Cold Diffusion Model for Seismic Denoising

Daniele Trappolini[1,4], Laura Laurenti[1], Giulio Poggiali[2], Elisa Tinti[2,4], Fabio Galasso[5], Alberto Michelini[4], Chris Marone[2,3]

[1]Department of Computer, Control and Management Engineering, Sapienza University of Rome, Rome, Italy
[2]Department of Earth Science, Sapienza University of Rome, Rome, Italy
[3]Department of Geosciences, Pennsylvania State University, University Park, Pennsylvania, USA
[4]Istituto Nazionale di Geofisica e Vulcanologia, Roma, Italy
[5]Department of Computer Science, Sapienza University of Rome, Rome, Italy

Key Points:

- First application of a Deep Learning (DL) model using cold diffusion to denoise seismic data;
- Cold Diffusion Model for Seismic Denoising (CDiffSD) demonstrates a superior performance in scenarios with low SNR, which is particularly challenging and crucial for effective seismic data analysis;
- CDiffSD model outperforms existing methods thereby establishing a new standard in seismic data denoising.

Corresponding author: Daniele Trappolini, dtrappolini@diag.uniroma1.it

Abstract

Seismic waves contain information about the earthquake source, the geologic structure they traverse, and many forms of noise. Separating the noise from the earthquake signal is a critical first step in seismic waveform analysis. This is, however, a difficult task because optimal parameters for filtering noise typically vary with time and, if chosen inappropriately, they may strongly alter the original seismic waveform. Diffusion models based on Deep Learning (DL) have demonstrated remarkable capabilities in the restoration of images and audio signals. However, those models assume a Gaussian distribution of noise, which is not the case for typical seismic noise. Diffusion models trained on Gaussian noise do not perform well in seismic applications; therefore, we introduce a "cold" variant of diffusion models in which both clean and noisy seismic traces are restored. Here, we describe the first Cold Diffusion Model for Seismic Denoising (CDiffSD), including key design aspects, model architecture, and noise handling. We demonstrate that CDiffSD provides a new standard in performance, outperforming existing methods. Our model provides a significant advance for seismic data denoising and establishes a new state-of-the-art in the field.

Plain Language Summary

Seismic waves contain information about earthquakes and the earth's structure but any seismic waveform is, to a variable extent, contaminated by noise. Separating noise from earthquakes is important in order to enhance signals quality and, as a consequence, improve subsequent analyses. However, this task can be challenging because not only noise characteristics change in time, frequency and amplitude, but also because an incorrect denoising procedure might significantly alter important features of the seismic waves. Recently, deep learning techniques have proven to be valuable tools in enhancing images and audio signals. But these techniques usually expect the noise to follow a certain pattern that doesn't match the more complex noise found in seismic data. To solve this, we've developed a new approach called the Cold Diffusion Model for Seismic Denoising (CDiffSD). This model, specifically designed to handle the types of noise found in seismic data, shows better performances than previous methods in removing noise and restoring seismic signals, ultimately setting a new high standard in the field.

1 Introduction

Seismograms contain signals generated by earthquakes and by other unidentified sources categorized in general as 'noise' (e.g., oceanic waves, wind, vehicular traffic, sonic booms, quarry activities, and instrument malfunctions.) It is standard practice in seismology to denoise waveforms to improve the performance of the subsequent analyses, such as P- and S-wave onset picking, earthquake source moment tensor inversion, and techniques of exploration seismology. Most commonly and in routine analysis, denoising is performed through bandpass filtering. However recent works have proposed several more sophisticated schemes to "clean" seismic traces. These include methods based on the independent component analysis (ICA) (Comon, 1994; Cabras et al., 2010; Moni et al., 2012), beamforming methods (Gibbons et al., 2008; Boué et al., 2013; Brooks et al., 2009), and MUltiple SIgnal Classification(MUSIC) (Schmidt, 1986; Bear et al., 1999). All of these methods, however, can fall short when the noise shares frequencies with the earthquake generated signal.

Denoising models have evolved to incorporate time-frequency methods, with techniques like the Wavelet transform (Gaci, 2014; Siyuan & Xiangpeng, 2005; W. Liu et al., 2016; Mousavi & Langston, 2016b; Mousavi et al., 2016; Mousavi & Langston, 2017), the Short-Time Fourier Transform (STFT) (Mousavi & Langston, 2016a), the S-transform (Tselentis et al., 2012), and other transformation-decomposition methods (Hennenfent & Herrmann, 2006; Bekara & der Baan, 2009; Neelamani et al., 2008; Han & van der Baan,

2015; Y. Liu et al., 2013; Chen & Ma, 2014; Shan et al., 2009; Tang & Ma, 2011). These techniques have proven useful but the emergence of deep learning (DL) has provided new strategies with improved performance. A notable development in this arena is the Deep Denoiser (DD) model (Zhu et al., 2019). The DD approach is based on a variant of the Variational Autoencoder (VAE)(Kingma & Welling, 2019), which generates dual masks for seismic and noise signals, enhancing waveform extraction. Another notable approach is that of van den Ende et al. (2021) who employed a DL to denoise Fiber-optic Distributed Acoustic Sensing (DAS) data. They demonstrate the potency of DL to enhance the quality of DAS and seismic data. Similarly, the Novoselov et al. (2022) project, utilizing a Dual-Path Recurrent Neural Network(DPRNN), led to another substantial stride in the application of deep learning for seismic signal denoising. These studies not only validate the efficacy of deep learning methods in seismic noise reduction but also pave the way for further innovations in this field.

Here we built on this topic, drawing parallels with techniques used in speech enhancement, a field closely related to seismic denoising. Speech enhancement has recently seen the use of models such as GANs (Pascual et al., 2017; Donahue et al., 2018; Cao et al., 2022; Kim et al., 2021) and VAEs (Fang et al., 2021; Leglaive et al., 2020, 2018; Bie et al., 2022). However, the recent trend points to the growing success of Diffusion Models (Sohl-Dickstein et al., 2015; Ho et al., 2020), which are now outperforming their predecessors. Using techniques like cold diffusion or Gaussian diffusion for denoising presents several advantages over approaches that use binary masks, especially in terms of flexibility, reconstruction quality, and the ability to handle complex noise; while binary generally retain advantages in terms of simplicity, speed, interpretability, and computational efficiency. Here, we investigate the application of diffusion models for seismic denoising. These models typically transform the input into an isotropic Gaussian distribution through the consistent addition of Gaussian noise. In the reverse process, diffusion probabilistic models aim to remove the anticipated noise from the corrupted input, thus recovering the original signal. However, given the non-Gaussian nature of seismic noise, traditional diffusion models are not directly applicable.

This challenge led us to explore the emerging Cold Diffusion model (Bansal et al., 2022; Yen et al., 2023), which adapts the diffusion process by replacing Gaussian noise with other types of degradation processes. The Cold Diffusion model demonstrates how diffusion models can effectively restore signals impaired by various types of degradation. Its inherent properties make it particularly suitable for tasks such as speech source separation in practical settings with non-Gaussian noise. Building on this, our research aims to adapt the cold diffusion paradigm for seismic trace denoising. This adaptation involves specific modifications, primarily in the sampling algorithm, to suit the unique challenges of seismic data. The result is a Cold Diffusion Model for Seismic Denoising (CDiffSD).

2 Methods

2.1 Formalization of the problem

We begin with problem formulation and application of a novel diffusion model specifically designed for seismic denoising. The primary challenge in seismic signal processing is to extract the earthquake signal, denoted as x_0, from a noisy signal $y = x_0 + x_n$. This signal y consists of the desired seismic signal x_0 and an unwanted noise component x_n. The goal is to develop a model that can effectively learn to approximate the function $f(y) = x_0$, thereby isolating the earthquake signal from the noise. To address this challenge, we introduce a diffusion probabilistic model, which utilizes both forward and reverse processes for noise reduction:

1. **Diffusion process** (or Forward process) is defined as a T-step Markov chain that gradually adds Gaussian noise to the recorded earthquake x_0:

$$D(x_{1:T}|x_0) := \prod_{t=1}^{T} D(x_t|x_{t-1}) = \prod_{t=1}^{T} N(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I) \tag{1}$$

2. **Reverse process** (or Backward process) aims to restore x_0 from the latent variable x_T based on the following Markov chain:

$$R_\theta(x_{0:T}) := R(x_T) \prod_{t=1}^{T} R_\theta(x_{t-1}|x_t) := R(x_T) \prod_{t=1}^{T} N(x_{t-1}; \mu_\theta(x_t, t), \sum_\theta(x_t, t)) \tag{2}$$

where $\beta = 1 - \alpha$ serves as a key parameter that controls the process of adding and removing noise in signal during the training and inference process. In particular, the Markov formulation asserts that a given distribution depends only on the previous timestep, hence we can rewrite (2) as:

$$R_\theta(x_{t-1}|x_t) := N(x_{t-1}; \mu_\theta(x_t, t), \sum_\theta(x_t, t)) \tag{3}$$

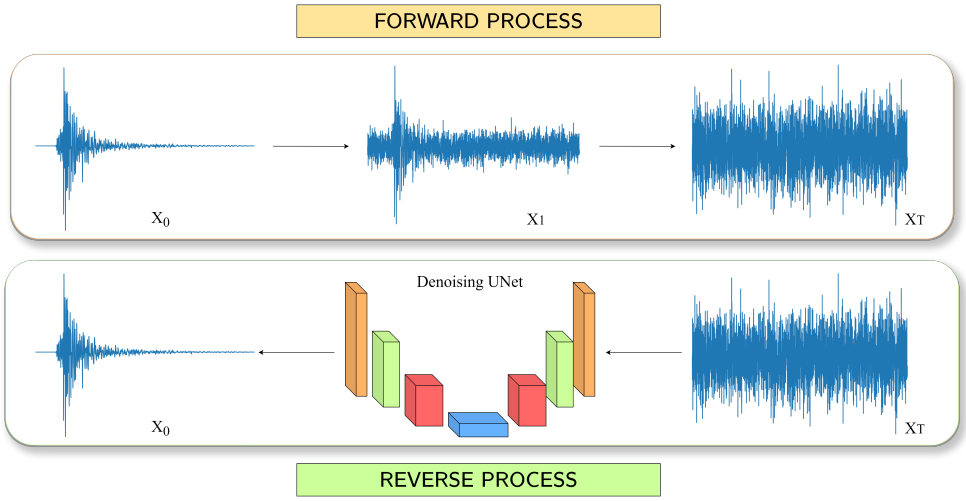

Figure 1. Sketch of how the Diffusion Process is adapted for seismic data. The standard Forward Process, which typically adds Gaussian noise, is modified to incorporate real noise, which defines so-called Cold Diffusion. The Reverse Process then employs neural networks to recover the recorded earthquake from the noise-enhanced data, illustrating the transition from noisy data back to the recorded data.

2.2 Diffusion Models

We explore diffusion models in some detail in order to elucidate key aspects of the training phase for our DL model and its operational principles. Understanding these elements is essential for appreciating how diffusion models achieve effective noise reduction and signal recovery in complex data sets. Starting from Equation (3), we can define:

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta t}{\sqrt{1-\hat{\alpha}_t}}\epsilon_\theta(x_t, t)) \tag{4}$$

The function $\mu_\theta(x_t, t)$ predicts the mean of x_{t-1} by removing the estimated Gaussian noise $\epsilon_\theta(x_t, t)$ in x_t, and the variance of x_t is fixed to a constant $\hat{\beta}_t = \frac{1-\hat{\alpha}_{t-1}}{1-\hat{\alpha}_t}\beta_t$.

The employed strategy is as follows: during training, a random time step t is sampled, and the signal is progressively degraded with Gaussian noise until reaching time

Algorithm 1 Diffusion Model Training

repeat
  $x_0 \sim D(x_0)$
  $t \sim$ Uniform$(\{1, ..., T\})$
  $\epsilon \sim N(0, I)$
  Take gradient descent step on $\nabla_\theta \|\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon_t\ t)\|^2$
until converged

t, after which the signal is restored. Once the model is trained in this manner, the sampling process follows. The model removes noise step by step from time step t to T, as in Equations 1 and 2. This process is generally motivated by two factors. First, diffusion models can be harnessed to generate novel synthetic data starting from a strongly degraded step. The second motivation is that, particularly in denoising tasks, the step-by-step noise removal approach is expected to yield superior performance compared to a direct procedure. Our initial experiments involved adding Gaussian noise to earthquake

Algorithm 2 Diffusion Model Sampling

$x_T \sim N(0, I)$
for $t = T, ..., 1$ **do**
  $z \sim N(0, I)$ if $t > 1$, else $z = 0$
  $x_{t-1} = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{1-\alpha_t}{\sqrt{1-\alpha_t}}\epsilon_\theta(x_t, t)) + \sigma_t z$
end for
return x_0

seismic traces, but we found that most seismic noise does not conform to Gaussian noise characteristics. Therefore, we adopted a more general and effective approach using a model based on Cold Diffusion (Bansal et al., 2022). This model generalizes diffusion models by replacing Gaussian noise with real noise. By using real seismic noise patterns, the model can more accurately and effectively perform denoising tasks, reflecting the actual complexities and variations found in seismic data, aligning better with the behavior of seismic traces to be denoised.

This approach marks a significant leap forward in applying diffusion models to practical tasks, integrating the use of real noise. Such integration not only confirms the findings of previous research (Bansal et al., 2022; Yen et al., 2023), which successfully applied the cold diffusion model in computer vision and speech enhancement respectively, but also highlights the limitations of traditional methods dependent on Gaussian noise. This is because Gaussian noise may not adequately represent the real-world characteristics of seismic data.

2.3 Proposed Method: Cold Diffusion Seismic Denoising (CDiffSD)

The core of our model involves degrading a one-dimensional earthquake, in the form of a seismic record, x_0 (the target), with recorded seismic noise x_n, to produce an out-of-domain sample (noisy signal): $x_T = x_0 + x_n * NRF$. Here, x_0 represents an earthquake recorded by a seismometer. While x_0 serves as a 'clean' sample in our context, it's important to note that it inherently contains some level of noise, given its real, non-synthetic origin.

The "Noise Reduce Factor" (NRF) is a key element in our specific analysis. It's responsible for calibrating the amplitude of the noise signal (x_n) in relation to the earthquake signal's amplitude, often indicated by the amplitude of S-waves in the data. By

choosing a NRF value within the range 0.4 to 0.65, we ensure that the noise does not dominate the trace compared to the earthquake. We work with data from different stations that independently record noise and earthquake signals. It's worth mentioning that we mix earthquake x_0 and noise x_n recorded from different seismic stations, to improve generalizability and robustness. This setup provides our model with an input x_T and a ground truth x_0, enabling effective backpropagation of loss and performance measurement.

Therefore, concerning the specific degradation introduced in diffusion models, we can rephrase the degradation at time T as follows:

$$x_t = D_{x_T}(x_0, t) = \sqrt{\alpha_t} x_0 + \sqrt{1 - \alpha_t} x_T \tag{5}$$

where x_0 is the recorded earthquake $x_T = x_0 + x_n * NRF$ and $\alpha \in [0, 1]$ is the parameter interpolation weight. α can also be regarded as the amount of information retained in the diffusion process, and it can alternatively be defined as $1 - \beta$, where beta represents the amount of noise introduced in the degradation, such parameters are defined a priori by a scheduler.

Regarding the specific operation of cold diffusion models, our approach is delineated using the improved training algorithm proposed by (Yen et al., 2023):

Algorithm 3 Cold Diffusion Enanched Training

for $n = 1, ..., N_{iter}$ **do**
  Sample clean data x_0
  Sample $t \sim$ Uniform$(\{1, ..., T\})$
  $x_t \leftarrow D(x_0, t), \hat{x}_0 \leftarrow R_\theta(x_t, t)$
  Sample $t^{'} \sim$ Uniform$(\{1, ..., t\})$
  $\hat{x}_{t'} \leftarrow D(\hat{x}_0, t^{'}), \hat{\hat{x}}_0 \leftarrow R_\theta(\hat{x}_{t'}, t^{'})$
  Take gradient descent step on $\nabla_\theta(||\hat{x}_0 - x_0||_1 + ||\hat{\hat{x}}_0 - x_0||_1)$
end for

This approach enhances the robustness of the training phase when applied in the presence of non-Gaussian noise. During the training phase, the model randomly selects a time step t within the range $[0, T]$. At this point, the signal is degraded by introducing recorded noise, after which a restoration operation is applied. This step is crucial as it simulates the process of denoising, where the model learns to reverse the effects of noise on the signal. The algorithm further deepens its learning by reiterating this process with a new time step t' where $t' < t$. At this stage, the signal undergoing degradation is not the recorded earthquake, but rather the one that has already been restored in the previous step. The signal is then degraded again up to the new time step t' and subsequently restored. This iterative process of degrading and restoring at various time steps allows the model to learn more robustly, adapting to the complexities introduced by real noise patterns. The improved training algorithm is tolerant to shifting errors during the sampling process. As we can observe, Algorithm 3, the training process incorporates $\hat{x}_{t'}$, which is the denoised signal. This results in $\hat{x}_{t\prime}$, which now contains the misalignment error that may occur during the sampling process.

2.3.1 Input Assumptions

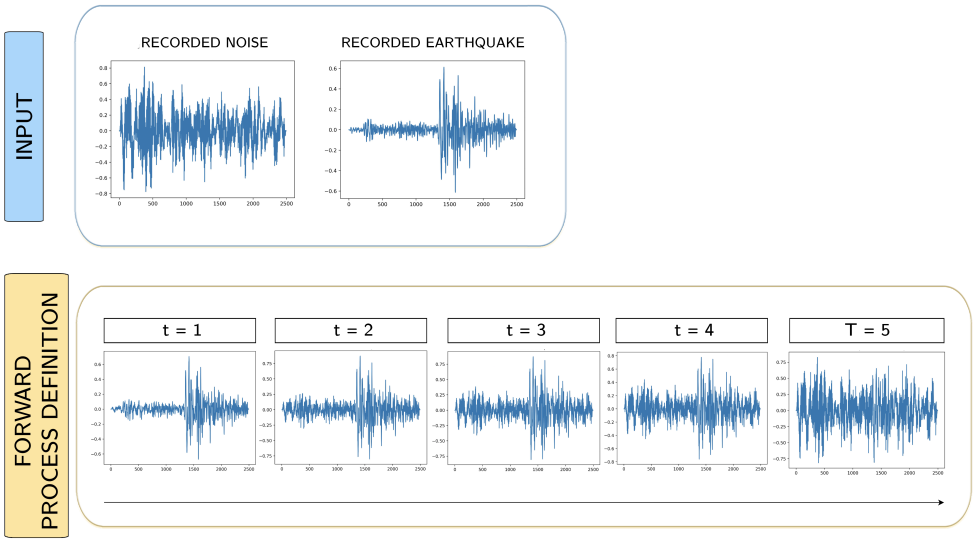

Figure 2. An illustration of the forward process with real noise for $T = 5$. The recorded noise, the recorded earthquake, and their various combinations are presented according to Equation 5. Notably, the combinations of noise and earthquake magnitudes are dictated by a scheduler, which pre-determines the levels of β and α. The level of noise at level T=5 is defined by the NRF.

In our seismic denoising approach, we separately normalize the noise and earthquake data. We adopt a trace-specific method, normalizing each seismic trace (earthquake and noise) across its East-West (E), North-South (N), and Vertical (Z) channels. This normalization process aligns the maximum and minimum values within these channels, standardizing the data to a range of [-1,1]. Such an approach ensures that each component retains its relative amplitude, enabling precise and balanced analysis. This also enhances generalizability for each type of seismic trace that the end user wants to denoise.

In the training phase for each seismic trace, we begin by merging a normalized earthquake trace with a normalized noise trace. The noise component is scaled using the NRF, adjusting its intensity in the noisy signal before the forward process is applied.

The creation of the 'noisy signal' x_T, a combination of the earthquake and scaled noise signals, leads to the 'forward process'. Here, a stochastic variable t, ranging from 0 to a predetermined maximum T, is chosen for further noise modulation. At $t = 0$, we have a recorded earthquake signal with no additional noise, whereas at $t = T$, the noise is at its full scale, set at 1.

2.3.2 Model Configuration

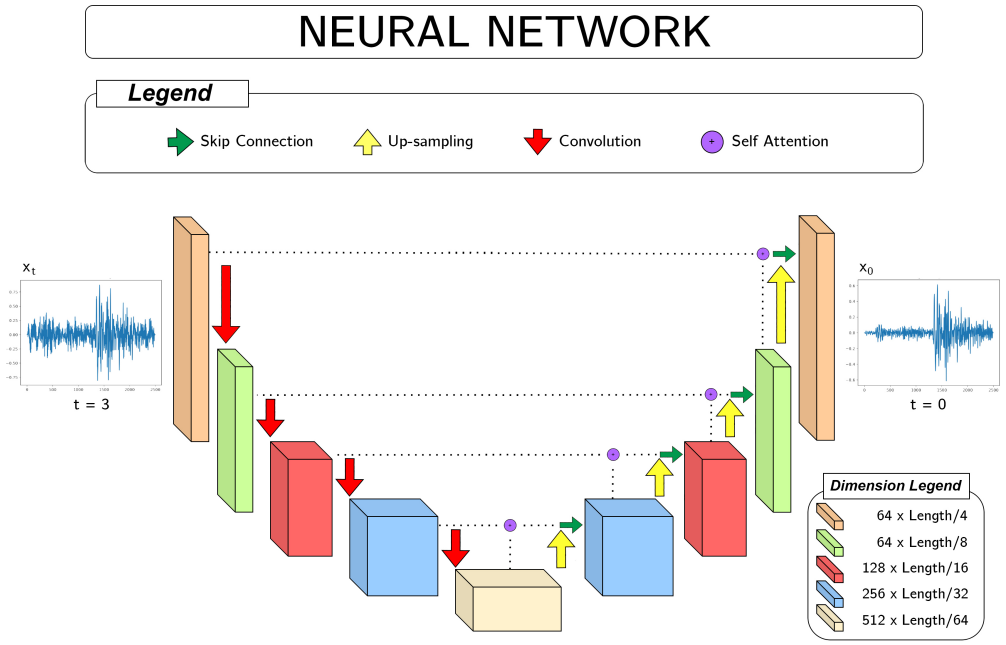

Figure 3. CDiffSD employ combinations of convolutional layers, ResNet blocks, and attention mechanisms to process one-dimensional data efficiently. The attention mechanisms are particularly important as they allow the models to selectively weigh the importance of different parts of the input data, aiding in extracting meaningful features for downstream tasks.

As a building block of the diffusion model, we adopt a neural network model inspired by the 1D U-Net (Ronneberger et al., 2015) design, for the processing of one-dimensional data streams such as time series or audio signals. The network begins with 1D convolutional layers, each equipped with 64 filters of a kernel dimension of 7, instrumental for the initial extraction of salient features. This is followed by the integration of temporal processing units that leverage sinusoidal positional encoding to effectively capture the temporal intricacies inherent within the data. These units then employ two linear layers dedicated to feature refinement and are paired with the Gaussian Error Linear Unit (GELU) (Hendrycks & Gimpel, 2016) activation function to instill the requisite non-linearity. As the architecture progresses, it introduces dimensionality manipulation layers consisting of ResNet modules (He et al., 2016) pivotal for feature conservation during downsampling and 1D convolutional layers for further data refinement. Post-downsampling, a series of upsampling layers are implemented, designed to elevate data dimensionality by merging ResNet blocks with dedicated upsampling operations. A noteworthy feature of our design is the mid-level blocks, each outfitted with dual residual units. They exploit attention mechanisms crucial for highlighting pertinent data characteristics. The network culminates with terminal residual blocks that are succeeded by 1D convolutional layers, making definitive outputs typically manifest as singular channels.
The U-Net block is applied for each iteration of the diffusion model from each t_i to 0 and then again from t_{i-1} to 0 and so on until the end of the process.

We trained models with 3 configurations: $T = 20, T = 100, T = 300$. These diverse scheduler assumptions allowed us to evaluate how performance metrics vary with increasing T, highlighting the trade-off between model performance and computation time,

which is a crucial consideration in seismic monitoring room operations where balancing processing speed and precision is essential.

Particularly in the inference phase, understanding the impact of T on both model performance and computational efficiency is vital. For applications requiring rapid trace processing, like real-time seismic monitoring, a preference for speed may be necessary, though it could impact precision. Conversely, in tasks where accuracy is the priority, such as dataset cleaning, a greater emphasis on precision may be warranted, even at the expense of longer processing times.

We compared our approach using the same seismic dataset with DD, that we consider as the reference for the state of the art. For this task, DD underwent comprehensive training for 400 epochs, while our model completed its training in just 150 epochs. This difference was due to our model's learning dynamics and efficiency. We initiated our model's training with a learning rate of 1e-3 and employed a scheduler to reduce this rate gradually, ensuring controlled and stable convergence.

2.3.3 Inference with Direct and Sampling Reconstruction

Cold diffusion models involve distinct methods to reconstruct the signal including the adoption of direct or sampling reconstruction. These methods represent approaches within the framework of diffusion models, each with unique operational mechanisms and implications for model performance. Understanding the nuances of these methods is crucial for comprehending the overall efficacy and application potential of diffusion models.

For the range of configurations used in training our models ($T = [20, 100, 300]$), we applied these configurations to both direct and sampling reconstruction. In the context of diffusion models, the distinction between 'direct' and 'sampling' approaches is pronounced, marked by their differing operational mechanisms.

The **'direct'** method involves applying the reverse process using the U-Net architecture to transition from a specific timestep t_n directly to zero. Conversely, the **'sampling'** method incrementally applies this reverse transition from a specific timestep t_n to zero, but crucially, it traverses through all intermediate timesteps t_i, where $i \in [n-1, 0]$. This results in applying the U-Net architecture multiple times (n).

A key aspect of the cold diffusion paradigm is evaluating the effectiveness of the sampling procedure, which is hypothesized to outperform the direct approach. If the direct method, particularly using U-Net alone, yields comparable results, it would call into question the necessity of the complex training infrastructure typically associated with diffusion models. We provide a detailed comparison between the direct and sampling methods in section 4.

2.3.4 Metrics

For enhanced clarity, we define here the metrics used in our study now and then in Section 4 we provide a detailed commentary on the results.

1. **Signal to Noise Ratio (SNR)** is a measure used to compare the level of a signal (earthquake in this case) to the level of background noise. A higher SNR indicates that the seismic signal stands out clearly from the background noise, facilitating accurate analysis and interpretation. We defined SNR as in (Zhu et al., 2019):

$$10 \log_{10} \frac{\sigma_{signal}}{\sigma_{noise}}.$$

where σ_{noise} and σ_{signal} are the standard deviation of waveforms before and after the P arrival, respectively.

2. **Cross-correlation** is a widely used measure of similarity between two signals. We compute the zero-lag cross-correlation (CC) between the recorded earthquake signals (before noise is added) that represents our ground truth x_0 and the denoised ones to evaluate the performance of the different models in reconstructing the recorded waveform.
3. To evaluate the **picking** performances of the proposed method, we applied the deep learning phase picker PhaseNet (Zhu & Beroza, 2019) to the waveforms and compared the retrieved arrival times with the labeled picked phases of the catalog ($\sim 70\%$ of manually picked and $\sim 30\%$ of automatic picked). In this way we can assess the impact of the denoiser on P and S arrival determination, the accuracy of which enables the calculation of a well constrained location.

3 Data Sources and Selection

In our study, we focus on a subset extracted from the STanford EArthquake Dataset (STEAD) (Mousavi et al., 2019). This section is dedicated to elucidating the composition of the subset, detailing the following components:

1. We selected specific seismic stations to gather earthquakes and others for noise, with some overlap, providing a clear trace of the data's origin for our analysis (Figure 4).
2. The distribution of seismic events across the globe (Figure 4) is mapped out, with these events sorted into training, validation, and test sets. This classification helps us to assess the model's effectiveness and its generalizability across different regions.
3. We applied constraints to the dataset, including the magnitude and proximity to the seismic stations.

STEAD features a significantly larger number of stations for earthquake data compared to those used for noise. Moreover, the majority of these stations are concentrated within the U.S. territory. In our study, we utilize a ratio of (1786/2613) stations for the extraction of earthquake data, representing a fraction of the total available. For seismic noise, we have selected a subset corresponding to 306 stations dedicated to noise recording.

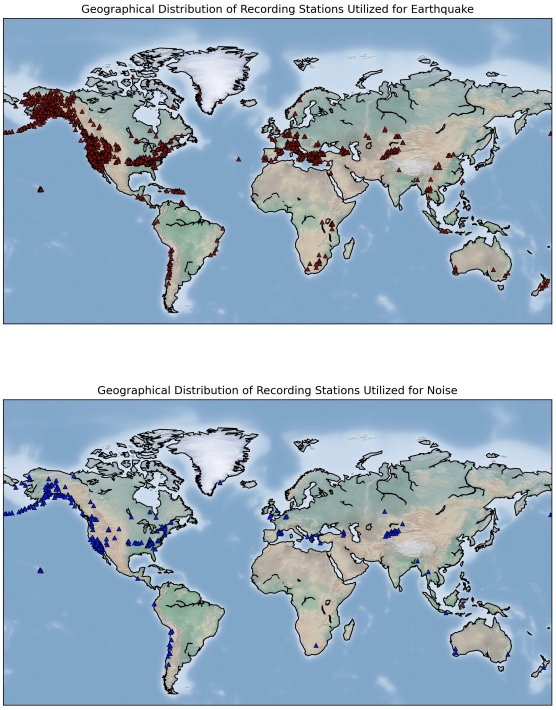

Figure 4. The maps show the subset of stations of the STEAD (STanford EArthquake Dataset) used for the recorded earthquake signal (upper) and the recorded noise (bottom).

Throughout our analysis, we consistently sample seismic traces of 30-second durations, based on the following criteria:magnitude > 2,earthquake-station distance < 100

km, and P-wave arrival after 7 seconds. Figure 5 shows the frequency-magnitude statistics for our data set.

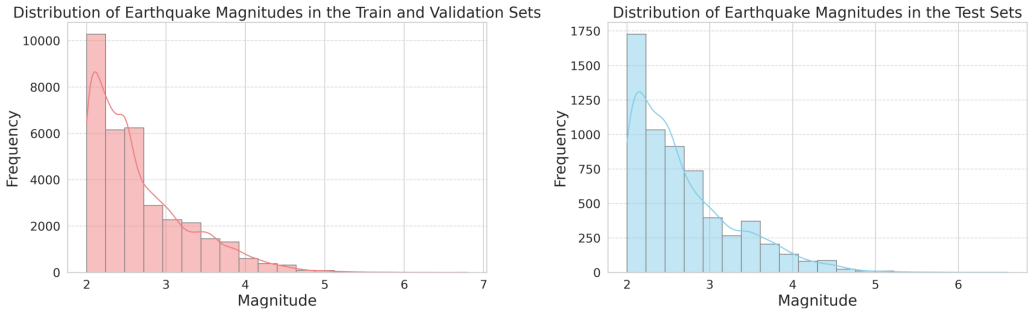

Figure 5. The histograms illustrate the frequency distribution of earthquake magnitudes within our dataset, with the left panel representing the training and validation sets and the right panel the test set.

We chose an inclusive approach for training, leveraging the full spectrum of available data, without any SNR selection criteria. While this might seem disadvantageous initially, a model that performs well under these conditions can be versatile across various scenarios. For researchers looking to retrain this model on their datasets, especially when specific datasets are limited, it may be advantageous not to put restrictive filters such as SNR.

Our dataset was divided into training (30491 traces), validation (3441 traces), and test (5994 traces) as illustrated in Figure 6. Such a division in machine learning ensures model reliability and generalizability. The training set aids the model's primary learning, the validation set is used for hyperparameter adjustments, and the test set objectively evaluates the model's performance on unseen data.

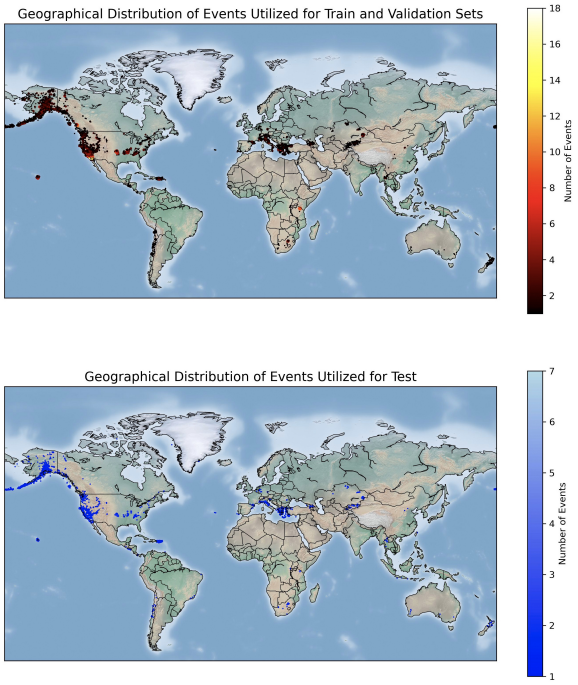

Figure 6. The image presents two maps of the geographical distribution of seismic events used in our study, with the upper map illustrating the events for the training and validation sets marked in red, and the bottom map showing the events for the test set in blue. The color intensity on each map corresponds to the number of events, with darker shades indicating a lower concentration of events in that location.

For more details on the specific train, validation and test configurations, please refer to our GitHub repository at the following link:

`https://github.com/Daniele-Trappolini/Diffusion-Model-for-Earthquake`.

4 Results

In the following we present our results and discuss the validity of our model by adopting quantitative and qualitative catagories. The metrics used for each are provided section 2.3.4.

4.1 Quantitative Results

4.1.1 Signal to Noise Ratio (SNR)

A comparison of the SNR metric for the denoised waveforms obtained with different models and configurations is shown in Fig. 7. Note that Figure 7 includes the same metric for the original earthquake signals (labeled "earthquake") and those with added noise (labeled "eqk + noise"). The latter are the inputs to the denoiser algorithm. The performances of the different models appear aligned, with DD differing by a slightly lower median but greater variability in output SNR. In Fig. 8 we classified the noisy obser-

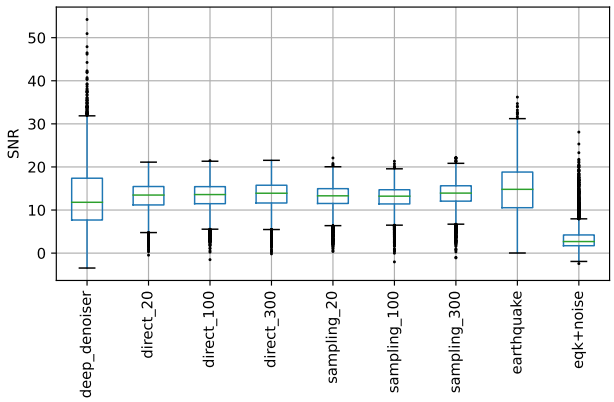

Figure 7. SNR comparisons using box-plots for various model configurations applied to the test set. The original signals (earthquake) and the ones with added noise (eqk+noise) have respectively the higher and lower SNR, as expected. The different denoising models appear overall aligned, with direct and sampling showing slightly higher median values and tighter distributions with respect to DD.

vations as a function of the SNR before denoising to highlight the effectiveness of our models in cleaning the seismic traces. The performance of our CDiffSD are consistently superior with respect to DD in low SNR scenarios. This aspect is crucial, given that low SNR conditions correspond to more complex and heavily noisy seismic traces precisely where an effective denoising solution is most needed. The high-quality performance of our model in these low SNR environments is demonstrated in Fig. 8. We note in particular model reliability and efficacy in extracting correct signals from noisy data. This proficiency is important in real-world seismological applications, especially for discovering lower magnitude earthquakes often hidden in the noise.

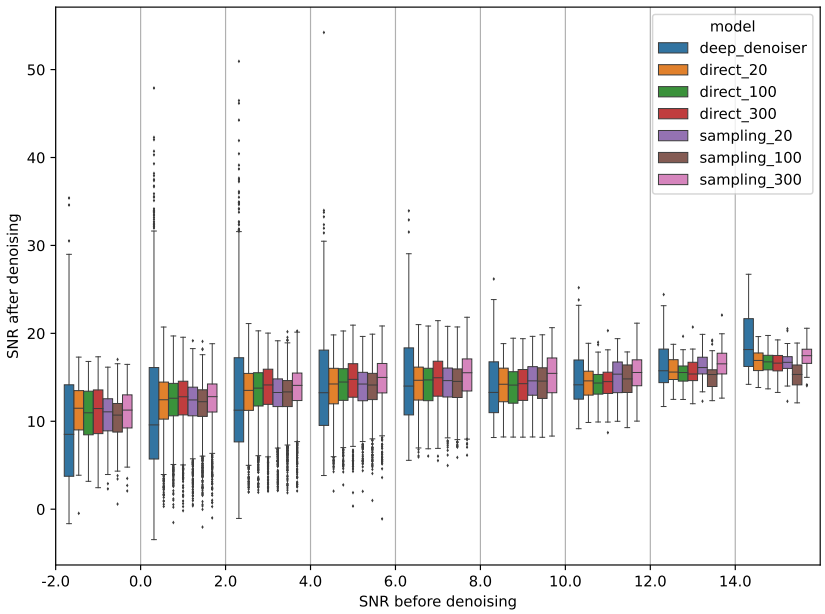

Figure 8. Distributions of SNR values of denoised waveforms for different ranges of input SNR. The SNR statistics after denoising are computed on 2dB wide ranges of input SNR. CDiffSD models show higher performances in low SNR scenarios, while DD is superior for the higher SNR signals. We study the range: $SNR < -2.0 \cup SNR > 16.0$, which covers 99% of real data. Solid bars within each model (color) show the median value.

While the cold diffusion approach excels in low SNR scenarios, the binary mask-based method DD exhibits greater variability and tends to perform better in higher SNR conditions, benefiting from its ability to provide a clear-cut signal delineation (Fig. 7 and Fig. 8). In particular, DD shows improved performance when the input SNR is higher than $\sim$ 14 and is get worse at lower input SNR while our models remain consistently effective for a large range of input SNR. An example of high input SNR conditions can be found in the Supporting Information.

4.1.2 Cross Correlation

We evaluate the similarity between original signals and denoised signals, by showing the statistics of the computed CC values, in Fig. 9. A higher CC indicates a greater similarity between the denoised trace and the original signal. In this figure, we see that all CDiffSD models show similar performance and they are all consistently higher than DD. To better highlight the variability of CC values obtained from the different traces of the test set, in Fig. 10 we show the distribution of CC values between denoised and original traces as a function of CC of the noisy traces with original signals (x axis), that is, CC of traces before denoising is applied. The performances for both direct and sampling are higher than DD for every considered range of CC before denoising.

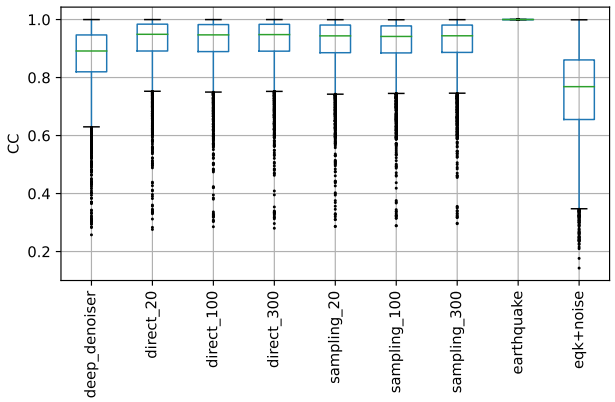

Figure 9. Cross-correlation (CC) comparisons for various model configurations applied to the test set. Higher CC values indicate greater similarity between the denoised trace and the original signal. All CDiffSD models show similar performance and that they are consistently higher than DD.

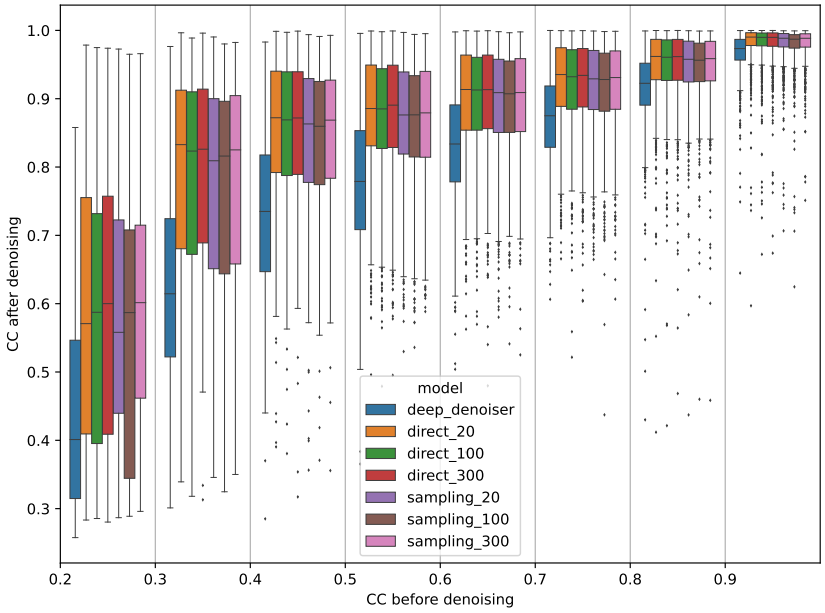

Figure 10. Distribution of CC values between original and denoised signals (y axis) as a function of CC before application of denoising. The statistics are computed for ranges of 0.1. CDiffSD models show better performances with respect to DD for all the ranges considered. The difference is more noticeable especially at low CC values before denoising.

For each model considered we see better performance, with higher values of CC after denoising (Fig. 10). Another noteworthy aspect is that at higher noise levels, thus lower CC before denoising (values from 0.2 to 0.3), models with T = 300 outperform their counterparts. As expected, these performance disparities tend to converge with an in-

crease in CC before denoising, corresponding to a relative reduction in noise compared to the signal.

4.1.3 Phase arrival picks

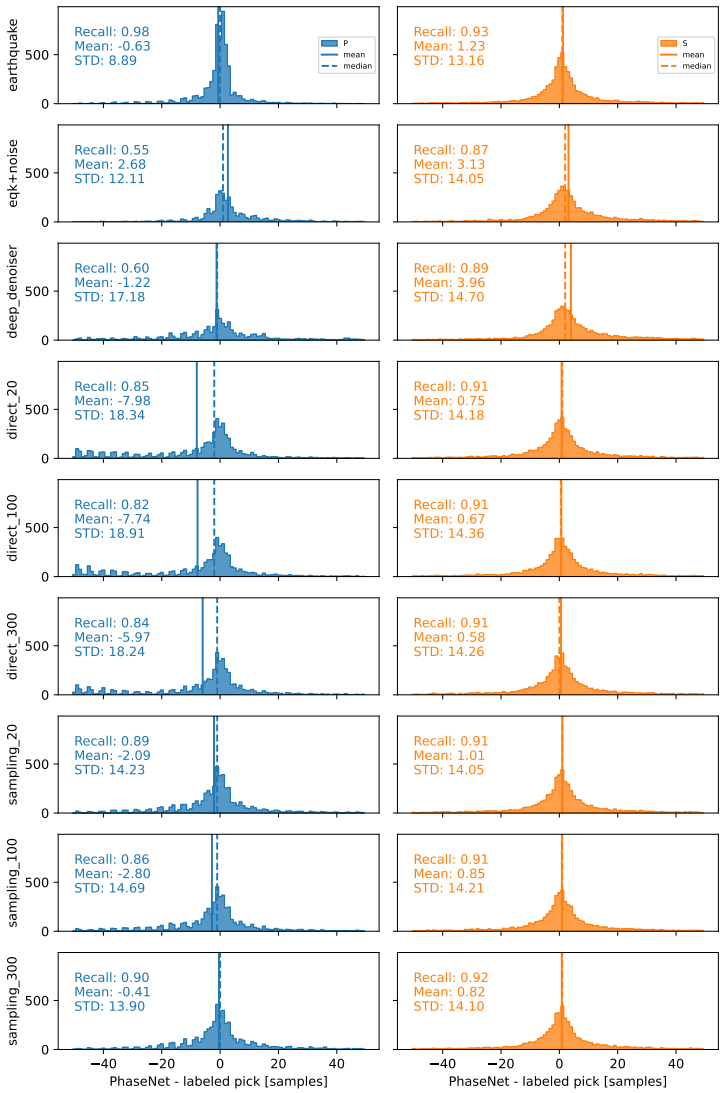

Figure 11. The histograms display the distributions of P-wave (blue) and S-wave (orange) arrival time differences between automated PhaseNet detections and label picks (in samples). The results obtained using the original seismograms, eqk+noise and DD are shown for comparison in the first, second and third row, respectively. The remaining rows show the results for different CDiffSD models applied to the same data subset, offering insights into the accuracy of wave arrival time detection by each model. Central tendency metrics, such as mean and median, are indicated in these histograms, highlighting any potential skewness in the distribution towards either early or late picks for both P and S waves.

The histograms in Fig. 11 provide a visual representation of the efficacy of different seismic signal denoising methods — "direct", "sampling", and DD — in retrieving a signal and preserve P- and S-wave onsets. The accuracy of automated P and S-wave arrival time picks by PhaseNet is compared to label picks. The histograms are organized by method and parameter variations, displaying the distribution of arrival time discrepancies measured in samples.

In the case of earthquake (i.e., no noise added, top histogram), the P-wave pick difference distribution exhibits spreads that are narrower than those of the S-wave and this is in full agreement with the expected behavior.

When noise is introduced, the pick difference distributions for P-waves and S-waves tend to converge towards a more similar pattern. This convergence can be attributed to the primary impact of noise on P-waves, owing to their lower amplitude compared to S-waves. As a result, the performance with added noise on P waves detection is much more degraded than on S waves detection with the same level of noise because P-waves have also smaller amplitudes. This observation is further supported by the recall values for S waves, which remain greater than 0.85 not only for all the denoising methods, but also for the noisy traces (earthquake + noise). In contrast, the recall rate for P-waves is consistently lowered by the presence of noise (Fig. 12). For these reasons we focus our analysis on P-wave picks.

As seen in Fig. 11 the distribution of the "direct" methods show pronounced negative skews, with mean values far from 0. This indicates a tendency of PhaseNet to pick P-waves slightly before the labeled picks for the waveforms denoised with "direct" methods. The reason of this behavior is most likely to be attributed to noise remaining in the denoised traces processed with the "direct method". This in turn can mislead PhaseNet to an early detection (see the "direct" example in Fig. 13). This tendency, however, is mitigated completely by the CDiffSD "sampling" method, as shown in Fig. 13. In particular, we see that the "sampling" methods display recall rates that are consistently high for both P and S, especially the 300 configuration, indicating a good denoising performance and the ability to recover the labeled phases.

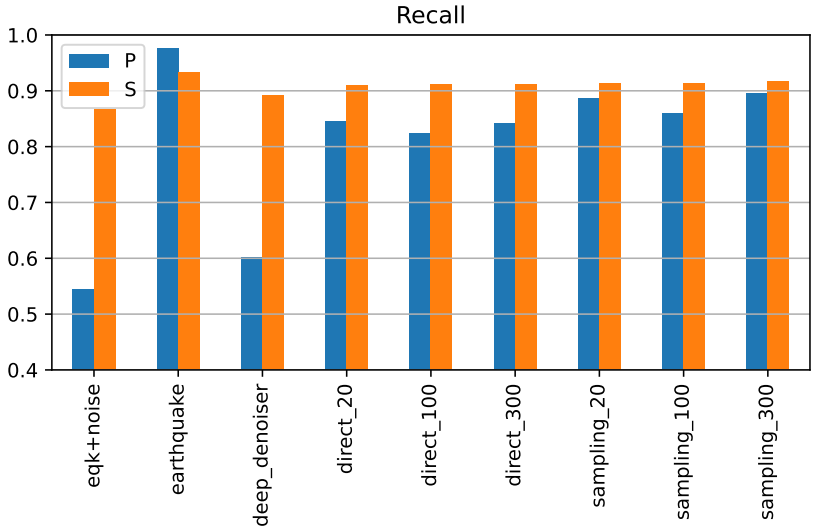

Figure 12. Comparison of recall rates for P and S waves between the different methods within a fixed window of 50 samples. S-waves recall rates are aligned almost for all models, indicating that the noise level is not enough to affect the S-waves because of the greater amplitude. P-waves recall rates instead show significant disparities between the DD approach and other methods, suggesting a lower performance of DD in preserving the P onset in these cases. The 'sampling-300' method is confirmed as the one with better performances.

From the comparison of the results obtained with the "direct", "sampling", and DD methods, it is evident that each method influences the automated pick accuracy differently. The "sampling" method, particularly at higher parameter settings, demonstrates a notable alignment with label picks, suggesting its superiority in mitigating noise and enhancing the precision of automated picking systems. It is also noteworthy that the recall values for P-waves shown in Fig. 12 are higher than DD for both "sampling" and "direct" methods, which suggests that in these cases DD does not preserve accurate P-wave onsets.

4.2 Qualitative Results

Qualitative factors are useful as side-by-side comparisons between the outcomes of different approaches. In the supplementary materials, we delve deeper into the analysis of seismic traces, examining the impact of denoising on picking. This includes a thorough examination of both the strengths and limitations of our model. We highlight instances where our model excels in denoising, as well as situations in which it does not perform optimally.

The examples below and those in the supplementary are organized with the same layout: in the top panel we compare the noisy signal (grey) with the denoised signal (black); in the middle panel we compare the original signal (green) with the denoised signal (black); the bottom panel is a zoom on the P-wave arrival.

4.2.1 Qualitative Picker Analysis: Direct Vs Sampling

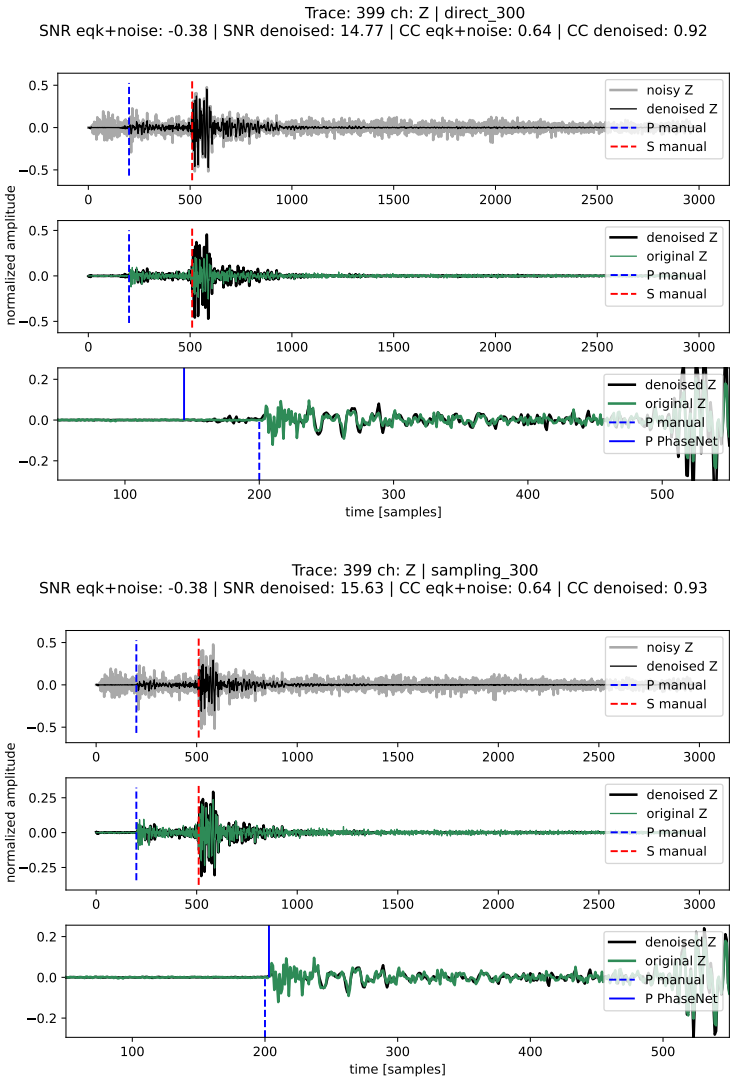

Figure 13. Comparison of a trace processed using 'direct_300' and 'sampling_300' methods. Notably, the 'direct_300' retains some of the noise preceding the P wave arrival, which is instead filtered out in the 'sampling_300' results. This noise before the P-wave retained with the the 'direct_300' method explains the tendency for this model to cause early picks (as seen in Fig. 11), as the residual noise can lead to earlier detections.

In the first example shown in Fig. 13 we compare the 'direct_300' and the 'sampling_300' methods. Here "sampling" method is found to be more effective than the "direct" method in denoising the seismic signal, and this is particularly evident from the middle and bottom panels, where the denoised signal in the "sampling" method match more closely the original signal. In contrast, the "direct" method shows more significant deviation from the original, especially before the P-wave arrival. This example is also useful because it provides insight into the tendency of the "direct" methods to cause spurious early P-picks. The direct method, in fact, retains some pre-arrival noise, which can trigger an early pick in automatic approaches such as PhaseNet. This is less of an issue

in the sampling method, as seen in the lower set of traces, where the denoised signal is cleaner, and the P-wave arrivals are closer to the labels. The implication for seismic processing is significant since the sampling method appears to produce cleaner signals and more accurate P-wave arrival times as a direct consequence. We note that this is crucial for various seismological applications such as earthquake location and tomographic imaging.

4.2.2 Qualitative Picker Analysis: DD Vs Sampling

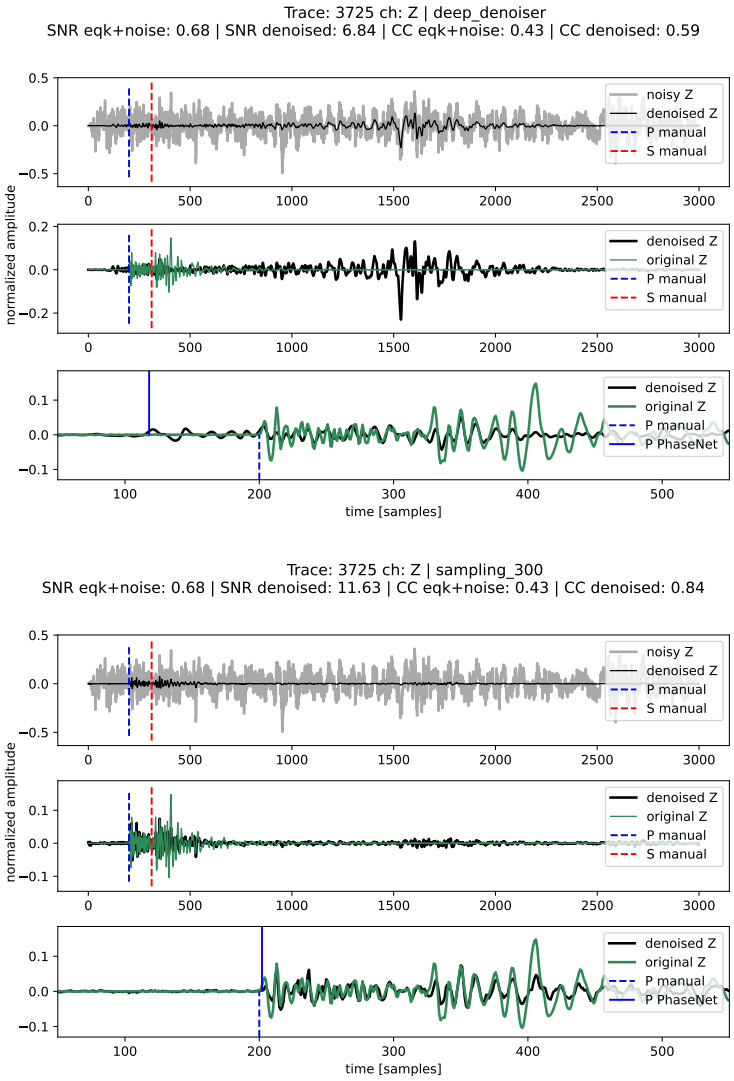

Figure 14. Comparison between a seismic trace processed with 'deep_denoiser' and 'sampling_300' methods. The 'sampling_300' method demonstrates a closer match to label phase picks and a more precise amplitude preservation, despite the substantial noise present in the original signal. DD also retains a high amplitude noise signal at around 1500 samples that 'sampling_300' manages to filter out almost completely.

Fig. 14 exemplifies the concepts previously discussed in Fig. 8, highlighting the performance of our model compared to that of the 'deep denoiser' in scenarios with very low Signal-to-Noise Ratio (SNR) before denoising. The figure demonstrates clearly how an extreme noise situation can lead to an error in phase picking for the 'deep denoiser', whereas the 'sampling' method is capable to reconstruct accurately the correct P wave arrival despite the presence of significant noise.

4.2.3 Qualitative Amplitude Analysis: Direct Vs Sampling

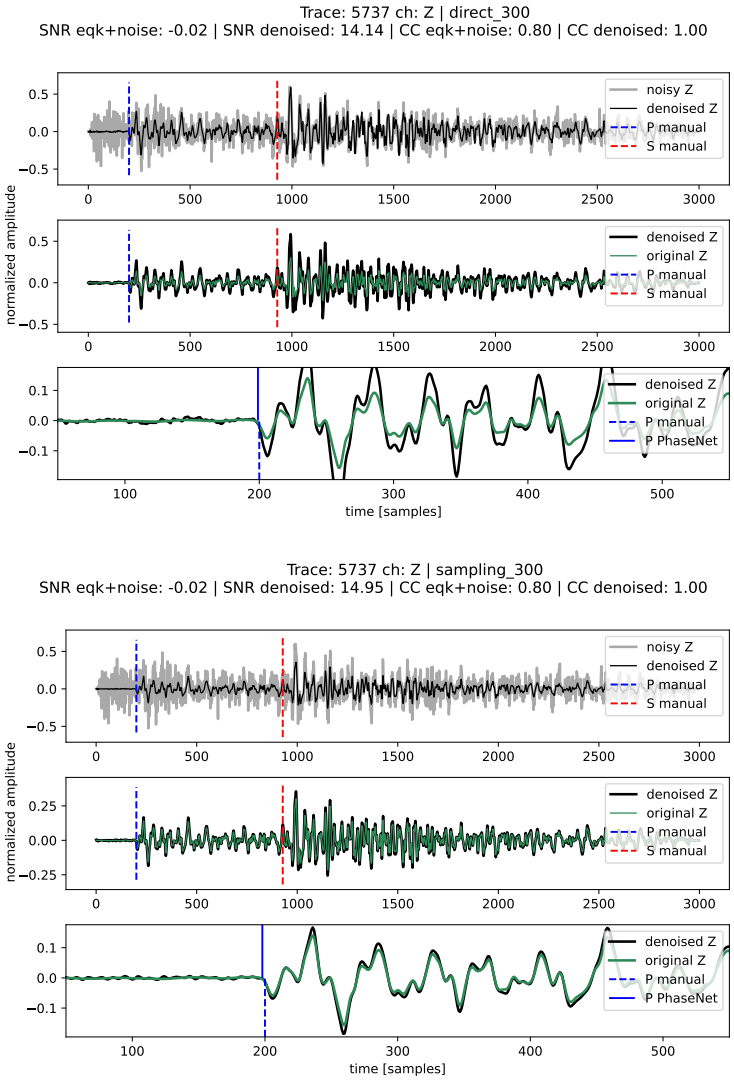

Figure 15. Comparison of a seismic trace processed with 'direct_300' and 'sampling_300' methods. It is particularly significant that the 'sampling_300' technique demonstrates an enhanced ability for amplitude reconstruction compared to the 'direct_300' method.

In Fig. 15 we show a comparative analysis of seismic signal denoising methods to investigate the importance of amplitude preservation. The Cold Diffusion Model employing a sampling strategy ('sampling_300') demonstrates a superior performance in maintaining the amplitudes of the seismic signal. In practice, the denoised signal aligns more

accurately with the original waveform, preserving the integrity of the amplitude across the signal's duration. This is particularly evident in the detailed zoomed-in analysis, where the 'sampling_300' method displays remarkable congruence with the original signal, as evidenced by the minimal and consistent residuals. In contrast, the direct application of a U-Net model ('direct_300') displays a slight but discernible attenuation in amplitude, most noticeable in segments with higher amplitude peaks. The increased residuals associated with the 'direct_300' method suggest a more significant alteration of the signal after the denoising process. Therefore, the Cold Diffusion Model with sampling stands out as the most effective method for seismic data denoising (amongst those tested here), especially where the preservation of amplitude is critically important.

5 Model assessment: Assessing the Impact of Exclusive Noise Input

In this section we aim to test the behaviour of the model in no-earthquake scenarios, i.e. with inputs containing only noise. This is done in order to verify whether the model doesn't generates any artifacts in the absence of signal generating false earthquakes.

Cold Diffusion is based on the model's ability to learn the broad data distribution during training, which generally includes a variety of seismic traces with different levels of noise. Therefore, the model should be able to generalize and identify traces that are entirely dominated by noise, even without direct exposure to specific types of earthquake samples where there is no earthquake signal. Based on these assumptions, we seek to verify if our results align with the theoretical expectations.

We have used the entire noise test set as input, without combining it with the earthquake data. Theoretically, with a perfect denoising, the expected output would be a trace composed exclusively of zeros, in the real context the trace should approach zero.

We applied the model without retraining, meaning the model's weights have never been exposed to the absence of earthquake traces as ground truth. To assess the correctness of the output we set an amplitude threshold between ± 0.02 to decide whether the output could resemble a trace of zeros. The direct and sampling methods have correctly reconstructed the expected signal in 60.3% and 88.6% of cases, respectively. This different performance highlights the sampling method's superior capability in recognizing the absence of earthquake signals and adapting to it.

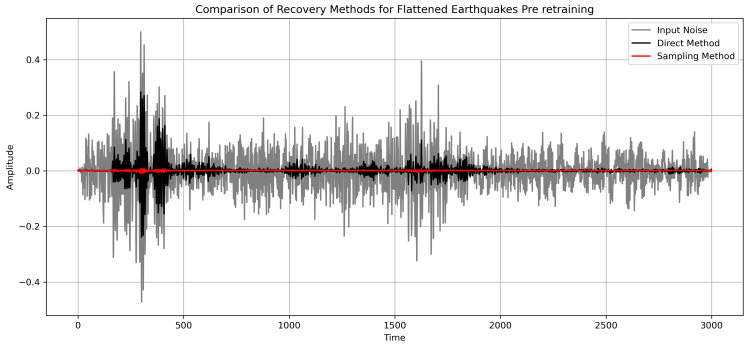

Figure 16. Example of the outputs of direct (in black) and sampling (in red) methods in case of a noise only input (in grey). No retraining is performed here, i.e. the models have never been exposed to zero-traces as ground truth for noise-only input. The direct method fails in recovering a zero-trace since it introduces artificial signals. In contrast, the sampling method reconstructs successfully an output that resembles a zero-trace.

Given the promising results just described, we further explored this scenario by retraining the model including no-signal traces as ground truth. We focused only on a single channel for this test and incorporated 3% of the entire training set with zeroed traces to represent the absence of seismic events. The results align with our expectations, indicating an improvement in performance in the presence of noise alone. Specifically, the cases where zero traces are retrieved increases to 68.2% and 90.5% for direct and sampling methods, respectively. The direct method exhibits a more substantial improvement, starting from a lower baseline performance, whereas the sampling method shows a smaller increase, likely due to its performance already approaching saturation.

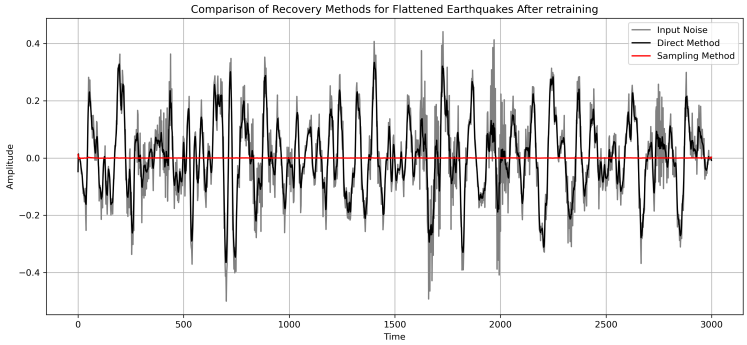

Figure 17. Example of the comparison between the sampling method (in red) and the direct method (in black), the input (in gray) for both methods is only noise. In this case the models have been retrained with zero-traces as ground truth for noise-only traces. The sampling method succeeds in reconstructing a zero-trace. On the other hand, the direct method outputs noise, indicating a less accurate reconstruction in this scenario.

Regarding the results post-retraining, it should be noted that the output trace of the sampling method shown in Figure 17 is indeed close to the expected zero-trace. On the contrary, low amplitude noise was still present in the output of the non retrained-case shown in Figure 16. This highlights the importance of including flat traces during the training.
In this evaluation of the CDiffSD on these cases comprised solely of noise, we proved that it is not imperative to include such examples in training to accurately discern between noise and genuine seismic signals. However, including these kind of signals in training, improves the capability of effectively identifying traces that are comprised solely of noise.

6 Conclusion

Our study has demonstrated promising results in pick accuracy (4.1.3), Signal-to-Noise Ratio (SNR) enhancement (4.1.1), and Cross-Correlation metrics (4.1.2), thus affirming the validity of cold diffusion denoising for seismological applications. In addition to these achievements, it is important to emphasize that, despite SNR and Cross-Correlation metrics aligning with other models, the sampling with $T = 300$ demonstrates its superiority in practical, applied contexts, such as P-phase onset picking.

While SNR and Cross-Correlation are critical metrics for assessing the quality of the reconstructed signal, not every part of the seismic trace holds equal significance. In fact, the preservation of the integrity of the P- and S-wave arrivals is fundamental. As

highlighted in Section 4, the most effective model in this regard is the one utilizing sampling with $T = 300$. This model's ability to maintain the aspects of the seismic trace, particularly the arrival times of these key waveforms, underscores its practical superiority in applied seismological contexts.

The findings discussed in Section 4, while serving as a good base, should be regarded, however, as a preliminary step towards addressing a broader spectrum of open questions and potential model enhancements.

A significant direction for future advancement lies in the broadening of our dataset. Our initial explorations aimed to establish the feasibility of these methods. Moving forward we could potentially develop a more generalized model by retraining on the full STEAD and INSTANCE (Michelini et al., 2021) datasets, encompassing collectively several million traces compared to the $\sim$ 40k traces used in this study. This expanded model would be capable of effectively treating noise in a wide range of seismological contexts without the need for further retraining, thus significantly boosting its applicability and robustness across diverse seismic scenarios.

In conclusion, the model presented exhibits significant potential for enhancing seismic traces, facilitating more accurate onset picking of P- and S-waves. Moreover, it holds promise for extracting earthquakes from noise—events that may have eluded human detection or other approaches. Such capability could contribute to expanding seismic catalogs. While further refinements are conceivable, this method, which is borrowed from speech enhancement tasks, has proven its validity in the intricate domain of seismological analysis. This cross-disciplinary innovation underscores the model's versatility and suggests broader applicability in extracting and analyzing subtle seismic signals.

Acronyms

CC Cross Correlation
CDiffSD Cold Diffusion Model for seismic denoising
DAS Distributed Acoustic Sensing
DD Deep Denoiser
DL Deep Learning
DM Diffusion Model
DPRNN Dual-Path Recurrent Neural Network
E East-West
eqk Earthquake
ERC European Research Council
GAN Generative Adversarial Network
GELU Gaussian Error Linear Unit
ICA Independent Component Analysis
INGV Istituto Nazionale di Geofisica e Vulcanologia
INSTANCE Italian Seismic Dataset For Machine Learning
MUSIC MUltiple SIgnal Classification
N North-South
NRF Noise Reduce Factor
ResNet Residual Neural Network
SNR Signal to Noise Ratio
STEAD STanford EArthquake Dataset
STFT Short-Time Fourier Transform
VAE Variational Autoencoder
Z Vertical

Open Research Section

The STEAD (Mousavi et al., 2019)(Seismological Tools for Earthquake Analysis and Detection) dataset is openly accessible at the following link: `https://github.com/smousavi05/STEAD` or by utilizing ObsPy, a Python library for processing seismological data (for more information on ObsPy, refer to their official site: `https://docs.obspy.org/`).

To replicate the data accurately, it is necessary to apply the filters described in the Section 3 to chunk2 of the STEAD dataset. Furthermore, specific data related to this research will soon be made available on the GitHub repository at `https://github.com/Daniele-Trappolini/Di`

Acknowledgments

This research was made possible by the generous support of both the Istituto Nazionale di Geofisica e Vulcanologia (INGV) and the European Research Council (ERC) grant 835012 (TECTONIC). We acknowledge partial funding from the MUR PNRR FAIR (PE00000013) project. Complementary funding was provided by the project INGV Pianeta Dinamico 2021 Tema 8 SOME (CUP D53J1900017001) funded by the Italian Ministry of University and Research "Fondo finalizzato al rilancio degli investimenti delle amministrazioni centrali dello Stato e allo sviluppo del Paese, legge 145/2018". The funding and resources provided by these institutions have been instrumental in advancing the scope and depth of our study. We extend our sincere gratitude to the INGV for their valuable contributions and to the ERC for their commitment to fostering scientific research and innovation.

References

Bansal, A., Borgnia, E., Chu, H.-M., Li, J. S., Kazemi, H., Huang, F., ... Goldstein, T. (2022). Cold diffusion: Inverting arbitrary image transforms without noise. *arXiv preprint arXiv:2208.09392*.

Bear, L. K., Pavlis, G. L., & Bokelmann, G. H. (1999). Multi-wavelet analysis of three-component seismic arrays: application to measure effective anisotropy at pinon flats, california. *Bulletin of the Seismological Society of America*, *89*(3), 693–705.

Bekara, M., & der Baan, M. V. (2009). Random and coherent noise attenuation by empirical mode decomposition. *Geophysics, vol. 74, no. 5, pp. V89–V98, 2009*.

Bie, X., Leglaive, S., Alameda-Pineda, X., & Girin, L. (2022). Unsupervised speech enhancement using dynamical variational autoencoders. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *30*, 2993–3007.

Boué, P., Roux, P., Campillo, M., & de Cacqueray, B. (2013). Double beamforming processing in a seismic prospecting context. *Geophysics*, *78*(3), V101–V108.

Brooks, L. A., Townend, J., Gerstoft, P., Bannister, S., & Carter, L. (2009). Fundamental and higher-mode rayleigh wave characteristics of ambient seismic noise in new zealand. *Geophysical Research Letters*, *36*(23).

Cabras, G., Carniel, R., & Wasserman, J. (2010). Signal enhancement with generalized ica applied to mt. etna volcano, italy. *Bollettino di Geofisica Teorica ed Applicata*, *51*(1).

Cao, R., Abdulatif, S., & Yang, B. (2022). Cmgan: Conformer-based metric gan for speech enhancement. *arXiv preprint arXiv:2203.15149*.

Chen, Y., & Ma, J. (2014). Random noise attenuation by f-x empiricalmode decomposition predictive filtering. *Geophysics, vol. 79, no. 3, pp. V81–V91, 2014*.

Comon, P. (1994). Independent component analysis, a new concept? *Signal processing*, *36*(3), 287–314.

Donahue, C., Li, B., & Prabhavalkar, R. (2018). Exploring speech enhancement with generative adversarial networks for robust speech recognition. In *2018*

ieee international conference on acoustics, speech and signal processing (icassp) (pp. 5024–5028).

Fang, H., Carbajal, G., Wermter, S., & Gerkmann, T. (2021). Variational autoencoder for speech enhancement with a noise-aware encoder. In *Icassp 2021-2021 ieee international conference on acoustics, speech and signal processing (icassp)* (pp. 676–680).

Gaci, S. (2014). The use of wavelet-based denoising techniques to enhance the first-arrival picking on seismic traces. *IEEE Trans. Geosci. Remote Sens., vol. 52, no. 8, pp. 4558–4563, Aug. 2014*.

Gibbons, S. J., Ringdal, F., & Kværna, T. (2008). Detection and characterization of seismic phases using continuous spectral estimation on incoherent and partially coherent arrays. *Geophysical Journal International*, *172*(1), 405–421.

Han, J., & van der Baan, M. (2015). Microseismic and seismic denoising via ensemble empirical mode decomposition and adaptive thresholding. *Geophysics, vol. 80, no. 6, pp. KS69–KS80, . doi: 10. 1190/geo2014-0423.1*.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 770–778).

Hendrycks, D., & Gimpel, K. (2016). Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*.

Hennenfent, G., & Herrmann, F. J. (2006). Seismic denoising with nonuniformly sampled curvelets. *Comput. Sci. Eng., vol. 8, no. 3, p. 16, May 2006*.

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

Kim, H. Y., Yoon, J. W., Cheon, S. J., Kang, W. H., & Kim, N. S. (2021). A multi-resolution approach to gan-based speech enhancement. *Applied Sciences*, *11*(2), 721.

Kingma, D. P., & Welling, M. (2019). An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*, *12*(4), 307–392.

Leglaive, S., Alameda-Pineda, X., Girin, L., & Horaud, R. (2020). A recurrent variational autoencoder for speech enhancement. In *Icassp 2020-2020 ieee international conference on acoustics, speech and signal processing (icassp)* (pp. 371–375).

Leglaive, S., Girin, L., & Horaud, R. (2018). A variance modeling framework based on variational autoencoders for speech enhancement. In *2018 ieee 28th international workshop on machine learning for signal processing (mlsp)* (pp. 1–6).

Liu, W., Cao, S., & Chen, Y. (2016). Seismic time-frequency analysis via empirical wavelet transform. *IEEE Geosci. Remote Sens. Lett., vol. 13, no. 1, pp. 28–32, Jan. 2016*.

Liu, Y., Li, Y., Lin, H., & Ma, H. (2013). An amplitude-preserved time–frequency peak filtering based on empirical mode decomposition for seismic random noise reduction. *IEEE Geoscience and Remote Sensing Letters*, *11*(5), 896–900.

Michelini, A., Cianetti, S., Gaviano, S., Giunchi, C., Jozinović, D., & Lauciani, V. (2021). Instance–the italian seismic dataset for machine learning. *Earth System Science Data*, *13*(12), 5509–5544.

Moni, A., Bean, C. J., Lokmer, I., & Rickard, S. (2012). Source separation on seismic data: Application in a geophysical setting. *IEEE Signal Processing Magazine*, *29*(3), 16–28.

Mousavi, S. M., & Langston, C. A. (2016a). Adaptive noise estimation and suppression for improving microseismic event detection. *Appl. Geophys., vol. 132, pp. 116–124, Sep. 2016. doi: 10.1016/j.jappgeo.2016.06.008*.

Mousavi, S. M., & Langston, C. A. (2016b). Hybrid seismic denoising using higher-order statistics and improved wavelet block thresholding. *Bull. Seismolog. Soc. Amer., vol. 106, no. 4, pp. 1380–1393, 2016*.

Mousavi, S. M., & Langston, C. A. (2017). Automatic noise-removal/signalremoval

based on general cross-validation thresholding in synchrosqueezed domain and its application on earthquake data. *Geophysics, vol. 82, no. 4, pp. V211–V227, 2017. doi: 10.1190/geo20160433.1*.

Mousavi, S. M., Langston, C. A., & Horton, S. P. (2016). Automatic microseismic denoising and onset detection using the synchrosqueezed continuous wavelet transform. *Geophysics, vol. 81, no. 4, pp. V341–V355, 2016. doi: 10.1190/geo2015-0598.1*.

Mousavi, S. M., Sheng, Y., Zhu, W., & Beroza, G. C. (2019). Stanford earthquake dataset (stead): A global data set of seismic signals for ai [dataset]. *IEEE Access*, *7*, 179464–179476.

Neelamani, R., Baumstein, A. I., Gillard, D. G., Hadidi, M. T., & Soroka, W. L. (2008). Coherent and random noise attenuation using the curvelet transform. *The Leading Edge*, *27*(2), 240–248.

Novoselov, A., Balazs, P., & Bokelmann, G. (2022). Sedenoss: Separating and denoising seismic signals with dual-path recurrent neural network architecture. *Journal of Geophysical Research: Solid Earth*, *127*(3), e2021JB023183.

Pascual, S., Bonafonte, A., & Serr'a, J. (2017). Segan: Speech enhancement generative adversarial network. *in Proc. Interspeech, 2017*.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–miccai 2015: 18th international conference, munich, germany, october 5-9, 2015, proceedings, part iii 18* (pp. 234–241).

Schmidt, R. (1986). Multiple emitter location and signal parameter estimation. *IEEE transactions on antennas and propagation*, *34*(3), 276–280.

Shan, H., Ma, J., & Yang, H. (2009). Comparisons of wavelets, contourlets and curvelets in seismic denoising. *Appl. Geophys., vol. 69, no. 2, pp. 103–115, Oct. 2009. doi: 10.1016/j.jappgeo.2009.08.002*.

Siyuan, C., & Xiangpeng, C. (2005). The second-generation wavelet transform and its application in denoising of seismic data. *Appl. Geophys.,vol. 2, no. 2, pp. 70–74, Jun. 2005*.

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

Tang, G., & Ma, J. (2011). Application of total-variation-based curvelet shrinkage for three-dimensional seismic data denoising. *IEEE Geosci. Remote Sens. Lett., vol. 8, no. 1, pp. 103–107, Jan. 2011*.

Tselentis, G.-A., Martakis, N., Paraskevopoulos, P., Lois, A., & Sokos, E. (2012). Strategy for automated analysis of passive microseismic data based on s-transform, otsu's thresholding, and higher order statistics. *Geophysics*, *77*(6), KS43–KS54.

van den Ende, M., Lior, I., Ampuero, J.-P., Sladen, A., Ferrari, A., & Richard, C. (2021). A self-supervised deep learning approach for blind denoising and waveform coherence enhancement in distributed acoustic sensing data. *IEEE Transactions on Neural Networks and Learning Systems*.

Yen, H., Germain, F. G., Wichern, G., & Le Roux, J. (2023). Cold diffusion for speech enhancement. In *Icassp 2023-2023 ieee international conference on acoustics, speech and signal processing (icassp)* (pp. 1–5).

Zhu, W., & Beroza, G. C. (2019). Phasenet: A deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, *216*(1), 261–273.

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