SEISMIC NOISE MODELING WITH GENERATIVE AI

Sayan Kumar Swar *
Department of Electrical and Computer Engineering
University of Rochester
Rochester, NY 14620, USA
sswar@ur.rochester.edu

ABSTRACT

The receiver function (RF) technique is a way to study the structure of the Earth's crust and upper mantle by using data from teleseismic earthquakes. When seismic signals are recorded they are always convolved with inter crustal reverberations and noise. These makes the singal anaysis harder and in order to analyze Seismologist needs to deconvolve those noises to generate accurate receiver functions which can then be used to analyse velocity structure beneath the receiver. Though a variety of deconvolution techniques have been developed, they are all adversely affected by background and signal-generated noise. In order to take into account the unknown noise characteristics Seismologist has previously used Bayesian inference in which both the noise magnitude and noise spectral character are parameters in calculating the receiver functions. In this project I propose to build a generative deep learning diffusion model which can better estimate the noise information and thus in turn can imporve a more accurate receiver function realization.

1 INTRODUCTION

The Receiver Function (RF) technique is one of the most widely employed methods for imaging crustal and upper mantle structures using teleseismic earthquake data. This technique provides insights into the lateral variability of significant velocity boundaries, such as the Moho and other major discontinuities within the Earth's crust. The method primarily utilizes data from teleseismic Primary (P) and Secondary (S) waves. When a teleseismic P-wave encounters a boundary with a strong velocity contrast, it generates converted S-waves. The travel-time difference between the original P-wave and the converted S-wave carries critical information about the depth and velocity structure of the crust and upper mantle beneath the recording station. Receiver functions are derived by deconvolving the "parent" wave (initial P-wave) from the "daughter" wave (converted S-wave) to isolate the converted component and remove the effect of source and propagation path. These converted waves, recorded at the surface by seismometers, are subsequently analyzed to infer Earth structure beneath the receiver.

One of the primary challenge in RF analysis is to obtain receiver functions that are minimally affected by background noise and signal-generated noise. This noise often cannot be effectively modeled through a simple convolution operator, as the energy is scattered by heterogeneous structures that vary by path. Consequently, methods that are robust to noise are essential for enabling receiver function analysis at seismic stations in noisy environments. Previous approaches have applied Bayesian inference, where both the noise amplitude and spectral characteristics are treated as parameters for generating an ensemble of receiver functions. Researchers have found that seismic noise when represented as a function of lag time, follows a certain pattern which can be captured approximately through a parametric model. However, noise parameters in such models are often determined through empirical equations, which may not adequately capture the specific noise characteristics at a given location. Accurate noise parameterization is essential, as improved noise modeling directly enhances the resolution of extracted receiver functions.

*Use footnote for providing further information about author (webpage, alternative address)—*not* for acknowledging funding agencies. Funding acknowledgements go at the end of the paper.

In this project, I propose an alternative approach that employs a deep learning generative model, specifically a diffusion-based model, to learn the complex patterns of pre-seismic event noise at specific stations. Rather than relying on empirical equations, this model will be trained on seismic noise information to recognize noise characteristics, accurately estimating noise parameters that can then be used to clean noisy seismic signals. The deep learning model is expected to provide a more robust and station-specific representation of noise, ultimately facilitating the extraction of higher-resolution, noise-free receiver functions for detailed subsurface imaging.

2 RELATED WORK

To generate noise-free seismic signals, which are critical for producing accurate receiver functions, researchers have focused on modeling the behavior of noise as a function of lag time. Given that seismic signals are time-series data, constructing a covariance matrix C_D over the temporal axis is essential for accurately capturing the covariance characteristics of the noise. In this study, we adopt a direct parametrization of C_D that approximates the characteristics observed in pre-event noise recorded by broadband seismometers across various settings. Bodin et al. (2012) propose two parametrizations for the covariance matrix: type 1, an exponentially decaying correlation (Equation 7), and type 2, a Gaussian correlation (Equation 8). Their findings indicate that the Gaussian (type 2) parametrization is more realistic, as an exponentially decaying correlation diminishes too rapidly, resulting in a noise profile with higher frequency content than is typically observed in actual pre-event noise.

To assess the suitability of these noise parametrizations for representing real-world noise, other researchers have analyzed pre-event noise at station ANMO. Covariance matrices were constructed from 250-second samples of noise preceding recorded seismic events to represent the behviour of noise as function of lag time. This involved normalizing the noise time series, arranging them in a matrix, and computing the matrix product with its transpose. Each row of the resulting matrix was then normalized by dividing by the corresponding diagonal entry. The averaged values along the diagonal provided an estimate of the correlation as a function of lag time, illustrated in Figure. The analysis has revealed that the noise exhibits a decaying sinusoidal correlation with periodic anticorrelations, which cannot be adequately captured by the type 1 and type 2 parametrizations. In light of these observations a new noise parametrization of type 3 (eq. 1) was intrduced by Kolb & Lekić (2014) which models the covariance as a decaying exponential multiplied by a cosine function.

$$R_{ij} = e^{-\lambda|t_j - t_i|} cos(\lambda\omega_0|t_j - t_i|) \tag{1}$$

To limit the parameter count in this third parametrization, they fixed the value of ω, thereby anchoring the decay rate to the oscillation frequency of the covariance. For each noise model, they determined the parameters that best matched the actual noise, and subsequently plotted the modeled and observed correlations. The results demonstrated that the type 3 parametrization in Figure 1(c) right, aligns more closely with observed noise characteristics over a broader range of lag times, particularly at extended lag times.

In this study, the focus is on Noise Parametrization type 3, with the hypothesis that a deep learning-based generative model can outperform the conventional approach outlined in Equation 1 in estimating noise parameters. As Equation 1 requires the parameters λ and ω to be scalar values, it is hypothesized that a deep learning model trained on noise data can provide dynamically parametrized representations, resulting in a more accurate and comprehensive characterization of noise behavior.

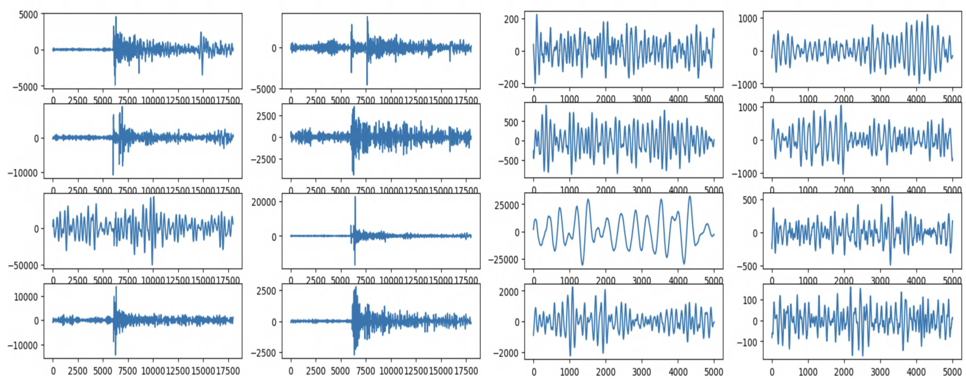
3 DATA SET

Seismic data used for modeling noise characteristics are sourced from the Data Management Center of the Incorporated Research Institutions for Seismology (IRIS), specifically under network code IU and station ANMO. This particular seismic station was selected due to the availability of published studies that provide noise parameter benchmarks, enabling an effective comparison of our neural network model's performance against traditional mathematical models. The dataset comprises approximately 2.8 GB of data, encompassing around 4200 Z-seismic records.

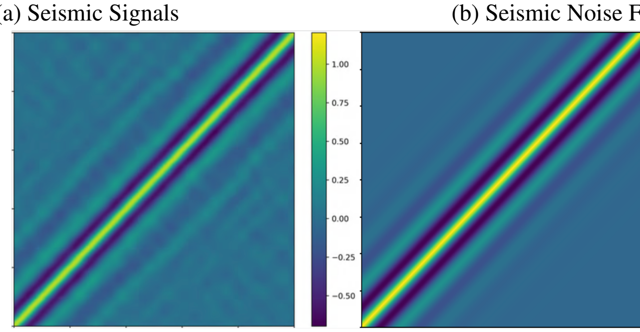
For the initial phase of this project, the neural network models will be trained and evaluated on a single channel of data to assess preliminary performance. Upon obtaining promising results, the analysis will be expanded to incorporate three-channel data, allowing for a more comprehensive and accurate modeling of noise characteristics. This progressive approach ensures that the neural network model is optimized before engaging in the more complex, multi-dimensional noise characterization required for robust receiver function extraction.

4 METHOD

To replicate their findings, I obtained the ANMO dataset and conducted the same experiments to synthetically generate noise, employing a curve-fitting approach to estimate the parameters λ and ω to align with the results reported in Kolb & Lekić (2014). My estimates yielded $\lambda = 0.2$ and $\omega_0 = 4.26$ closely matching the paper's reported values of $\lambda = 0.2$ and $\omega_0 = 4.2$. Figure 1(a) presents selected examples of seismic signals, while Figure 1(b) highlights the first 250 seconds if the entire signal that constitutes the pre-arrival seismic noise.



(a) Seismic Signals (b) Seismic Noise Focus



(c) Left: Pre-arrival noise Original, Right: Parameterized Noise Type 3

Figure 1: (a) shows some randomly selected seismic signal from our dataset. (b) focuses on the pre-arrival seismic noise of the same signals. (c) the left figure depicts the actual noise as a function of lag time through covariance matrix and the right figure depicts the parameterized noise type 3

The main focus point of this study is the covariance matrix in Figure 1(c, left) that shows the characteristics of pre-arrival noise as a function of lag time. The focus will be to train a generative deep learning model so that it can reproduce seismic noise which will follow original characteristics more closely compared to the one in Figure 1(c, right).

Given the state of the art performance of Diffusion Models as generative frameworks, this study employs them as the primary methodology. Diffusion Models operate by reversing a stochastic process in which clean data is incrementally corrupted with increasing levels of random noise Permenter & Yuan (2023). After training, the model learns to generate new samples by progressively de-noising a randomly sampled Gaussian distribution over T iterative steps, thereby reproducing outputs that closely adhere to the original data distribution. Although Diffusion models, initially popularized for image and video generation it has also demonstrated remarkable performance in non-visual domains, such as speech signal processing Kong et al. (2020). While differing in nature, seismic data, being another form of time-series signals, are expected to benefit from the application of diffusion models, potentially yielding accurate and robust results. The analysis of seismic data involves processing extended waveforms, resulting in extensive datasets that present significant computational challenges. Diffusion models, being resource-intensive, require substantial computational memory, further exacerbating the demand for high-performance computational resources thus it is wise to always keep the efficiency of an algorithm in mind while processing with such large scale datasets. Researchers from the Toyota Research Institute Permenter & Yuan (2023) have advanced diffusion model denoising techniques by formulating them as a projection problem under the manifol hypothesis, introducing an optimized convergence strategy based on the projection error of the denoiser. Additionally, they proposed a novel gradient-estimation sampler, which accelerates the diffusion sampling process and achieves better Fréchet Inception Distance (FID) scores compared to other models. The faster convergence and computationally efficient design of this model make it well-suited for seismic noise modeling, enabling effective training through architectural optimization and hyperparameter fine-tuning.

As diffusion model during the training learns to predict the diffused noise at each time-step thus the loss function $L(\theta)$ as shown in eq. 2 tries to minimize the $L2\ norm$ between added noise $\sigma_t \epsilon$ and predicted noise ϵ_θ for a particular time step t during each iteration. The noise intensity level at each time step t is determined by a defined schedule σ_t and ϵ is randomly sampled from a distribution $\mathcal{N}(0, I)$.

$$L(\theta) := \mathbb{E}||\epsilon_\theta(x_0 + \sigma_t \epsilon, \sigma_t) - \epsilon)||^2 \tag{2}$$

With the defined noise schedule and grounded in the fundamental principles of the manifold hypothesis, the sampling technique Permenter & Yuan (2023) is optimized by iteratively performing gradient steps, enabling the generation of the original data x_0 more efficiently, with fewer iterations, and improved approximation accuracy. The sampler enhances gradient estimation by incorporating aggregated information $\bar{\epsilon}_t$ from multiple prior denoising steps, as expressed in Eq. 3. This aggregation significantly reduces gradient estimation errors, thereby improving the stability and overall quality of the sampling process.

$$\bar{\epsilon}_t = \gamma \epsilon_\theta(x_t, \sigma_t) + (1 - \gamma)\epsilon_\theta(x_{t+1}, \sigma_{t+1}) \tag{3}$$

Furthermore, the method Permenter & Yuan (2023) postulates that incorporating noise during the generation process enhances sampling quality. To achieve this, while adhering to the original noise schedule, denoising is performed to a reduced $\sigma_{t'}$ followed by reintroducing noise $w_t = \mathcal{N}(0, I)$. The parameter η is carefully selected to ensure that the norm of the update remains constant in expectation, thus maintaining consistency in the sampling process and further refining the quality of the generated data. The Eq. 4. thus finally generates the sampled denoised data at a a timestep $t - 1$.

$$x_{t-1} = x_t + (\sigma_{t-1} - \sigma_{t'})\bar{\epsilon}_t + \eta w_t \tag{4}$$

4.1 Experiment 1: Sanity Testing with Synthetic Waveform

To evaluate the applicability of the selected diffusion model for waveform data, a synthetic dataset was generated for initial testing. The dataset consisted of a sinusoidal waveform with 200 samples at a frequency of 12 Hz.

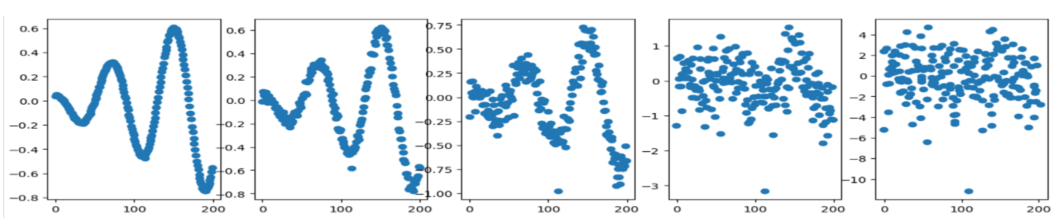
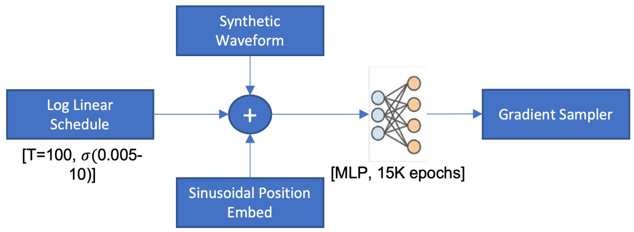
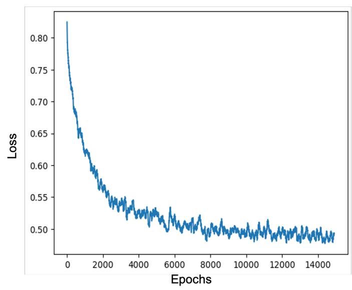


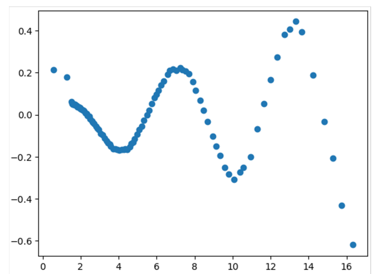
Figure 2: Noise addition at Individual Timesteps

The signal was preprocessed into a 2D matrix by incorporating its corresponding time-step details and sampling frequency. The forward diffusion process employed a log-linear noise schedule with 100 time steps and a noise intensity σ ranging from 0.005 to 10 as shown in Fig 2. To enhance model performance, sinusoidal positional embeddings were concatenated with the noise levels at each step of the forward process. The resulting data was input to a multi layer perceptron (MLP) with hidden dimensions structured as (16, 128, 128, 128, 128, 16). Finally the gradient sampler will be used to generate waveforms.



(a)





(b)

Figure 3: (a) is the architecture adopted for the sanity testing, key difference is MLP for noise prediction. (b) left shows the training loss progression and (b) right shows the resulting data after sampling from the learned model

After training the model for 15,000 epochs, the proposed sampler, which incorporates gradient estimation, was utilized for waveform generation. The model successfully reproduced the synthetic waveforms (Fig. 3(b, right), confirming its potential to handle waveform data effectively. The model was subsequently trained on a target dataset comprising approximately 4,000 seismic waveforms. However, the performance on this dataset was suboptimal, indicating the need for further optimization and adjustments to the model architecture or training procedure to improve its ability to generalize to real-world seismic data.

4.2 Refined Model Architecture

The model architecture was thus upgraded where the core of the model consists of a U-Net architecture with multiple downsampling and upsampling blocks. The downsampling path uses convolutional layers with normalization and activation functions to encode features progressively into smaller spatial resolutions, supported by residual blocks to maintain information flow. It is important to note that a log linear schedule is used in this technique which controls the noise intensity level added during each training step. The noise intensity is incorporated with a sinusoidal positional embedding which is added to the information flow in the residual blocks. Skip connections are established between the downsampling and upsampling paths and also during each residual block to retain spatial details and improve gradient flow. Attention mechanisms, integrated in the bridge connection enhance the model's ability to focus on relevant features during generation. The upsampling path mirrors the downsampling structure, reconstructing the input resolution while incorporating skip connections to refine the output.

The output layer applies normalization, activation, and a final convolution to produce the denoised or generated waveform. This design emphasizes efficiency in handling noise schedules and leverages

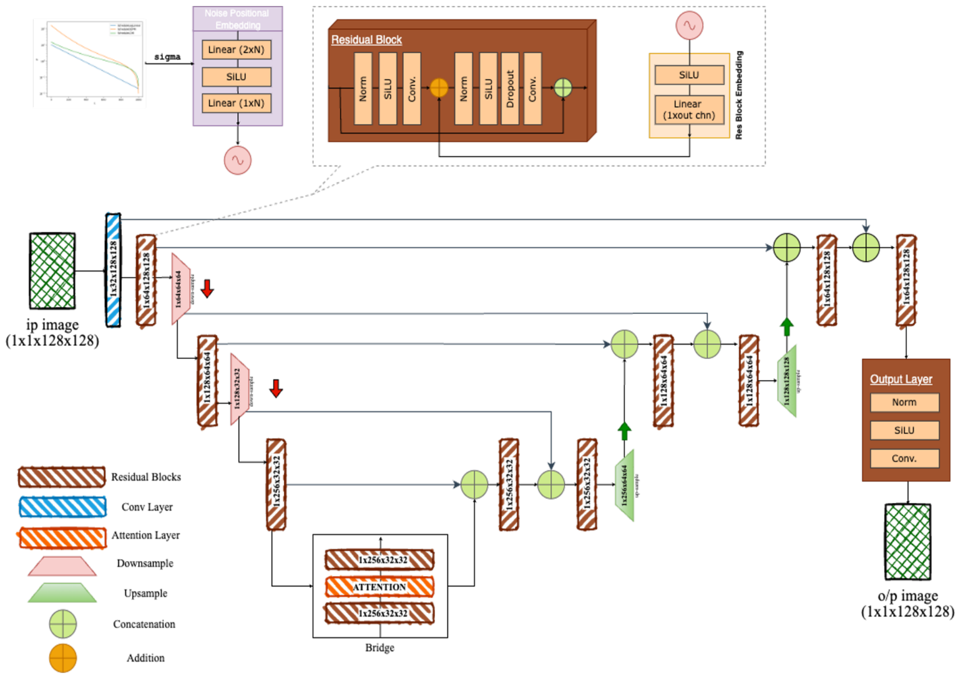
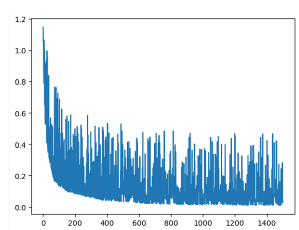


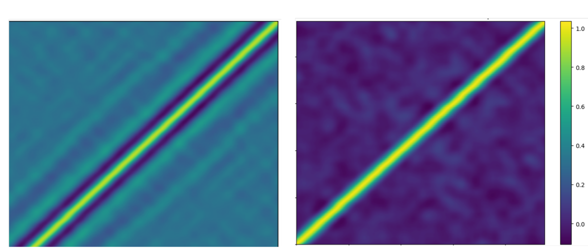
Figure 4: Depicts the Model Architecture used. Although the figure is highlighted with an image input, the focus here is to highlight the principal design of the diffusion model with the core components and workflow

gradient refinement strategies to optimize waveform generation. Overall, the model integrates foundational principles like residual learning, attention mechanisms, and hierarchical feature extraction for robust signal reconstruction.

4.3 Experiment 2: Learning to Generate Waveforms



(a)



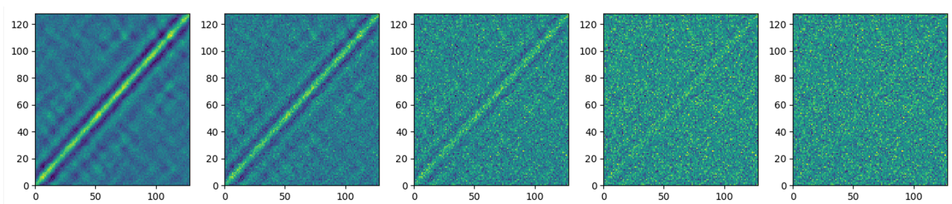
(b)

Figure 5: (a) Loss progression while training the seismic noise waveforms (4200×5000) with each row representing a unique seismic noise waveform. (b) left is the original noise covariance matrix and (b) right is the covariance matrix of the generated waveforms from the diffusion model

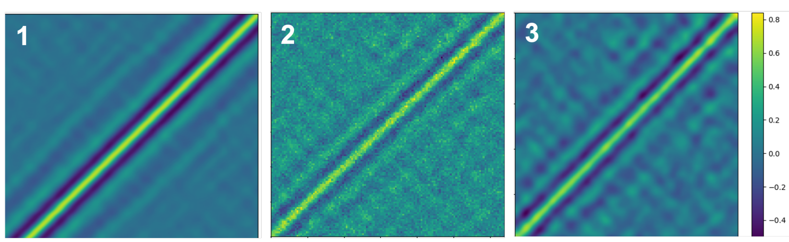
The complete seismic noise dataset was formatted into a tensor of dimensions 4200×5000, where each row represented a seismic noise waveform consisting of 5000 samples. The model was trained over 15,000 epochs using a full-batch training approach. However, the training process proved to be suboptimal, as reflected in the unstable loss progression curve in Fig. 5(a). Attempts to mitigate this instability by reducing the batch size were unsuccessful, as they not only failed to improve training performance but also significantly increased the overall training duration. As shown in Figure 5(b), a notable discrepancy was observed between the covariance matrix of the generated waveforms and that of the original waveforms, indicating a mismatch in statistical properties. Furthermore, the waveforms generated by the diffusion model failed to exhibit the expected lag-time characteristics inherent to the original seismic noise data, further emphasizing the limitations of the model's performance in accurately capturing the desired noise behavior.

4.4 Experiment 3: Learning to Generate Covariance Matrix

In the final experiment, the covariance matrix was directly utilized as input to the diffusion model. Seismic noise waveforms were processed to compute the covariance matrix as outlined in the Related Works section. Initially, the covariance matrix had dimensions of 1000×1000, corresponding to the first 1000 samples of the waveforms. This matrix was subsequently downsampled to a size of 128×128, reshaped into a single-channel image with dimensions 1×128×128. The model was trained using a single-batch approach over 15,000 epochs, with 100 noise addition timesteps scheduled with a log-linear scale ranging from 0.1 to 10 shown in Fig. 6(a).



(a)



(b)

Figure 6: (a) Loss progression while training the seismic noise waveforms (4200×5000) with each row representing a unique seismic noise waveform. (b) left is the original noise covariance matrix and (b) right is the covariance matrix of the generated waveforms from the diffusion model

This approach yielded significantly improved results, as observed in the sampled output Fig. 6 (b-2). The generated covariance matrix was then upsampled back to its original dimension of 1000×1000 using a Gaussian blurring technique as depicted in Fig. 6 (b-3). A visual inspection of the output demonstrated that the diffusion model effectively captured the noise characteristics, as reflected in the lag-time behavior, and produced results that outperformed the empirically defined Type-3 noise parameterization equation. These findings underscore the capability of the diffusion model to accurately learn and replicate the statistical properties of seismic noise.

4.5 Hyperparameter Optimization with WanDB

To fine-tune the model and enhance its performance, the Weights and Biases (WanDB) Sweep methodology was employed. The model's tunable hyperparameters included the number of convolutional layers, the number of residual blocks during downsampling and upsampling, the learning rate, the number of filters in the initial convolutional layer, the dropout rate, and the resolution of the noise positional embeddings. For this study, the primary hyperparameters selected for optimization were the learning rate (lr_rate), the number of filters in convolutional layers (ch_res), and the number of residual blocks (num_res_block). A configuration file specifying the range of hyperparameter values to be explored was prepared, and the WanDB Sweep tool was integrated to consume this configuration and execute the model with various combinations of hyperparameters over 10 runs. The optimization objective was defined as minimizing the training loss. The WanDB Sweep algorithm adaptively selected combinations of hyperparameters based on the results of previous runs, progressively steering the training process toward the optimal configuration that reduced the loss. The results are depicted in Fig. 7, which highlights the optimal hyperparameter combination yielding the minimum loss. Additionally, the importance of various hyperparameters is analyzed, revealing that lr_rate and num_res_block played pivotal roles in improving the model's performance, thereby driving it toward achieving optimal results.

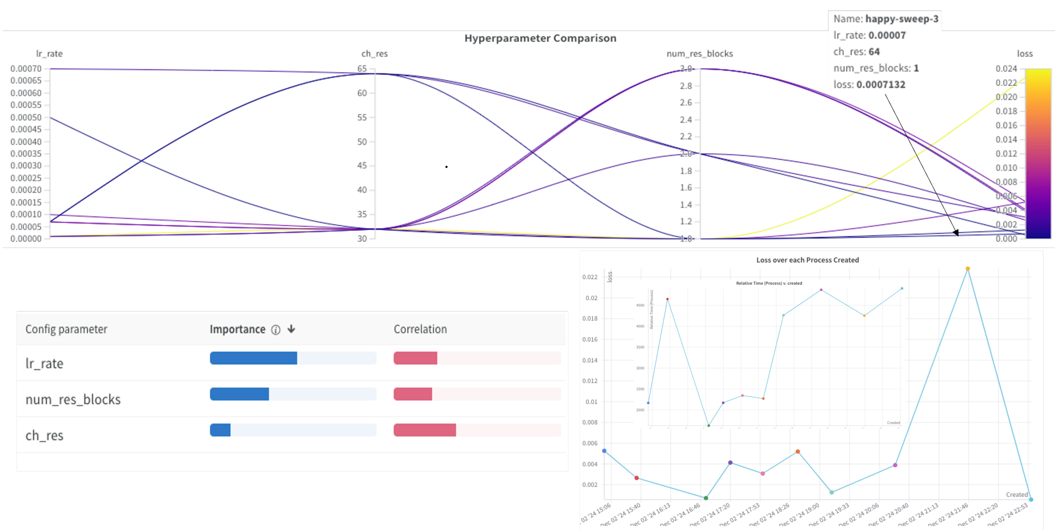


Figure 7: A snapshot of the Wandb Optimization report showing the optimal combination of hyperparameters, individual hyperparameter importance and the change in loss through each successive tests. The full report of the WanDb optimization can be found at: Seismic Noise Modelling Full Report of Hyperparameter Optimization in WanDB

5 Results

The model was finally trained using the most optimal hyperparameters determined through the optimization process, with the loss progression shown in Fig. 8(1).

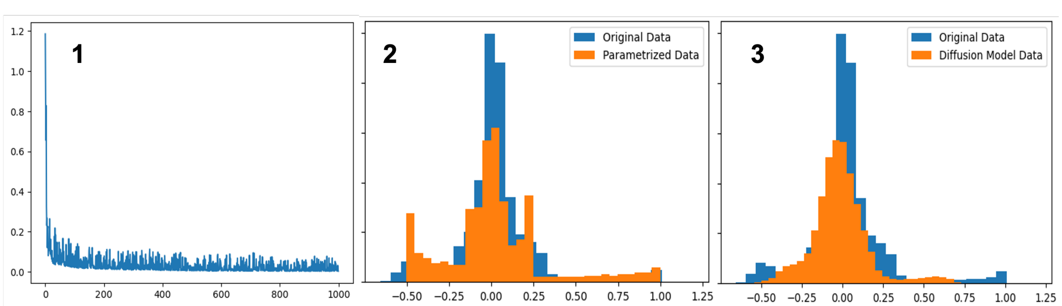


Figure 8: Loss Curve Progression in 1 and comparison between the data distribution of original data, type 3 parameterized data and diffusion model data in 2 and 3 respectively

Table 1: Quantitative Comparison of Results

Method	**KL-Divergence Loss**
Original Vs Diffusion	0.00105
Original Vs Param Type 3	0.01596

The output of the diffusion model was compared against the original seismic noise covariance dataset to evaluate its performance. Fig. 8(3) visually demonstrates that the diffusion model's output aligns more closely with the original data distribution compared to the Noise Parameterization Type 3 model, as depicted in Fig. 8(2).

To quantitatively assess this improvement, a comparison was performed using Kullback-Leibler (KL) Divergence, which measures the difference between two probability distributions. The KL divergence between the diffusion model's output and the original data distribution was found to be approximately 0.001, which is ten times smaller than the KL divergence between the Noise Parameterization Type 3 model and the original data distribution, measured at 0.01. This significant reduction in KL divergence highlights the superior capability of the diffusion model in learning and reproducing the noise characteristics, offering a far more accurate representation of the original seismic noise distribution compared to conventional parameterization methods.

6 Conclusion

In conclusion, this study explored the limitations of traditional noise parameterization techniques, particularly Noise Parameter Type 3, in accurately modeling seismic noise characteristics and proposed the use of a diffusion model to address these challenges. The experiments demonstrated that the diffusion model leveraging its generative capabilities, could effectively learn the underlying noise covariance structure in relation to lag time with reduced reliance on empirical assumptions. The proposed model exhibited significantly better performance, as reflected by a tenfold reduction in KL divergence compared to Noise Parameter Type 3, indicating a much closer approximation to the original seismic noise distribution. By incorporating modern optimization techniques like WandB sweeps, the hyperparameters of the model were fine-tuned, further improving its training stability and accuracy. Visual inspections of the covariance matrix and quantitative evaluations consistently highlighted the model's ability to capture the noise characteristics across lag times more effectively than traditional methods. This work establishes the potential of diffusion models as a robust and scalable alternative for seismic noise modeling, paving the way for more accurate receiver function estimations and seismic imaging in future research.

7 Acknowledgment

I sincerely acknowledge Prof. Chenliang Xu for the CSC 445 ”Deep Learning” Course and his advise for conducting this project. I also deeply acknowledge my PhD advisor Dr. Tolulope Olugboji for his constant support and motivation.

REFERENCES

Thomas Bodin, Malcolm Sambridge, Hrvoje Tkalčić, Pierre Arroucau, Kerry Gallagher, and Nicholas Rawlinson. Transdimensional inversion of receiver functions and surface wave dispersion. *Journal of geophysical research: solid earth*, 117(B2), 2012.

JM Kolb and V Lekić. Receiver function deconvolution using transdimensional hierarchical bayesian inference. *Geophysical Journal International*, 197(3):1719–1735, 2014.

Zhifeng Kong, Wei Ping, Jiaji Huang, Kexin Zhao, and Bryan Catanzaro. Diffwave: A versatile diffusion model for audio synthesis. *arXiv preprint arXiv:2009.09761*, 2020.

Frank Permenter and Chenyang Yuan. Interpreting and improving diffusion models from an optimization perspective. *arXiv preprint arXiv:2306.04848*, 2023.