\epsilon,t}\|Z^{tgt} - \mathbf{x}_\theta(z_t^{tgt}, \vec{c}_{tgt}, t)\| \tag{15}$$

$$L'_{LDM} = \mathbb{E}_{(Z^{src},Z^{tgt},\vec{c}_{tgt}),\epsilon,t}\|Z^{tgt} - \mathbf{m}_\theta(z_t^{src}, \vec{c}_{tgt}, t)\| \tag{16}$$

F. GMPE formula

In this section, we express the PGA formula for GMPE analysis. Given waveform W, we obtain local magnitude M_L first, and compute the PGA value later.

F.1. SCEDC(SCEDC, 2013)

Given the waveform W, the local magnitude M_L of SCEDC can be computed by using the following formula (equations 1 to 6 of (Uhrhammer et al., 2011))

$$\begin{aligned} M_L &= logA - logA_0(R_{hypo}) \text{ where} \\ -logA_0(R_{hypo}) &= 1.11logR_{hypo} + 0.00189 \times R_{hypo} + 0.591 + \sum_{n=1}^{6} \text{TP}(n) \times T(n,z). \end{aligned} \tag{17}$$

where A is amplitude of W and $A_0(r)$ is attenuation function of southern california region. The station adjustment term was not applied due to a lack of values for recently installed stations.

The TP(n) coefficients are

$$\begin{aligned} &\text{TP}(1) = +0.056, \quad \text{TP}(2) = -0.031, \\ &\text{TP}(3) = -0.053, \quad \text{TP}(4) = +0.080, \\ &\text{TP}(5) = -0.028, \quad \text{TP}(6) = +0.015, \end{aligned} \tag{18}$$

When z is

$$z(r) = 1.11366 \times log(r) - 2.00574, \tag{19}$$

$8 \leq r \leq 500$ to $-1 \leq z \leq +1$, $T(n,z)$ is the Chebyshev polynomial

$$T(n,z) = cos[n \times \arccos(z)]. \tag{20}$$

After determining the local magnitude M_L, we obtain the PGA value by equation 1 of (Boore et al., 2014), with pynga (Wang, 2012) implementation. Since HEGGS doesn't exploit the focal mechanism information, we set $mech$ and $rake$ to be 0, which represents unspecified.

F.2. KMA(Emolo et al., 2015)

For the KMA dataset, the local magnitude M_L can be computed by the following equation (equations 1 and 6 of (Sheen et al., 2018)):

$$\begin{aligned} M_L &= logA - logA_0 + S \\ -logA_0 &= 0.5869log(R_{epi}/100) + 0.001680(R_{epi} - 100) + 3 \end{aligned} \tag{21}$$

where A is the peak amplitude of the Wood-Anderson simulated waveform and S is station-wise correction term and R_{epi} is epicentral distance in kilometers.

After determining local magnitude M_L we obtain PGA value Y by (Emolo et al., 2015) with following formula for South Korea peninsula:

$$\begin{aligned} logY = &-3.16 + 0.75M_L \\ &- 0.72log\left[\sqrt{R_{epi}^2 + 3.7^2}\right] - 0.0034R_{epi} \end{aligned} \tag{22}$$

F.3. INSTANCE(Lanzano et al., 2019a;b)

the local magnitude M_L on the INSTANCE dataset can be computed by the following equation (equation 1 to 14 of (Di Bona, 2016)):

$$\begin{aligned} M_L &= logA - logA_0(R_{hypo}) + C \\ &= logA + 1.749log(R_{hypo}/100) + 0.0016(R_{hypo} - 100) + 2.9445 + C \end{aligned} \tag{23}$$

where A is the peak amplitude of the Wood-Anderson simulated waveform and C is the station-wise correction term.

After determining the local magnitude M_L, we obtain the PGA value by equations 1 to 5 of (Lanzano et al., 2019b) and 7 to 8 of (Lanzano et al., 2019a). In these equations, Moment Magnitude(M_W) was used to compute PGA. However, (Di Bona, 2016) proposed a formula that satisfies $M_L = M_W$, on average. Therefore, this study used M_L instead of M_W. Also, HEGGS doesn't exploit the focal mechanism information, we set the style of faulting SOF to 0, representing the normal fault type.

G. Qualitative Analysis on Ablation Models

This section is dedicated to the qualitative analysis of the SCEDC of the models mentioned in Table 3. The figure compares the Real observation (Real), HEGGS(w/ ACM), end-to-end train (w/o ACM), and LDM + paired data. The human-labeled P/S arrival times of the earthquake are indicated by orange and black lines, while the P/S arrival times detected by EQT for each waveform are shown in red and blue.

G.1. Positive samples

We first list the positive samples, which are the results that all models generated realistic waveforms with accurate phase arrivals.

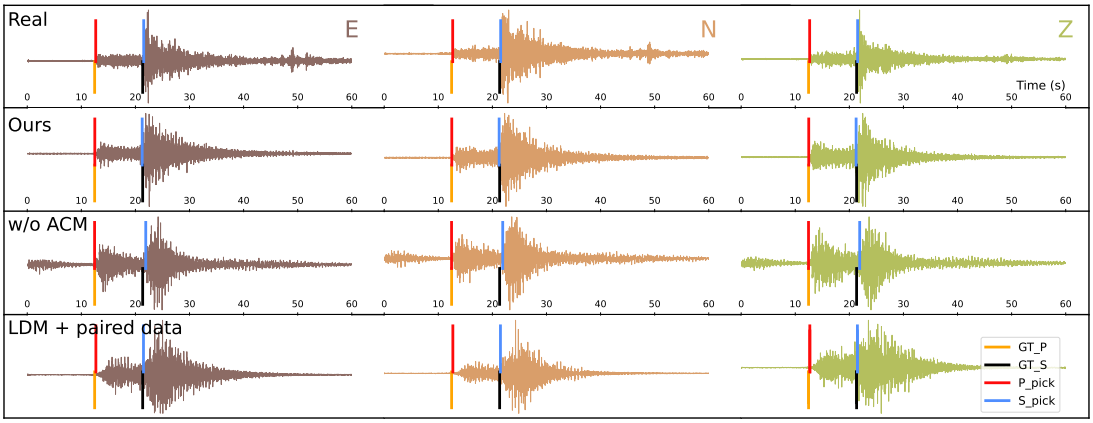


Figure 8. Positive synthesis results of our model and ablation models, compared to the real observation.

G.2. Negative Samples

We also include the results of the synthesis that at least one of the models failed to generate realistic and accurate waveforms.

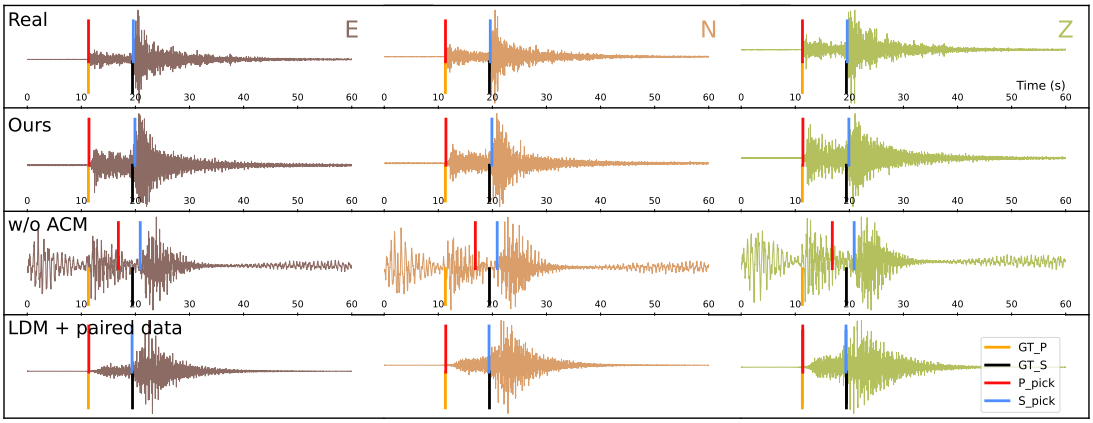
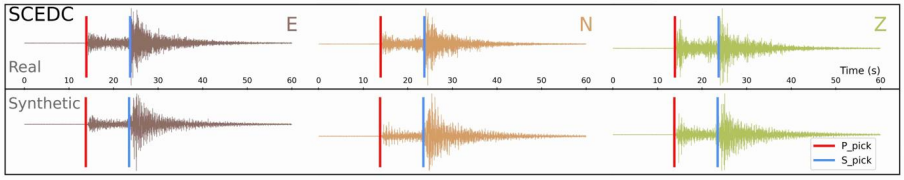


Figure 9. Negative synthesis results of our model and ablation models, compared to the real observation.

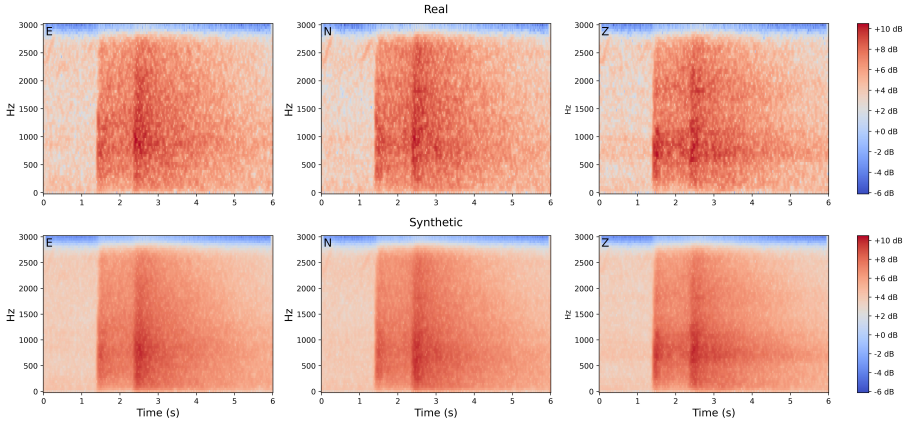
H. Additional Figures: Waveform and Spectrogram

This section presents the waveforms and spectrograms shown in Figure 3 and Figure 4. The seismic data we used consist of 3-components, ENZ. Each pair displays the same waveform and spectrogram, with the top representing the real observation and the bottom representing the synthetic generated HEGGS. The red and blue lines on the waveforms indicate the P/S arrival times.

H.1. SCEDC



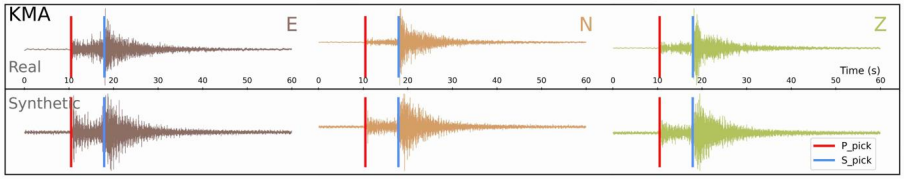
(a) waveform



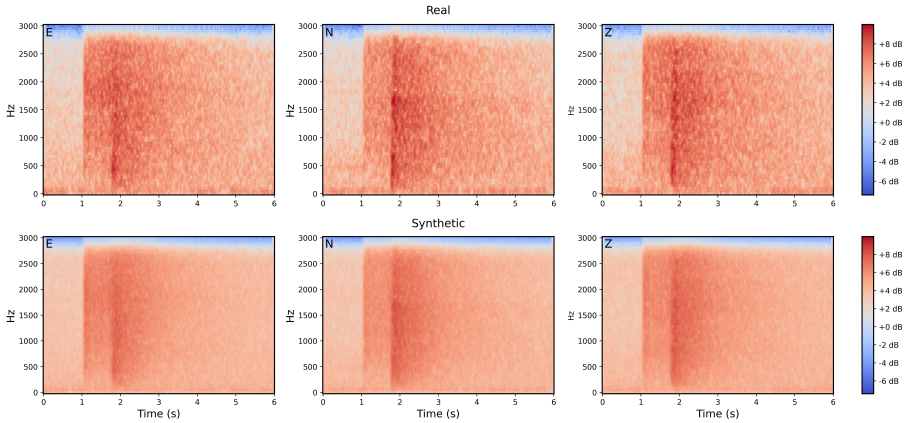
(b) spectrogram

Figure 10. Synthesis results of our model compared to the real observation.

H.2. KMA



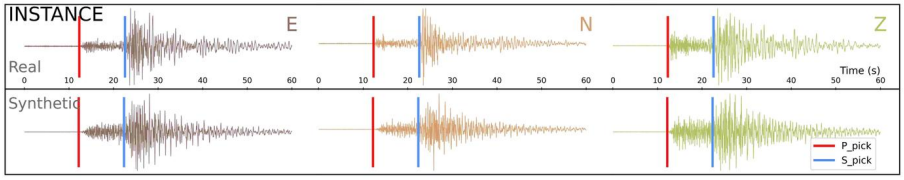
(a) waveform



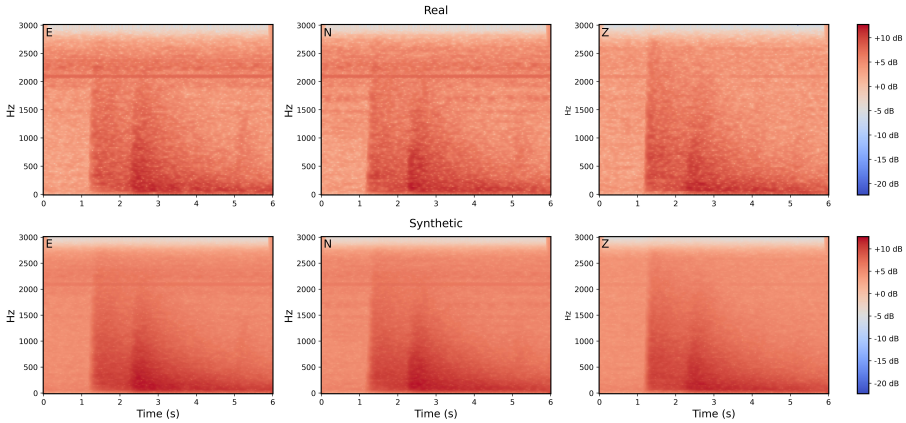
(b) spectrogram

Figure 11. Synthesis results of our model compared to the real observation.

H.3. INSTANCE



(a) waveform



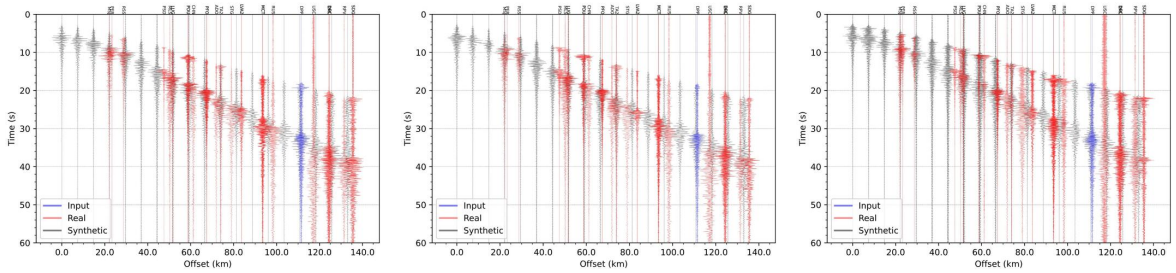
(b) spectrogram

Figure 12. Synthesis results of our model compared to the real observation.

I. Additional figures: Section plot

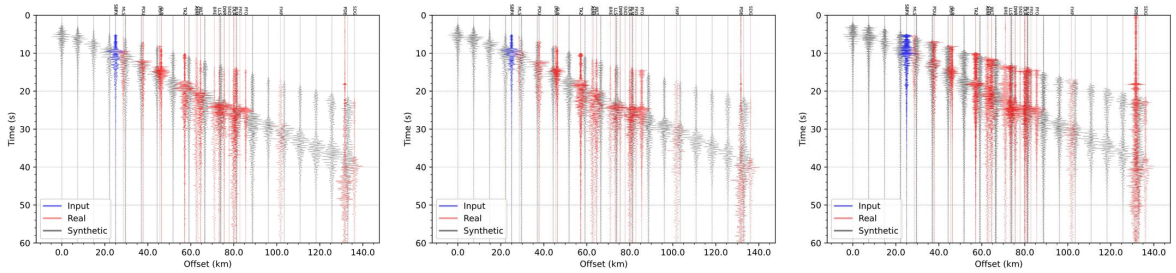
The section plot is constructed by following process. Initially, a specific earthquake event is chosen, and input data is randomly selected (indicated by the blue line). We set virtual stations established at equidistant intervals from the epicenter, generate waveforms, and plot together with real observations. The red lines represent ground truth observations and black lines are the synthesized waveforms. Note that the azimuth angle of observations varies, while the synthetic stations are set to have same values. This potentially affect the P/S wave arrivals and lead to mismatch in visualization, but the effect is not considered to be significantly large.

I.1. SCEDC



(a) Section plot on E axis (b) Section plot on N axis (c) Section plot on Z axis

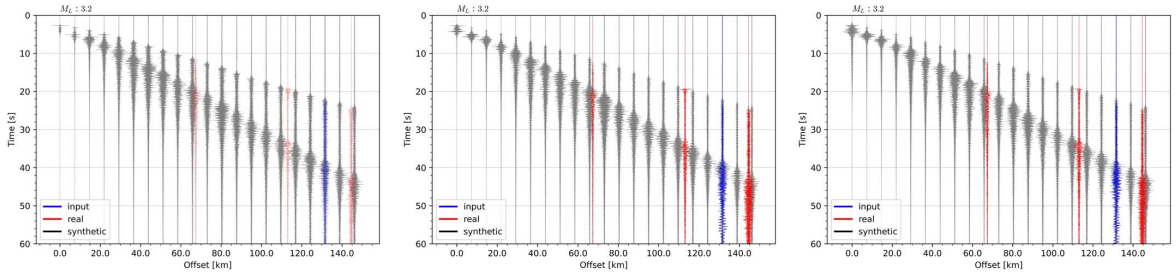
Figure 13. Section plot on synthetic stations.



(a) Section plot on E axis (b) Section plot on N axis (c) Section plot on Z axis

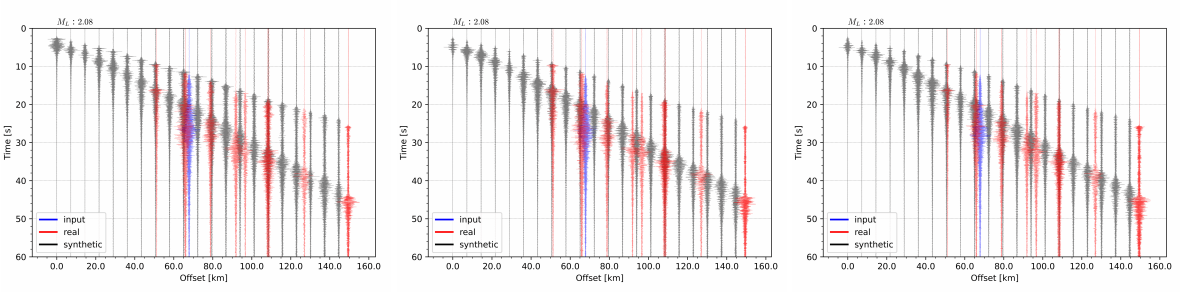
Figure 14. Section plot on synthetic stations.

I.2. KMA



(a) Section plot on E axis (b) Section plot on N axis (c) Section plot on Z axis

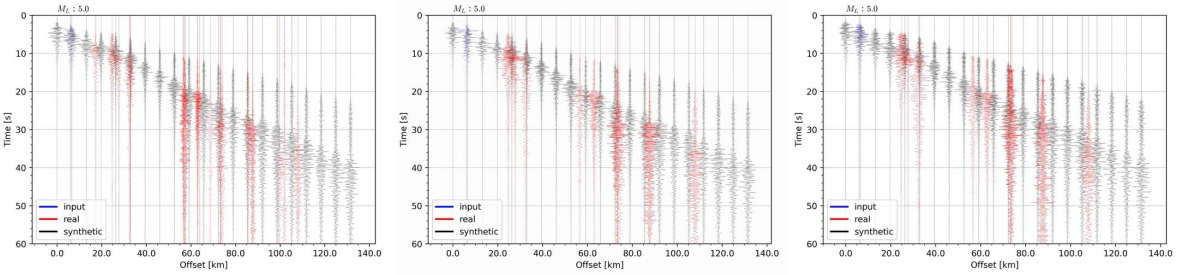
Figure 15. Section plot on synthetic stations.



(a) Section plot on E axis (b) Section plot on N axis (c) Section plot on Z axis

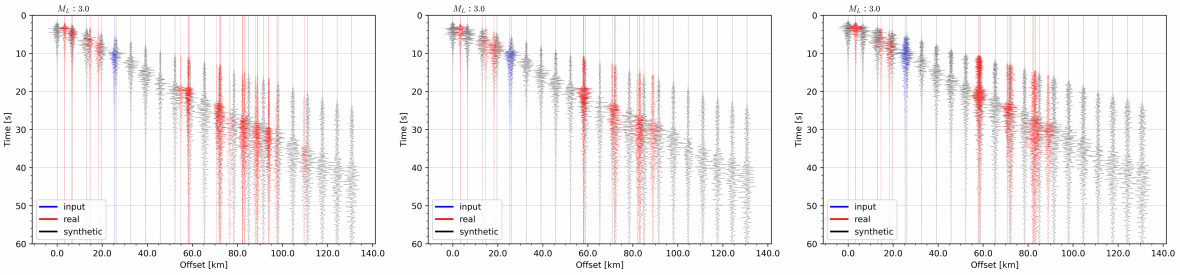
Figure 16. Section plot on synthetic stations.

I.3. INSTANCE



(a) Section plot on E axis (b) Section plot on N axis (c) Section plot on Z axis

Figure 17. Section plot on synthetic stations.



(a) Section plot on E axis (b) Section plot on N axis (c) Section plot on Z axis

Figure 18. Section plot on synthetic stations.

J. On pair-Exploiting Diffusion Model

In this appendix, we provide more detailed explanations about the training and inference of HEGGS for the clarification.

J.1. Remark:Prediction targets of diffusion model

Referring equation (2),(4),(6) and (7) of (Ho et al., 2020), the forward process $q(x_{1:T}; X)$ would be given by

$$q(x_t|x_{t-1}) = \mathcal{N}(\sqrt{1-\beta_t}x_{t-1}, \beta_t I), q(x_t|x_0) = \mathcal{N}(\sqrt{\overline{\alpha}_t}x_0, (1-\overline{\alpha}_t)I). \tag{24}$$

and thus we have

$$x_t = \sqrt{\overline{\alpha}_t}x_0 + \sqrt{1-\overline{\alpha}_t}\epsilon \text{ where } \epsilon \sim \mathcal{N}(0,1). \tag{25}$$

The backward process $q(x_{t-1}|x_t, x_0)$ would be

$$q(x_{t-1}|x_t, x_0) \sim \mathcal{N}(\tilde{\mu}(x_t, x_0), \tilde{\beta}_t I) \tag{26}$$

where

$$\tilde{\mu}_t(x_t, x_0) = \frac{\sqrt{\overline{\alpha}_{t-1}}\beta_t}{1-\overline{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}(1-\overline{\alpha}_{t-1})}{1-\overline{\alpha}_t}x_t \text{ and } \tilde{\beta}_t = \frac{1-\overline{\alpha}_{t-1}}{1-\overline{\alpha}_t}\beta_t. \tag{27}$$

In implementation, it is required to find $\tilde{\mu}_t(x_t, x_0)$ term in Equation (27). There are several methods for the prediction, with replacing x_0 by estimate $x_\theta(x_t, t)$. The HEGGS directly predicts x in sample space, as it would be more natural since we want to learn the morphology between paired data, compared to the alternatives which predicts the noise ϵ (Ho et al., 2020) or $\mathbf{v}$-prediction(Salimans & Ho, 2022).

Algorithm 1 HEGGS training

Input: Seismic dataset $\mathbb{D}$, diffusion steps T
repeat
  $(W^{src}, W^{tgt}, \vec{c}_{tgt}) \sim \mathbb{D}$
  convert (W^{src}, W^{tgt}) to (X^{src}, X^{tgt})
  $t \sim Uniform(1, \cdots, T)$
  $\epsilon \sim \mathcal{N}(0,1)$
  Take gradient descent step on
    $\nabla \|X^{tgt} - \mathcal{D}_{AE}(\mathbf{m}_\theta(z_t^{src}, \vec{c}_{tgt}, t))\|^2$
    where $z_t^{src} = \sqrt{\overline{\alpha}_t}\mathcal{E}_{AE}(X^{src}) + \sqrt{1-\overline{\alpha}_t}\epsilon$
until converged

Algorithm 2 Generation

Input: Diffusion steps T,condition vector $\vec{c}_{tgt}$, source waveform W^{src} (optional)
if W^{src} is given **then**
  convert W^{src} to spectrogram X^{src}
  $z_T = \mathcal{E}_{AE}(X^{src})$
else
  sample $z_T \sim \mathcal{N}(0,1)$
end if
for $t = T, \cdots, 1$ **do**
  sample $\mathbf{z} \sim \mathcal{N}(0,1)$
  compute $\tilde{z} = \mathbf{m}_\theta(z_t, \vec{c}_{tgt}, t)$
  compute $z_{t-1} = \tilde{\mu}_t(z_t, \tilde{z}) + \sqrt{\tilde{\beta}_t}\mathbf{z}$ (Eq. 27)
end for
$X^{tgt} = \mathcal{D}_{AE}(z_0)$
Convert X^{tgt} to waveform W^{tgt}
Return: W^{tgt}

J.2. Training with pairs

As described in Section 3, we consider the paired data (X^{src}, X^{tgt}) with corresponding condition vector $\vec{c}_{src}$ and $\vec{c}_{tgt}$. Note that $\vec{c}_{src}$ is not in use.

Since X^{src} and X^{tgt} are the observations of same earthquake, we make assumption that there exist a morphology η which maps the latent x_t^{src} of X^{src} at time t, to x_t^{tgt} using $\vec{c}_{tgt}$, as a random variable. We formulate this assumption with Equation (1), as follows:

$$\eta(x_t^{src}, \vec{c}_{tgt}, t) \sim q(x_t^{tgt}|X^{tgt}) \tag{1}$$

This assumption includes the intuition that the broadband waveform signal is a combination of earthquake information, which is considered to be included in X^{src}, and local geological features near observatory, encoded by positional information from $\vec{c}_{tgt}$.

For training, we aim to train the neural network $\mathbf{m}_\theta$ which is a composition of η and denoising model $\mathbf{x}_\theta$. Precisely, $\mathbf{m}_\theta$ would be written by

$$\mathbf{m}_\theta(x, \vec{c}, t) = \mathbf{x}_\theta(\eta(x, \vec{c}, t), \vec{c}, t). \tag{28}$$

Since $\eta(x_t^{src}, \vec{c}_{tgt}, t) = x_t^{tgt}$, we have $\mathbf{m}_\theta(x_t^{src}, \vec{c}_{tgt}, t) = \mathbf{x}_\theta(x_t^{tgt}, \vec{c}_{tgt}, t)$ for the paired latents (x_t^{src}, x_t^{tgt}), the loss function of diffusion model Equation (2) would be equivalent to Equation (3):

$$\mathcal{L}'_{DM} = \mathbb{E}_{(X^{src}, X^{tgt}, \vec{c}_{tgt}), \epsilon, t} \|X^{tgt} - \mathbf{m}_\theta(x_t^{src}, \vec{c}_{tgt}, t)\|^2 \tag{29}$$

After that, we consider same procedure in latent space (the z_t^{tgt} for the clarification) with autoencoder consist of the encoder $\mathcal{E}_{AE}$ and decoder $\mathcal{D}_{AE}$, we obtain the loss function Equation (9), with end-to-end training.

$$\mathcal{L}_{ours} := \mathbb{E}_{(X^{src}, X^{tgt}, \vec{c}_{tgt}), \epsilon, t} \|X^{tgt} - \mathcal{D}_{AE}(\mathbf{m}_\theta(z_t^{src}, \vec{c}_{tgt}, t))\|^2 \tag{9}$$

In Algorithm 1, we present an training algorithm for the HEGGS training with $\mathcal{L}_{ours}$. The paired waveforms and corresponding condition vector of target waveform would be sampled from the dataset, and the gradient descent would update all modules $\mathbf{m}_\theta, \mathcal{E}_{AE}$ and $\mathcal{D}_{AE}$ together.

Remark J.1. For the training process of diffusion model with Equation (9), several details below are considered for the loss and model design.

1. During the training, the noise is designed to be added to the Z^{src} instead of Z^{tgt}. This would provide robustness against site-specific noise which already included in observation W^{src} and its latent vector Z^{src}.
2. When t is small, z_t^{src} would be almost same to Z^{src} (this is also because X^{src} itself is already noisy) and thus the model would learn the transformation η with more attention.
3. Regarding the intuition that z_t^{src} and z_t^{tgt} will be identified (in distribution) when t is sufficiently large, the training loss Equation (9) would be equivalent to the conventional training loss for $\mathbf{x}_\theta$ training when we disregard the end-to-end training. Hence, the model learns to generate from the noise w/o W^{src} too, during the training.
4. Since η and $\mathbf{m}_\theta$ does not take $\vec{c}_{src}$ as input. Therefore the model learns to extract common information from z_t^{src} through multiple pairs of observations of same earthquake during training, regardless the local information (encoded by location) of observatory. This makes the model can handle z_t^{tgt} as a input too, since it shares the information of earthquake.

J.3. Inference w/o W^{src}

Although the diffusion model is trained with paired data and takes W^{src} as an input, our model is capable to synthesize seismic waveform without the observation W^{src}.

Since η is defined to map the source latent z_t^{src} to target latent z_t^{tgt}, it also maps the target latent to itself, in distribution. Precisely, we can write

$$\eta(z_t^{tgt}, \vec{c}_{tgt}, t) = z_t^{tgt} \tag{30}$$

and thus the output of neural network would be

$$\mathbf{m}_\theta(z_t^{tgt}, \vec{c}_{tgt}, t) = \mathbf{x}_\theta(\eta(z_t^{tgt}, \vec{c}_{tgt}, t), \vec{c}_{tgt}, t) = \mathbf{x}_\theta(z_t^{tgt}, \vec{c}_{tgt}, t) \tag{31}$$

Therefore, we can use conventional reverse process

$$z_{t-1}^{tgt} = \tilde{\mu}_t(z_t^{tgt}, \mathbf{m}_\theta(z_t^{tgt}, \vec{c}_{tgt}, t)) + \sigma_t \mathbf{z}, \mathbf{z} \sim N(0, I) \tag{32}$$

even if z_T^{tgt} is the gaussian noise sampled from $\mathcal{N}(0, 1)$.

In Algorithm 2, we summarize the generation process of our model. Note that the diffusion steps are equivalent to LDM(Rombach et al., 2022) when W^{src} is not given.