Enhancing Generative Seismic Modeling via Paired Dataset Construction Method

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

Abstract

Observation in Earth Sciences encompasses not only what can be visually perceived but also what can be inferred through instrumental recordings. As such, seismic data, though not directly visible, fall within the domain of Earth Observation (EO). Earthquakes are inherently sparse events, and the limited availability of ground motion records and associated metadata poses significant challenges for predicting and responding to earthquake-induced hazards. Although numerous data augmentation techniques based on deep learning have been proposed, their effectiveness is often hindered by the scarcity of high-quality training data. We introduce a scalable framework for constructing training datasets from limited seismic observations, aimed at improving the performance of generative models. By training models on the paired dataset constructed using our proposed methodology, we demonstrate both quantitatively and qualitatively that the generated waveforms closely resemble real seismic signals, thereby validating the effectiveness of our approach.

1. Introduction

As a form of Earth Observation, seismic waveforms are not visually observed but are instead recorded through ground motions induced by seismic events. When an earthquake occurs, seismic waves are recorded independently at multiple observation stations, with each recording reflecting the local geological and geographical characteristics of the station's location.

Consequently, seismic waveforms recorded from the same earthquake event exhibit both unique local characteristics and shared underlying features, making them well-suited for the construction of paired datasets. Motivated by the dual nature of seismic waveforms, which exhibit both commonalities and unique characteristics, we introduce a novel methodology for constructing datasets by pairing seismic waveforms recorded from the same earthquake event. Such datasets not only facilitate effective data augmentation by introducing physically consistent variability, but also contribute to improved generative model generalization ability across diverse geographical and geological conditions.

We obtained two dataset groups, SCEDC(SCEDC, 2013) and INSTANCE(Michelini et al., 2021), via the SeisBench(Woollam et al., 2022) API to apply our dataset processing methodology. We also included an additional earthquake-related dataset to validate our approach. According to statistical analysis, although major earthquakes are not frequent, small seismic events occur often. However, the data required to cover the full range of events that require human prediction and response is still lacking. This underscores the need for data augmentation methods that reflect the distinct characteristics and distribution patterns of seismic waveforms, and supports the application of generative approaches that are well aligned with these data properties.

In this paper, we focus on two primary contributions:

Our Contribution

- We propose a method that utilizes only an essential set of conditions such as earthquake origin time, epicenter location, depth and magnitude to construct paired seismic waveforms by grouping those that share the same earthquake event. This approach leverages commonly available metadata from public earthquake catalogs ensuring both practicality and scalability.

- We demonstrate how generative modeling can serve as an effective data augmentation strategy by synthesizing realistic waveforms at new or underrepresented station locations. Leveraging the pairable nature of seismic data, our method enriches existing datasets without requiring detailed source or site-specific information.

[*]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 TerraBytes: Towards global datasets and models for Earth Observation Workshop at the 42^{st} International Conference on Machine Learning

Table 1. Features of each Regional paired seismic 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 M_L	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

2. Dataset Construction

Seismic datasets are well-structured and self-aligning. Each waveform is time-referenced to the earthquake origin time and accompanied by reliable metadata, such as earthquake origin time, magnitude, and hypocenter coordinates, Station coordinates and recorded waveform traces.

We curated a large-scale paired dataset from three continental regions : SCEDC (North America), KMA (East Asia), and INSTANCE (Europe). For each event, we extract 60-second waveforms aligned to origin time, apply a $1 \sim 45$ Hz bandpass filter, and associate each trace with essential metadata (coordinates, depth, magnitude). Paired samples are formed by randomly selecting two stations per event. For the EQTransformer(Mousavi et al., 2020) (EQT) evaluation in Table 4, we excluded waveforms which include multiple event signals, which are out of our scope.

In this section, we explain how each dataset was constructed. All datasets are collected from corresponding APIs. Plus, The Table 1 shows the details of datasets we used.

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.

2.1. SCEDC

We utilize earthquake catalog of SCEDC (SCEDC, 2013) provided by SeisBench(Woollam et al., 2022). We selected waveforms with a sampling rate of 100Hz. 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.

2.2. 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 consisted 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.

2.3. INSTANCE

We used the Seisbench-provided version INSTANCE dataset and created a subset by selecting only the traces satisfying which includes records for 60 seconds from the earthquake occurrence time, local magnitude is larger than 3.0 and P-arrival time is included in the metadata to ensure that the earthquake signal is observed.

3. Method

3.1. Pair-Exploiting Diffusion Model

For each earthquake event, we sample a pair of waveforms (W^{src}, W^{tgt}) from the dataset and convert them into spectrograms (X^{src}, X^{tgt}) and construct the conditional vector for the target station $\vec{c}_{tgt}$ by preprocessing recipe in Appendix A.1.

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:

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

Table 2. Results of quantitative analysis. Models were evaluated with W^{src} when it is trained with paired data, otherwise without W^{src}.

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
	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**
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
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**

We conducted train based on this assumption Equation (5) using the following Algorithm 1. More detailed description of the method is provided in the Appendix B.

Algorithm 1 Paired LDM 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
        $\nabla \| X^{tgt} - \mathbf{m}_\theta(x_t^{src}, \vec{c}_{tgt}, t) \|^2$
        where $\mathbf{m}_\theta(x_t^{src}, \vec{c}_{tgt}, t) = \mathbf{x}_\theta(\eta(x, \vec{c}, t), \vec{c}, t)$
until converged
```

3.2. Condition

In many Earth Observation (EO) applications, acquiring detailed contextual metadata such as local geology, instrument calibration, or expert annotations is often difficult, expensive, or infeasible, especially in under-observed regions. In contrast, seismic data inherently provide abundant quantitative metadata—such as event time, location, depth, and magnitude—which offers significant advantages in dataset construction compared to other Earth Observation (EO) modalities. Our methodology utilizes only the basic set of readily available metadata (**Event**: latitude, longitude, depth, magnitude **Station**: latitude, longitude). More details described in Appendix A

4. Empirical Verification

We evaluate the effectiveness of paired-data training for generative models in two synthesis scenarios: (1) Generate waveforms of known earthquakes at known stations using an observed waveform W^{src}, and (2) Generate waveforms from fictitious metadata $\vec{c}''_{tgt}$ without W^{src}. These experiments validate the fidelity, generalizability, and enhancement potential of our approach, key challenges in EO data generation.

4.1. Quantitative Evaluation

To assess model fidelity, we measure P-/S-phase arrival times and similarity metrics including envelope correlation, SNR, PSNR, and spectrogram MSE. We compare against baseline models (Seismogen(Wang et al., 2021), ConSeisgen(Li et al., 2024), and LDM(Rombach et al., 2022)) trained on the SCEDC dataset. Additionally, we validate generalization by training on KMA and INSTANCE datasets.

4.1.1. Phase Arrival Accuracy

Phase arrival times are fundamental seismic features used in earthquake detection, localization, and early warning systems. Thus, accurately reproducing P- and S-phase arrivals is a strong indicator of how well a model captures the essential characteristics of seismic events.

We use EQTransformer (EQT) (Mousavi et al., 2020), fine-tuned per dataset via SeisBench (Woollam et al., 2022), to extract P/S-wave labels from both real and generated waveforms. Results in Table 4 show high accuracy of the picker.

MAE scores in Table 2 show that our paired data LDM achieves significantly improved arrival prediction compared to baselines, highlighting its utility for seismic EO applications where accurate temporal alignment is essential.

4.1.2. Similarity Metrics

We further assess fidelity via envelope correlation, SNR, PSNR, and spectrogram similarity. As shown in Table 2, pairing improves the performance of LDM across all metrics.

This demonstrates that dataset structuring—specifically, pairing observations with minimal metadata—enhances waveform realism, and opens opportunities for spatio-temporal EO dataset expansion via synthetic generation.

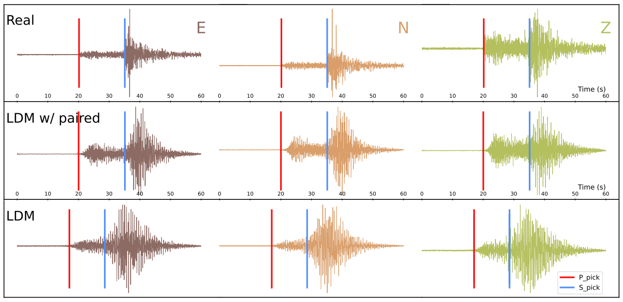


Figure 1. Comparison of Real waveform, LDM and LDM with paired data

4.2. Qualitative Evaluation

We qualitatively assess how well the paired-data trained model captures key seismological patterns, with a focus on fidelity, spatial consistency, and potential for dataset expansion in Earth Observation contexts.

4.2.1. Waveform and Spectrogram Analysis

We qualitatively assess the fidelity of generated waveforms by comparing them with real signals in both time and frequency domains. As shown in Figure 1, synthetic waveforms closely match real ones across all three components in terms of amplitude, duration, and phase arrival structure. The spectrograms of the generated and observed waveforms exhibit strong alignment, capturing key time-frequency characteristics such as energy distribution and seismic phase structure.

This alignment highlights the model's ability to learn and replicate event-specific morphology and spectral patterns crucial for downstream Earth Observation tasks like earthquake detection and ground motion modeling. Furthermore, spectrogram-level fidelity supports potential applications in vision-based EO dataset curation and frequency-aware augmentation strategies. More qualitative results as shown in Appendix D

4.2.2. Synthetic Station Analysis

To evaluate spatial generalization, we synthesize waveforms at virtual (unseen) station locations. Figure 2 illustrates how seismic energy propagates across these synthetic stations. The realistic spatial and temporal variations confirm the model's ability to generate coherent seismic fields, enabling synthetic data generation in under-monitored or uninstrumented regions a key step toward mitigating spatial bias in EO datasets.

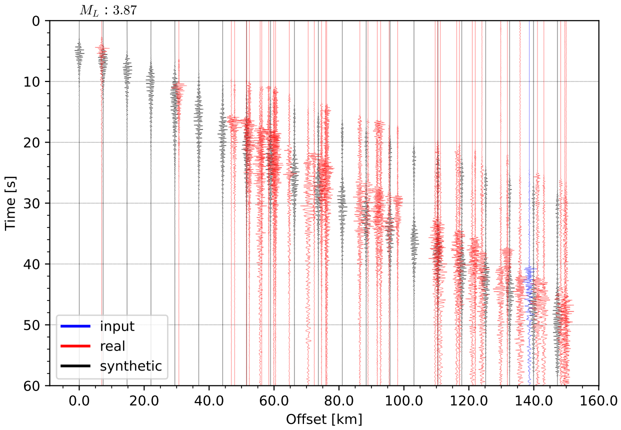


Figure 2. Section plots comparing synthetic and real waveforms.

5. Discussion

We show that models trained on our proposed paired dataset yield superior results both quantitatively and qualitatively. These findings suggest that the paired-dataset approach can play a critical role in constructing effective generative AI models for seismic applications.

While some discrepancies from the original signals were inevitable due to the exclusion of certain geophysical variables that were not available during the pairing process, the models trained on the paired dataset, constructed using commonly available geographic metadata, still outperformed those trained without pairing. These results indirectly confirm the effectiveness of the paired dataset, even in the absence of detailed site-specific labeling.

6. Conclusion

We present a dataset pairing methodology that simultaneously increases data volume and enhances informational richness for training generative models. This strategy leverages the structural characteristics of seismic waveforms, wherein recordings from a single seismic event share consistent metadata—such as origin time, epicenter, depth, and magnitude—while capturing station-specific geological features. Empirical results demonstrate performance improvements, suggesting the proposed approach as a viable framework for extending generative modeling to other data-scarce Earth Observation domains.

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

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.

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.

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.

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.

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.

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., 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.

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.

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

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, 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.

A. Conditioning

The six variables mentioned above were available for all data samples; however, several additional variables were frequently missing for certain seismic waveforms, resulting in their exclusion from the pairing process. Additionally, we derived and incorporated explicit metadata calculated from the given six variables such as, epicentral distance and back-azimuth to further enrich the dataset.

By demonstrating high-fidelity synthesis under available conditions, we show that realistic and spatially diverse EO data generation is feasible even in settings where entire rich metadata is unavailable, thereby expanding the applicability of generative EO methods to globally underrepresented regions.

A.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:

$$s'_{lat} = \frac{s_{lat} - l_{lat}}{u_{lat} - l_{lat}}, e'_{lat} = \frac{e_{lat} - l_{lat}}{u_{lat} - l_{lat}}, s'_{lon} = \frac{s_{lon} - l_{lon}}{u_{lon} - l_{lon}} \text{ and } e'_{lon} = \frac{e_{lon} - l_{lon}}{u_{lon} - l_{lon}} \tag{2}$$

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 3.

Table 3. 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:

$$\begin{aligned} c_{sta} &= (cos(s'_{lat})cos(s'_{lon}), sin(s'_{lat})cos(s'_{lon}), sin(s'_{lon})) \\ c_{epi} &= (cos(e'_{lat})cos(e'_{lon}), sin(e'_{lat})cos(e'_{lon}), sin(e'_{lon})) \end{aligned} \tag{3}$$

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

$$c_{azi} = (cos(Azi), sin(Azi)) \tag{4}$$

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

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.

	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	$(M_L - 2.0)/6.4$	$(M_L - 0.35)/5.24$	$(M_L - 3.0)/6.5$

A.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 $3 \times 64 \times 376$.

B. 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 to train both the diffusion process for spectrogram generation and the high-fidelity decoder for waveform generation.

B.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:

$$\eta(x_t^{src}, \vec{c}_{tgt}, t) \sim q(x_t^{tgt}|X^{tgt}). \tag{5}$$

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

$$\mathcal{L}_{DM} = \mathbb{E}_{(X^{tgt}, \vec{c}_{tgt}), \epsilon, t} \|X^{tgt} - \mathbf{x}_\theta(x_t^{tgt}, \vec{c}_{tgt}, t)\|^2. \tag{6}$$

while the SNR weight is simplified.

Considering the Equation (5), we rewrite the loss function as

$$\mathcal{L}'_{DM} = \mathbb{E}_{(X^{src}, X^{tgt}, \vec{c}_{tgt}), \epsilon, t} \|X^{tgt} - \mathbf{m}_\theta(x_t^{src}, \vec{c}_{tgt}, t)\|^2 \tag{7}$$

where

$$\mathbf{m}_\theta(x, \vec{c}, t) = \mathbf{x}_\theta(\eta(x, \vec{c}, t), \vec{c}, t). \tag{8}$$

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

$$x_{t-1}^{tgt} = \tilde{\mu}_t(x_t^{tgt}, \mathbf{m}_\theta(x_t^{tgt}, \vec{c}_{tgt}, t)) + \sigma_t \mathbf{z}, \mathbf{z} \sim N(0, I) \tag{9}$$

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

$$\eta(x_t^{tgt}, \vec{c}_{tgt}, t) \overset{d}{=} x_t^{tgt} \tag{10}$$

by assumption and thus

$$\mathbf{m}_\theta(x_t^{tgt}, \vec{c}_{tgt}, t) = \mathbf{x}_\theta(x_t^{tgt}, \vec{c}_{tgt}, t). \tag{11}$$

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 $\mathbf{m}_\theta$.

B.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

$$\mathcal{L}'_{LDM} = \mathbb{E}_{(Z^{src}, Z^{tgt}, \vec{c}_{tgt}), \epsilon, t} \| Z^{tgt} - \mathbf{m}_\theta(z_t^{src}, \vec{c}_{tgt}, t) \|^2 \tag{12}$$

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 (12) into an end-to-end loss function as shown below:

$$\begin{aligned} &\mathcal{L}_{ours} \\ &:= \mathbb{E}_{(X^{src}, X^{tgt}, \vec{c}_{tgt}), \epsilon, t} \| X^{tgt} - \mathcal{D}_{AE}(\mathbf{m}_\theta(z_t^{src}, \vec{c}_{tgt}, t)) \|^2 \end{aligned} \tag{13}$$

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

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. For the detailed implementation in the diffusion module, we used a U-Net backbone for $\mathbf{m}_\theta$, brought $\mathcal{E}_{AE}$ and $\mathcal{D}_{AE}$.

C. EQT

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}.

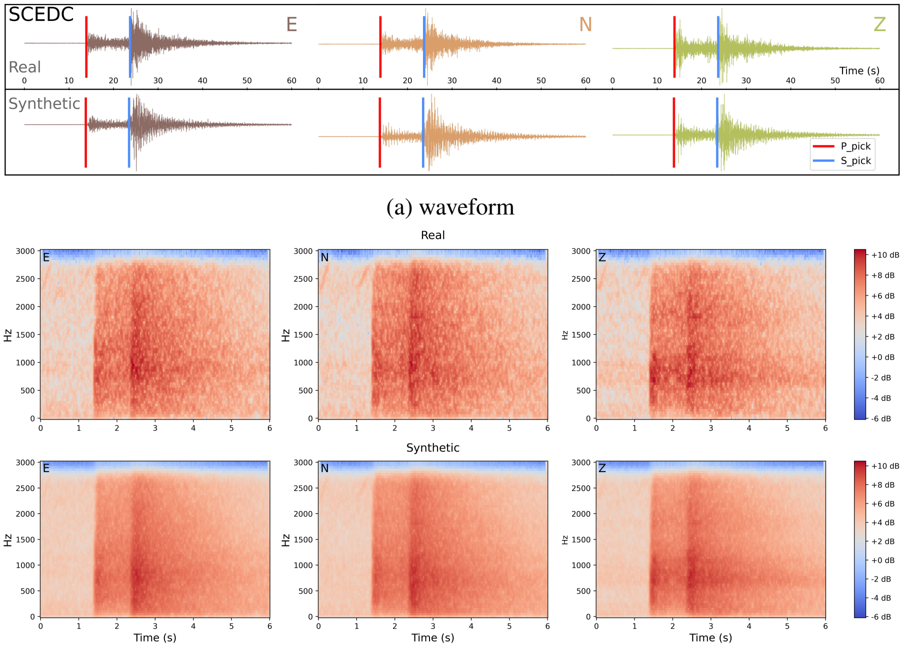
Table 4. Performance of EQT picker. F1: errors $<$ 0.5s counted as positive.

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

D. Additional Figures: Waveform and Spectrogram

This section presents the waveforms and spectrograms shown in Figure 1. 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.

D.1. SCEDC

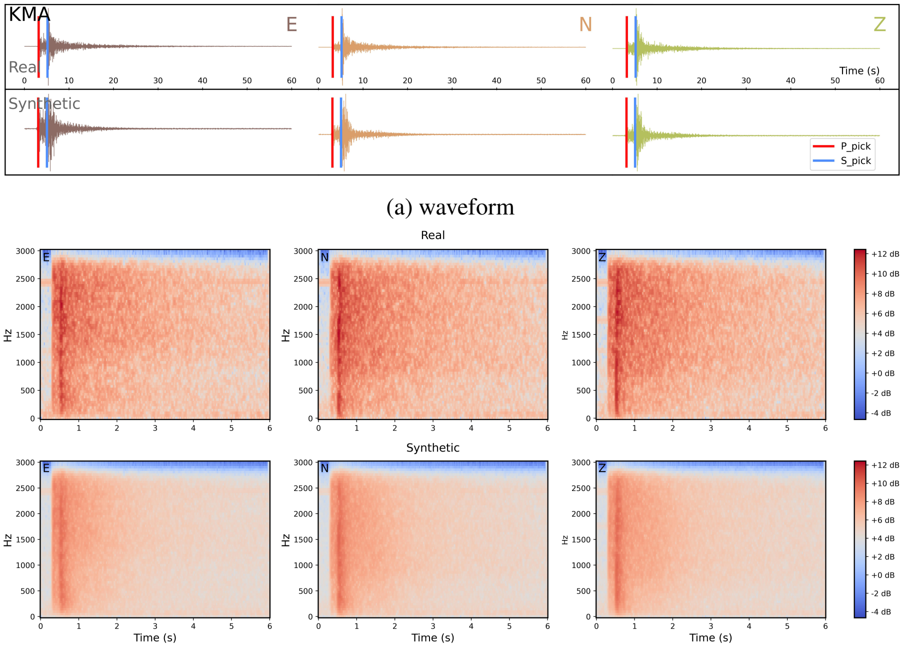


(a) waveform

(b) spectrogram

Figure 3. Synthesis results of our model compared to the real observation.

D.2. KMA

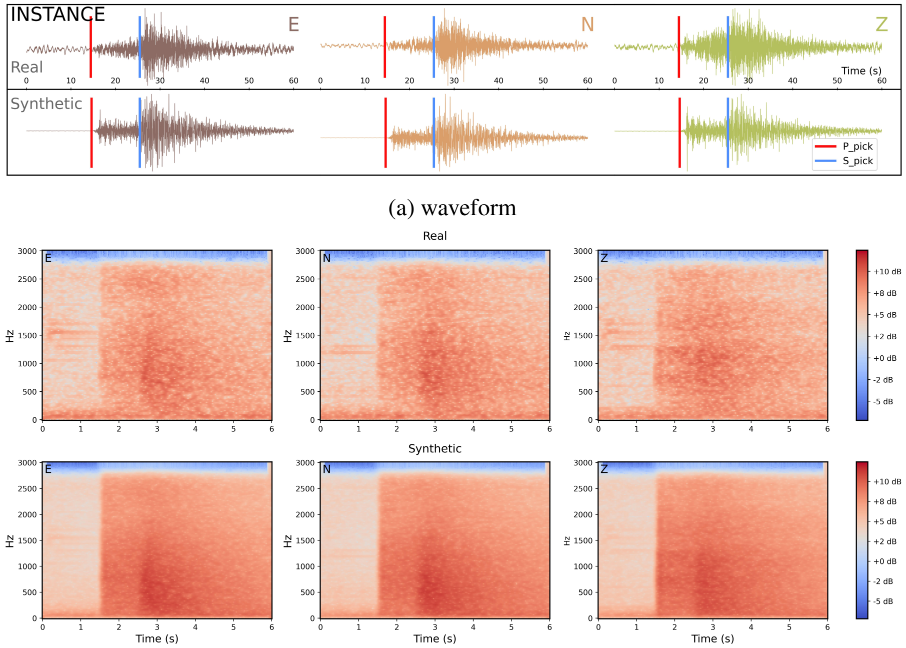


(a) waveform

(b) spectrogram

Figure 4. Synthesis results of our model compared to the real observation.

D.3. INSTANCE



(a) waveform

(b) spectrogram

Figure 5. Synthesis results of our model compared to the real observation.

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 (14) 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 (14) from pix2pix(Isola et al., 2017) as an additional loss term to improve training stability.

$$L_{\mathrm{L1}}(G) = \mathbb{E}_{x,y,z}\left[\|x_{tgt} - G(z,y)\|_1\right] \tag{14}$$

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:

$$L_{total} = min_{\mathcal{E}_{AE},\mathcal{D}_{AE}}, max_{\psi}[L_{rec}(x, \mathcal{D}_{AE}((x))) - L_{adv}(\mathcal{D}_{AE}(\mathcal{E}_{AE}(x))) + logD_{\psi}(x) + \lambda_{kl}KL] \tag{15}$$

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 5.

Table 5. 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 (16): not utilizing the characteristic of paired data
- Equation (17): 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.

$$L'_{LDM} = \mathbb{E}_{(Z^{tgt}, \vec{c}_{tgt}), \epsilon, t} \| Z^{tgt} - \mathbf{x}_\theta(z_t^{tgt}, \vec{c}_{tgt}, t) \|^2 \tag{16}$$

$$L'_{LDM} = \mathbb{E}_{(Z^{src}, Z^{tgt}, \vec{c}_{tgt}), \epsilon, t} \| Z^{tgt} - \mathbf{m}_\theta(z_t^{src}, \vec{c}_{tgt}, t) \|^2 \tag{17}$$