
ON EXTRACTING COHERENT SEISMIC WAVEFIELD USING VARIATIONAL SYMMETRIC AUTOENCODERS

A PREPRINT

✉ **Pawan Bharadwaj**
Centre for Earth Sciences
Indian Institute of Science
Bengaluru
pawan@iisc.ac.in

November 24, 2024

ABSTRACT

Many seismological applications, such as receiver function analysis, earthquake source imaging, and ambient noise tomography, rely on averaging waveforms to extract coherent information. While traditional linear stacking methods have been widely used, recent studies have demonstrated the advantages of nonlinear stacking techniques in enhancing signal-to-noise ratios. Our work effectively generalizes this concept and enables a more sophisticated form of non-linear stacking, using probabilistic generative models. We discuss the variational formulation of the Symmetric Autoencoder (SymAE) and its role in achieving disentanglement within the latent space to extract coherent information from a collection of seismic waveforms. Disentanglement involves separating the latent space into components for coherent information shared by all waveforms and components for waveform-specific nuisance information. SymAE employs a generative model that independently generates waveforms based on coherent and nuisance components, and an inference model that estimates these components from observed wavefield. By assuming the independence of waveforms conditioned on coherent information, the model effectively accumulates this information across multiple waveforms. After training, a metric based on Kullback-Leibler divergence is used to evaluate the informativeness of individual waveforms, enabling latent-space optimization and the generation of synthetic seismograms with enhanced signal-to-noise ratios.

To demonstrate the efficacy of our proposed method, we applied it to a data set of teleseismic displacement waveforms of the P wave from deep-focus earthquakes. By training the SymAE model on high-magnitude events, we successfully identified seismograms that contained robust source information. Furthermore, we generated high-resolution virtual seismograms enriched with relevant coherent source information and less influenced by scattering noise, allowing a deeper understanding of the characteristics of the earthquake source. Importantly, our method extracts coherent source information without relying on deconvolution, which is often used in traditional source imaging. This enables the analysis of complex earthquakes with multiple rupture episodes, a capability that is not easily achievable with conventional approaches.

Keywords coherent wavefield · seismic deconvolution · autoencoders; probabilistic models · deep learning · neural networks · unsupervised learning · representation learning · earthquakes · source estimation

1 Introduction

Probabilistic generative models [Goodfellow et al., 2020, Kingma and Welling, 2019, Ho et al., 2020, Rezende and Mohamed, 2015] have emerged as versatile tools across numerous scientific disciplines. Their applications span from basic image processing tasks such as compression, denoising, and inpainting to more complex problems like semi-supervised and unsupervised learning. While the core principles remain consistent, their implementation and application vary significantly across different fields. Recent advancements have witnessed a growing interest in applying

#	Code	Date	Time	Mw	Depth	Half Duration	Region
1	okt1	2013-05-24	05:44:48	8.30	598.00	35.7	Sea of Okhotsk
2	okt2	2013-05-24	14:56:31	6.70	624.00	5.6	Sea of Okhotsk
3	okt3	2008-07-05	02:12:04	7.70	633.00	17.3	Sea of Okhotsk
4	okt4	2008-11-24	09:02:58	7.30	492.00	10.9	Sea of Okhotsk
5	okt5	2013-10-01	03:38:21	6.70	573.00	5.6	Sea of Okhotsk
6	okt6	2012-08-14	02:59:38	7.70	583.00	17.8	Sea of Okhotsk
7	pb2	2015-11-24	22:45:38	7.60	606.00	16.6	Peru Brazil
8	fiji1	2018-08-19	00:19:40	8.20	600.00	30.8	Fiji Islands
9	fiji2	2018-09-06	15:49:18	7.90	670.00	21.4	Fiji Islands
10	fiji3	2014-11-01	18:57:22	7.10	434.00	-	Fiji Islands
11	fiji4	2018-11-18	20:25:46	6.80	540.00	5.8	Fiji Islands
12	fiji5	2021-04-24	00:23:38	6.50	301.00	4.5	Fiji Islands
13	bon1	2015-05-30	11:23:02	7.80	664.00	20.7	Bonin Island Japan Reg
14	van1	2014-03-05	09:56:57	6.30	638.00	3.8	Vanuatu Islands
15	spain	2010-04-11	22:08:12	6.30	609.00	-	Spain
16	nbl	1994-06-09	00:33:16	8.20	631.00	20.0	Northern Bolivia

Table 1: SymAE was trained using displacement seismograms of these deep focus earthquakes.

generative models to geophysical data [Song et al., 2021, Xiao et al., 2024, Trappolini et al., 2024, Ren et al., 2024, Sun et al., 2024]. In geophysics, a crucial aspect of generative model application lies in the physical interpretability of the generated data. In other words, the ability to generate samples conditioned on physically relevant variables is often paramount. This necessitates the development of generative models that can not only capture the statistical properties of the data, but also produce physically meaningful and informative outputs.

This study introduces a novel probabilistic generative model, the Symmetric Autoencoder (SymAE), designed to extract coherent information from seismological data. Traditional methods for extracting coherent energy often employ preprocessing steps, such as phase correction and linear stacking, to suppress nuisance variations. Recent advancements have introduced non-linear stacking [Schimmel et al., 2011, Olivier et al., 2015, Rückemann, 2012, Weaver and Yoritomo, 2018, Xie et al., 2020] techniques that assign weights to waveforms based on quality assessments. Our approach extends this concept by performing a non-linear decomposition of waveforms into constituent components. Subsequently, a neural network learns optimal weights for each component, enabling a more sophisticated form of non-linear stacking. We refer to this process as accumulating coherent information across waveforms, which we argue is more effective than traditional stacking methods.

SymAE [Bouchacourt et al., 2018, Bharadwaj et al., 2022] was developed to learn a disentangled representation of the seismic wavefield to generate waveforms conditioned on coherent information relevant to the task. However, generating waveforms enriched in coherent information while minimizing the effects of task-irrelevant nuisances was not addressed. This work introduces a metric for quantifying waveform informativeness based on its contribution to accumulated coherent information, enabling the generation of relevant waveforms. We pose the extraction of coherent information as a latent-space optimization problem. This process is data-driven and does not rely on knowledge of the underlying physical signal model. Additionally, an upgrade to SymAE that is based on spatial transformer networks [Jaderberg et al., 2015] and attend-infer-repeat networks [Eslami et al., 2016] is proposed to handle traveltimes shifts in the observed time domain waveforms, potentially obviating the need for preprocessing steps like cross-correlation.

Our numerical experiments concentrate on the unsupervised extraction of coherent earthquake source information [Dziewonski et al., 1981, Sipkin, 1982, Abercrombie, 2015] from a collection of seismograms. The seismograms of deep focus earthquakes, listed in Tab. 1, used are obscured by path scattering effects. After learning a representation of seismograms through SymAE, we demonstrate its application to quantify scattering effects in individual seismograms. Then, we generate seismograms with reduced path-related noises to gain deeper insight into the characteristics of the earthquake sources. In addition to real earthquake data, we considered two synthetic earthquake sources for generating training and testing datasets with roughly 5000 seismograms. These synthetic sources, along with representative path effects and sample seismograms, are illustrated in Fig. 1. To generate the seismograms, we convolved the synthetic sources with path effects representing P-arrivals and their coda. Random time shifts, band-limited noise, and normalization were applied to simulate realistic traveltimes errors and noises commonly observed in seismological data. The purpose of the synthetic experiment is to retrieve coherent source time functions from noisy seismograms, as plotted in Figs. 1g and 1h, without deconvolution using known path effects. To the best of our knowledge, this marks the first work to discuss the extraction of earthquake source information in an entirely model-free manner.

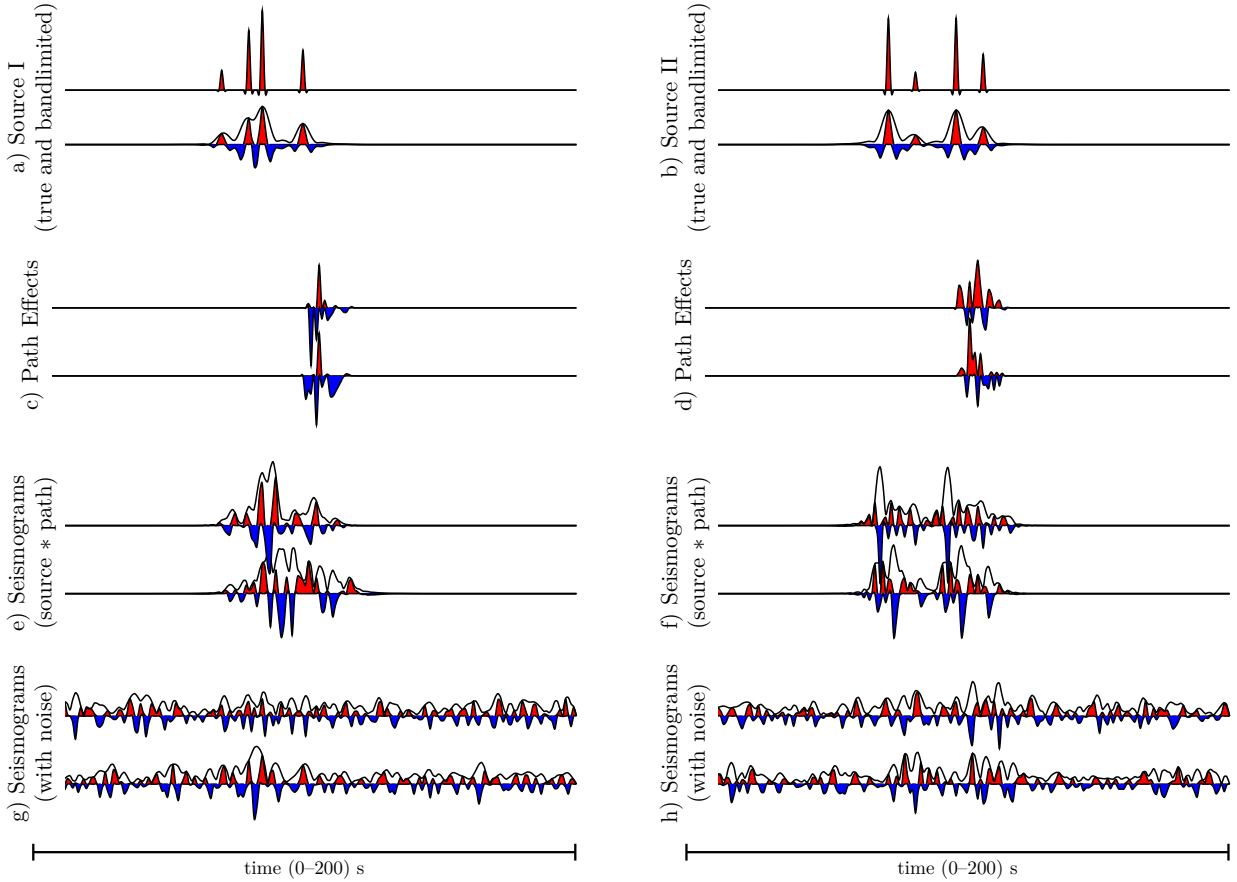


Figure 1: Extracting coherent source information from seismograms modeled as the convolution of band-limited source signatures with path effects, corrupted by noise and time shifts.

While this study focuses on earthquake source estimation, the proposed methodology is applicable to a broader range of seismological applications. Many seismic tasks involve extracting coherent information from waveforms, such as receiver function analysis, earthquake source imaging, and ambient noise tomography. These applications often rely on grouping waveforms based on shared characteristics, either source-related or path-related.

- Receiver function analysis [Zandt and Ammon, 1995, Zheng et al., 2018, Dalai et al., 2021, Bloch et al., 2023], where coherent information on the near-receiver Earth’s crustal structure should be extracted from waveforms related to numerous teleseismic earthquake events. Compared to linear stacking [Hu et al., 2015], where numerous events are required, a non-linear method to extract coherent information can allow for investigation of azimuthal and epicentral-distance dependence (attributed to anisotropy and 3D orientation) of the receiver functions.
- Coherent information from station-pair cross-correlations [Weaver, 2001, 2005, Campillo and Paul, 2003, Wapenaar, 2004, Snieder, 2004] is extracted linearly, where the emergence of this information is proportional to the square root of the recording time and inversely proportional to the square root of the distance between stations. The coherent information retrieved in this scenario pertains to the medium between the stations and is utilized for tomography purposes.

This work utilizes a D -dimensional vector \mathbf{x}_j^k to represent the measured seismic waveform in the time domain. The subscript j indicates the source of the waveform (the j th earthquake), and the superscript k indicates the location where it was measured (the k th receiver). To represent a group of waveforms recorded due to a single earthquake source (the j th earthquake) and measured at different receivers, we denote it as a set $\mathbf{X}_j = \{\mathbf{x}_j^k\}$, where we simply drop the receiver index k and use a capital letter. Although this paper focuses mainly on source-based groups \mathbf{X}_j , we can similarly establish a notation for receiver-based collections by omitting the source index, denoted as \mathbf{X}^k , to tailor the equations presented here to the application scenarios discussed previously. Finally, to denote a group of waveforms

for all sources and receivers, we remove the superscript and use the subscript to access the elements, i.e. $\mathbf{X} = \{\mathbf{x}_i\}$ represents the group of all measured waveforms, where i iterates over all the source-receiver combinations.

2 Variational Autoencoder

Variational autoencoders [Kingma and Welling, 2013, 2019, Prince, 2023, Bishop, 2023, Higgins et al., 2017, Burgess et al., 2018, VAE] are a class of probabilistic generative models designed to infer the distribution $P(\mathbf{X})$ from the observed wavefield. VAE is described as a nonlinear latent variable model, which expresses the probability distribution of each element \mathbf{x}_i as an integral over the latent variable \mathbf{z} of the joint probability distribution of \mathbf{x}_i and \mathbf{z} :

$$P(\mathbf{x}_i) = \int P(\mathbf{x}_i, \mathbf{z}) d\mathbf{z} = \int P(\mathbf{x}_i | \mathbf{z})P(\mathbf{z}) d\mathbf{z}. \quad (1)$$

Here, the *likelihood* $P(\mathbf{x}_i | \mathbf{z})$ is the conditional probability of \mathbf{x}_i given \mathbf{z} , modeled as a normal distribution \mathcal{N} , where its mean and diagonal covariance matrices are produced by a decoder function \mathbf{f} of \mathbf{z} and parameters ϕ : $P(\mathbf{x}_i | \mathbf{z}, \phi) = \mathcal{N}(\mathbf{x}_i | \mathbf{f}[\mathbf{z}, \phi])$. In this work, we parameterize \mathbf{f} as a neural network with convolutional layers. The *prior* probability of \mathbf{z} , denoted as $P(\mathbf{z})$, is also assumed to follow a normal distribution with zero mean $\mathbf{0}$ and the identity covariance matrix \mathbf{I} : $P(\mathbf{z}) = \mathcal{N}(\mathbf{z} | \mathbf{0}, \mathbf{I})$. In seismology, the latent or hidden variables can be considered as either the path or source parameters that must be specified to produce the wavefield. In this sense, the decoder \mathbf{f} can be seen as a network performing the *forward modeling* operation given the latent variables.

In order to determine the decoder parameters ϕ , which control the generation of the wavefield based on the latent variable setting, VAE aims to maximize the log likelihood of the wavefield \mathbf{X} . Assuming independently and identically distributed waveforms, we can write:

$$\log[P(\mathbf{X} | \phi)] = \sum_i \log[P(\mathbf{x}_i | \phi)] = \sum_i \log \left[\int P(\mathbf{x}_i | \mathbf{z}, \phi)P(\mathbf{z})d\mathbf{z} \right]. \quad (2)$$

Since the log likelihood expressed as an integral over \mathbf{z} is intractable [Bishop, 2023], we decompose it into two terms:

$$\log[P(\mathbf{X} | \phi)] = \sum_i \int Q_i(\mathbf{z}) \log \left[\frac{P(\mathbf{x}_i | \mathbf{z}, \phi)P(\mathbf{z})}{Q_i(\mathbf{z})} \right] d\mathbf{z} + \sum_i D_{\text{KL}}(Q_i(\mathbf{z}) \| P(\mathbf{z} | \mathbf{x}_i, \phi)). \quad (3)$$

It can be shown that this decomposition holds for any auxiliary distribution $Q_i(\mathbf{z})$, and it is commonly employed within the framework of variational inference to facilitate the approximation of intractable integrals. In eq. 3, $P(\mathbf{z} | \mathbf{x}_i, \phi)$ is the *posterior* distribution of \mathbf{z} given \mathbf{x}_i — it specifies the distribution of the path and the source latent variables responsible for the generation of the waveform \mathbf{x}_i . The second term here gives a summation over the Kullback-Leibler (KL) divergence

$$D_{\text{KL}}(Q_i(\mathbf{z}) \| P(\mathbf{z} | \mathbf{x}_i, \phi)) = \int Q_i(\mathbf{z}) \log \left[\frac{Q_i(\mathbf{z})}{P(\mathbf{z} | \mathbf{x}_i, \phi)} \right] d\mathbf{z}, \quad (4)$$

which measures the difference between each auxiliary distribution $Q_i(\mathbf{z})$ and the posterior distribution $P(\mathbf{z} | \mathbf{x}_i, \phi)$. As the KL divergence is always non-negative (can be shown using Jensen’s inequality), the first term in eq. 3, which is the Evidence Lower Bound (ELBO), serves as a lower bound on the log-likelihood of the data. VAE focuses on the maximization of this term, given that the estimation of the posterior $P(\mathbf{z} | \mathbf{x}_i, \phi)$ in the second term remains intractable. It has to be noted that the maximization of the ELBO inherently facilitates the minimization of the KL divergence between $Q_i(\mathbf{z})$ and $P(\mathbf{z} | \mathbf{x}_i, \phi)$.

VAE utilises an encoder network \mathbf{g} , parameterized by θ , to perform amortized inference of \mathbf{z} . Specifically, the network \mathbf{g} generates auxiliary distributions that are conditioned on the input waveform \mathbf{x}_i . These auxiliary or approximate posterior distributions are assumed to be Gaussian distributions. The mean and diagonal covariance of these Gaussian distributions are determined by the output of the network \mathbf{g} when given the input \mathbf{x}_i and the parameters θ . This is mathematically represented as $Q_i(\mathbf{z}) = Q(\mathbf{z} | \mathbf{x}_i, \theta) = \mathcal{N}(\mathbf{z} | \mathbf{g}[\mathbf{x}_i, \theta])$. The underlying rationale is that the encoder function \mathbf{g} , again parameterized using convolutional layers in our case, performs *inverse modelling* to estimate the distribution of the source and path parameters given a wavefield \mathbf{x}_i . Finally, this choice of $Q_i(\mathbf{z})$ makes the Evidence Lower Bound (ELBO) in Eq. 3 for VAEs a function of the parameters ϕ and θ , which is maximized to determine their optimal values:

$$\mathbb{L}(\theta, \phi) = \sum_i \int Q(\mathbf{z} | \mathbf{x}_i, \theta) \log \left[\frac{P(\mathbf{x}_i | \mathbf{z}, \phi)P(\mathbf{z})}{Q(\mathbf{z} | \mathbf{x}_i, \theta)} \right] d\mathbf{z}. \quad (5)$$

Using the product rule of the logarithm, the ELBO during the training process is further decomposed into two terms:

$$\mathbb{L}(\theta, \phi) = \sum_i \int Q(\mathbf{z} | \mathbf{x}_i, \theta) \log [P(\mathbf{x}_i | \mathbf{z}, \phi)] d\mathbf{z} - \sum_i D_{\text{KL}}(Q(\mathbf{z} | \mathbf{x}_i, \theta) \| P(\mathbf{z})). \quad (6)$$

The first term represents the reconstruction loss, which measures how accurately the model can reconstruct the original data from the latent variables. During training, the reconstruction loss term is approximated using a Monte Carlo estimate, such that

$$\mathbb{L}(\theta, \phi) = \sum_i \log [P(\mathbf{x}_i | \mathbf{z}^*, \phi)] - \sum_i D_{\text{KL}}(Q(\mathbf{z} | \mathbf{x}_i, \theta) \| P(\mathbf{z})), \quad (7)$$

where \mathbf{z}^* denotes a sample drawn from $Q(\mathbf{z} | \mathbf{x}_i, \theta)$. Maximizing this term improves the model’s ability to reproduce the original data. The second term in eq. 6 acts as a regularizer, minimizing the KL divergence between the approximate posterior distribution $Q(\mathbf{z} | \mathbf{x}_i, \theta)$ and the prior distribution $P(\mathbf{z})$.

In summary, VAE is a type of generative model that is well suited to learning latent representations of seismic data. They consist of two key components: an encoder \mathbf{g} and a decoder \mathbf{f} . The encoder maps the input waveforms to a compressed latent-space representation \mathbf{z} . Although VAEs are capable of producing virtual waveforms by drawing samples from the aggregated posterior distribution $\frac{1}{I} \sum_{i=1}^I Q(\mathbf{z} | \mathbf{x}_i, \theta)$ of I waveforms, and then processing these samples through the decoder, their utility is limited. The virtual waveforms generated by a basic VAE often lack a clear physical interpretation, making it difficult to extract task-relevant information. To address this limitation, recent research has focused on learning disentangled representations within the VAE framework Burgess et al. [2018]. The next section will discuss learning such disentangled representations using the symmetric autoencoder.

3 Variational Symmetric Autoencoder

This section discusses the variational formulation of the Symmetric Autoencoder (SymAE). We will explore how SymAE achieves disentanglement within the latent space, ultimately enabling the extraction of coherent information from seismic wavefield data. Disentanglement refers to the process of separating the latent space into distinct factors that correspond to physically meaningful components of the wavefield. This translates to separating the latent space into factors representing coherent source information and nuisance path information. We consider the task of representing each group \mathbf{X}_j of seismograms, corresponding to a particular earthquake, using a latent space variable \mathbf{z} partitioned into the following components:

- Source component (\mathbf{s}): this component captures the source information accumulated from all the seismograms of \mathbf{X}_j . It is shared by all the seismograms within the group, reflecting the common signal originating from the earthquake itself.
- Path component (\mathbf{p}^k): this component represents nuisance information, specifically the path scattering effects unique to each individual seismogram \mathbf{x}_j^k within the group \mathbf{X}_j . It is denoted by \mathbf{p}^k for the k th seismogram, highlighting its waveform-specific nature.

Fig. 2a illustrates the generative model assumed by SymAE, from which the following arguments are inferred. The source information and the path scattering information for distinct seismograms are assumed to be independent within the prior distribution. Consequently, the joint probability distribution of \mathbf{s} , \mathbf{p}^k and \mathbf{p}^l is expressed as the product of their marginal probability distributions

$$P(\mathbf{s}, \mathbf{p}^k, \mathbf{p}^l) = P(\mathbf{s})P(\mathbf{p}^k)P(\mathbf{p}^l), \quad (8)$$

for any two path codes of seismograms with indices k and l . In order to factor the likelihood $P(\mathbf{x}_i | \mathbf{z})$, we assume the independence of seismograms conditioned on the shared source information. For any two seismograms, indexed by k and l , of an earthquake, this independence can be expressed as follows:

$$P(\mathbf{x}_j^k, \mathbf{x}_j^l | \mathbf{s}, \mathbf{p}^k, \mathbf{p}^l) = P(\mathbf{x}_j^k | \mathbf{s}, \mathbf{p}^k)P(\mathbf{x}_j^l | \mathbf{s}, \mathbf{p}^l). \quad (9)$$

Intuitively, this indicates that, upon knowing the shared source information, the wavefield at one receiver does not inform the wavefield at another receiver. Consequently, for each source with index j , the likelihood can be ultimately factored as:

$$P(\mathbf{X}_j | \mathbf{z}, \phi) = \prod_k P(\mathbf{x}_j^k | \mathbf{s}, \mathbf{p}^k, \phi). \quad (10)$$

SymAE’s encoder generates the approximate posterior distribution $Q(\mathbf{z} | \mathbf{X}_j, \theta)$ of the latent variables given the input seismograms of the earthquake. As discussed previously, this is crucial because the true posterior can be mathematically intractable. Disentanglement in the latent space allows us to factorize this posterior distribution. To achieve factorization, we assume conditional independence during inference of latent variables. This means that

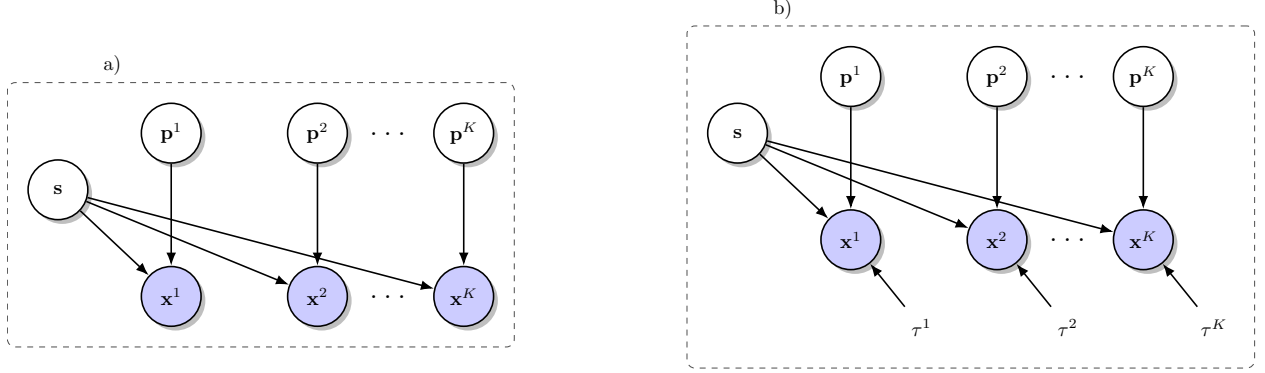


Figure 2: Generative models. (a) SymAE: this model assumes each waveform (denoted by \mathbf{x}^k) is generated from a common coherent component (represented by \mathbf{s}), along with waveform-specific nuisance components (represented by \mathbf{p}^k) that account for variations between waveforms. (b) SymAE with time-shift transformer: this enhanced model builds upon the regular SymAE by introducing a deterministic time-shift variable (denoted by τ^k). This variable allows for additional flexibility in capturing coherent information from waveforms affected by temporal shifts.

- the path effects are inferred independently using their respective seismograms

$$Q(\mathbf{p}^k, \mathbf{p}^l | \mathbf{X}_j, \theta) = Q(\mathbf{p}^k | \mathbf{X}_j, \theta)Q(\mathbf{p}^l | \mathbf{X}_j, \theta) = Q(\mathbf{p}^k | \mathbf{x}_j^k, \theta)Q(\mathbf{p}^l | \mathbf{x}_j^l, \theta); \quad (11)$$

- and when we consider a specific earthquake, the inference of source information \mathbf{s} is independent from that of the path information \mathbf{p}^k at any receiver

$$Q(\mathbf{s}, \mathbf{p}^k | \mathbf{X}_j, \theta) = Q(\mathbf{s} | \mathbf{X}_j, \theta)Q(\mathbf{p}^k | \mathbf{X}_j, \theta) = Q(\mathbf{s} | \mathbf{X}_j, \theta)Q(\mathbf{p}^k | \mathbf{x}_j^k, \theta). \quad (12)$$

By consolidating these assumptions, for each earthquake j , the approximate posterior distribution is given by

$$Q(\mathbf{z} | \mathbf{X}_j, \theta) = Q(\mathbf{s} | \mathbf{X}_j, \theta) \prod_k Q(\mathbf{p}^k | \mathbf{x}_j^k, \theta). \quad (13)$$

In the above equation, the term $Q(\mathbf{s} | \mathbf{X}_j, \theta)$ represents the coherent source information \mathbf{s} inferred by the accumulation of information across all seismograms, and $Q(\mathbf{p}^k | \mathbf{x}_j^k, \theta)$ is the path (nuisance) information derived from each seismogram. Finally, by substituting \mathbf{x}_i with \mathbf{X}_j in eq. 6, and by using the expressions for likelihood and posterior from equations 10 and 13, we derive the ELBO for SymAE which is optimized during the training:

$$\begin{aligned} \mathbb{L}(\theta, \phi) &= \sum_j \int Q(\mathbf{z} | \mathbf{X}_j, \theta) \log [P(\mathbf{X}_j | \mathbf{z}, \phi)] d\mathbf{z} - \sum_j D_{\text{KL}} (Q(\mathbf{z} | \mathbf{X}_j, \theta) \| P(\mathbf{z})) \\ &= \sum_j \sum_k \int Q(\mathbf{s} | \mathbf{X}_j, \theta)Q(\mathbf{p}^k | \mathbf{x}_j^k, \theta) \log [P(\mathbf{x}_j^k | \mathbf{s}, \mathbf{p}^k, \phi)] ds d\mathbf{p}^k \\ &\quad - \sum_j D_{\text{KL}} (Q(\mathbf{s} | \mathbf{X}_j, \theta) \| P(\mathbf{s})) \\ &\quad - \sum_j \sum_k D_{\text{KL}} (Q(\mathbf{p}^k | \mathbf{x}_j^k, \theta) \| P(\mathbf{p}^k)). \end{aligned} \quad (14)$$

In the above equation, the first term represents the reconstruction loss. During training, a Monte Carlo estimate is used for this term, and the next two terms are regularization terms.

3.1 Nuisance Encoder

The factored approximate posterior distribution $Q(\mathbf{z} | \mathbf{X}_j, \theta)$ in eq.13 allows us to employ separate encoder functions for each of the decomposed terms. These separate encoders operate on the input data \mathbf{X}_j to infer the individual latent variables that represent the source component (\mathbf{s}) and the path information components (\mathbf{p}^k) as depicted in Fig. 3. A separate convolutional network, denoted by \mathbf{g}_p with parameters θ_p , is employed to capture the path parameters specific to each seismogram. These parameters, denoted by \mathbf{p}^k for k th seismogram, account for variations or noise present in

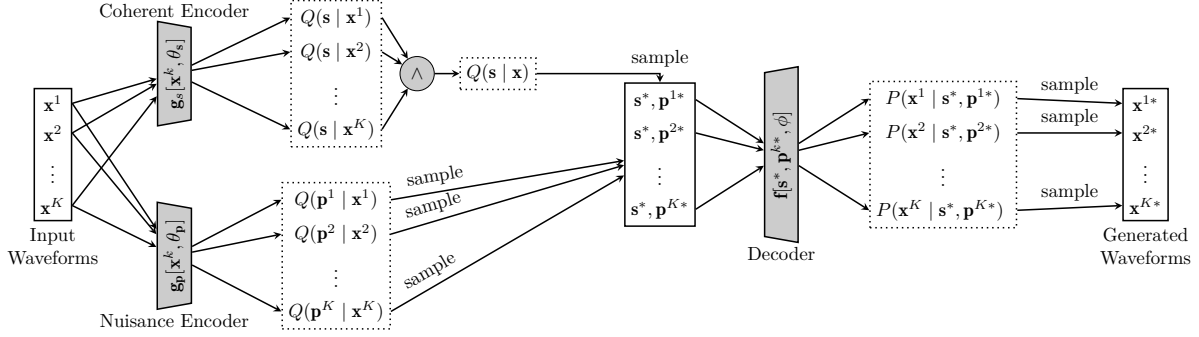


Figure 3: Disentangled representation learning for seismic waveforms using SymAE.

individual seismograms that are not part of the underlying coherent information. The network \mathbf{g}_p acts as a deterministic transformation, taking the seismogram \mathbf{x}_j^k and its own parameters θ_p as input. The output of this transformation defines the approximate posterior distribution $Q(\mathbf{p}^k | \mathbf{x}_j^k, \theta_p)$. This distribution represents the probability of the path parameters \mathbf{p}^k given the specific seismogram \mathbf{x}_j^k and the network parameters θ_p :

$$Q(\mathbf{p}^k | \mathbf{x}_j^k, \theta_p) = \mathcal{N}(\mathbf{p}^k | \mathbf{g}_p[\mathbf{x}_j^k, \theta_p]) \quad (15)$$

Again, \mathcal{N} denotes a normal distribution with mean and diagonal variance determined by the network output $\mathbf{g}_p[\mathbf{x}_i^k, \theta_p]$, where we denote the mean of this distribution using \mathbf{m}_j^k for future use.

3.2 Coherent Encoder: Accumulation of Coherent Information

The concept of combining information from multiple sources is crucial in various scientific fields. In Tarantola [2005] (example 1.12), an inverse problem is presented where independent measurements provide posterior probability distributions for a parameter. The combination of these independent probability distributions is achieved by first multiplying the individual distributions. Then, the resulting probability distribution is normalized to ensure that it sums up to one, representing the combined knowledge about the parameter after considering all measurements. This process, termed *conjunction* (denoted using \wedge) of information states, leverages the independence of the measurements, thereby accumulating the information provided by each distribution. It's important to remember that the process of conjunction used to combine the information from individual distributions is commutative (symmetry with respect to ordering).

Similarly, in SymAE, $Q(\mathbf{s} | \mathbf{X}_j, \theta)$ in eq. 13 represents the accumulated coherent information about the source parameter \mathbf{s} inferred from all seismograms in \mathbf{X}_j . Each seismogram is considered an independent observation that is informative about \mathbf{s} . The concept of conjunction is then used to accumulate this information across all seismograms:

$$Q(\mathbf{s} | \mathbf{X}_j, \theta) = Q(\mathbf{s} | \mathbf{x}_j^1, \theta) \wedge Q(\mathbf{s} | \mathbf{x}_j^2, \theta) \wedge \dots \propto \prod_k Q(\mathbf{s} | \mathbf{x}_j^k, \theta). \quad (16)$$

We model each $Q(\mathbf{s} | \mathbf{x}_j^k, \theta)$ as a Gaussian distribution with mean $\boldsymbol{\mu}_j^k$ and variance \mathbf{V}_j^k . These are determined by the output of the network \mathbf{g}_s with parameters θ_s , which takes the input seismogram \mathbf{x}_j^k as input: $Q(\mathbf{s} | \mathbf{x}_j^k, \theta_s) = \mathcal{N}(\mathbf{s} | \mathbf{g}_s[\mathbf{x}_j^k, \theta_s])$. Given that each individual posterior distribution $Q(\mathbf{s} | \mathbf{x}_j^k, \theta_s)$ is Gaussian with mean $\boldsymbol{\mu}_j^k$ and variance \mathbf{V}_j^k , it can be shown (see Bouchacourt et al. [2018]) that the accumulated distribution is also Gaussian, with mean ($\boldsymbol{\mu}_j$) and variance (\mathbf{V}_j) are derived using the following equations:

$$\mathbf{V}_j^{-1} = \sum_k (\mathbf{V}_j^k)^{-1}, \quad (17)$$

$$\boldsymbol{\mu}_j^T \mathbf{V}_j^{-1} = \sum_k (\boldsymbol{\mu}_j^k)^T (\mathbf{V}_j^k)^{-1}. \quad (18)$$

These equations combine the information from all individual distributions, effectively accumulating the knowledge about the source parameter \mathbf{s} across all seismograms. A crucial aspect of these equations lies in using the inverse (denoted by $^{-1}$) of variance, which represents the precision. Precision indicates how informative a distribution is. A

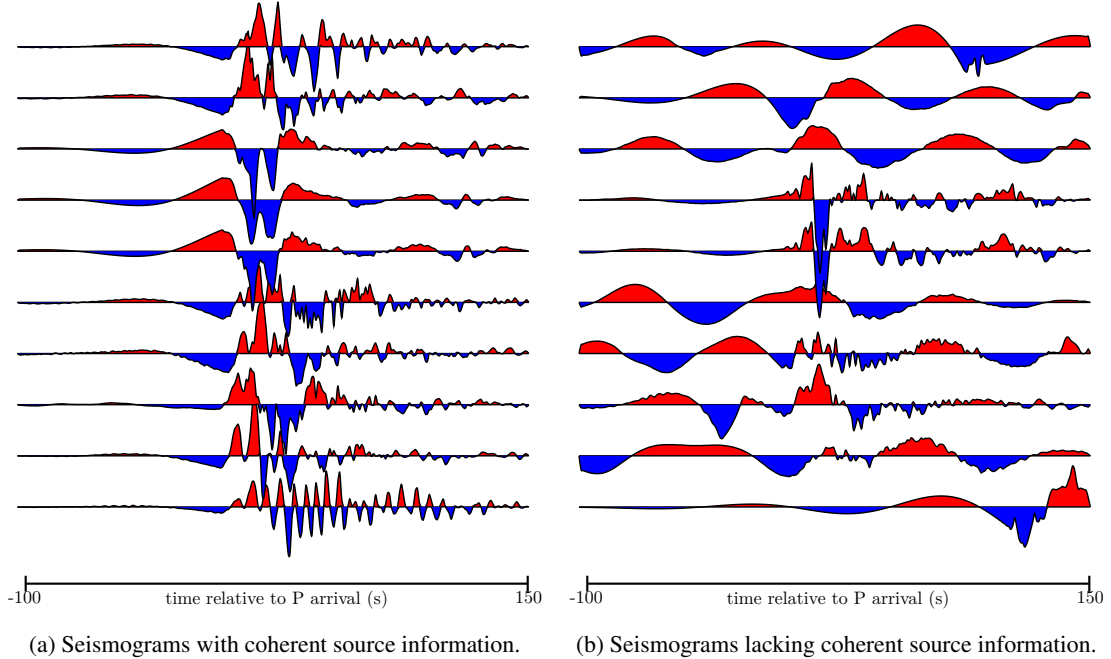


Figure 4: SymAE, applied here to the wavefield data from the 2013-05-24 Mw 8.3 Sea of Okhotsk earthquake, identifies seismograms that are more informative about the source by analyzing their contribution to the accumulated coherent information (see eq. 19). Notice that seismograms with a stronger contribution and lower influence from nuisance variations are considered more informative.

lower variance (higher precision) signifies a tighter distribution with more concentrated information around the mean. In contrast, a higher variance (lower precision) represents a broader and less informative distribution. The equations do not directly sum the variances because variance reflects the spread of the distribution, not its informativeness. Instead, they sum the inverses of the individual variances. This makes sense because precision tells us how much information each seismogram contributes. By adding these precisions, we effectively accumulate the total informative content from all observations. The second equation follows a similar logic for the mean. It calculates the product of the mean vector transposed μ_j^T with the inverse of the corresponding variance V_j^{-1} . This essentially weights the contribution of each mean based on the precision of its corresponding distribution. The sum of all seismograms then combines this weighted information. Seismograms with higher precision (more informative) have a greater influence on the final accumulated mean μ_j . This formulation allows the network to implicitly down-weight waveform features with high noise (high variance, low precision). These noisy features have a weaker influence on the accumulated mean μ_j . This mechanism generalizes non-linear stacking approaches where noisy samples receive lower weights. In this framework, the network effectively learns to discriminate between noisy samples and waveforms with good Signal-to-Noise Ratio (SNR), where *signal* refers to coherent information, alleviating the need for manual pre-processing steps.

3.3 Identifying Informative Seismograms

Our approach utilizes $Q(s | \mathbf{x}_j^k)$ to represent the knowledge about the coherent information obtained solely from the k -th seismogram. In contrast, $Q(s | \mathbf{X}_j)$ captures the accumulated coherent information derived from all seismograms of the earthquake. After training SymAE, we can identify which specific seismograms contribute most significantly to the accumulated source information. This is accomplished by calculating the Kullback-Leibler (KL) divergence between $Q(s | \mathbf{X}_j)$, which represents the accumulated distribution, and $Q(s | \mathbf{x}_j^k)$, the distribution for an individual seismogram \mathbf{x}_j^k :

$$H[\mathbf{x}_j^k] = D_{\text{KL}}(Q(s | \mathbf{X}_j) \parallel Q(s | \mathbf{x}_j^k)). \quad (19)$$

Since KL divergence measures the distance between two distributions, seismograms that contribute more significantly to the accumulated information will have lower H values. In simpler terms, waveforms with lower H values are less affected by irrelevant path or nuisance variations and hence contribute more to the overall understanding represented by $Q(s | \mathbf{X}_j)$. For instance, Fig. 4a displays seismograms that provide valuable source information of a deep earthquake, whereas Fig. 4b shows seismograms that lack earthquake source content; it is evident that the uninformative seismograms

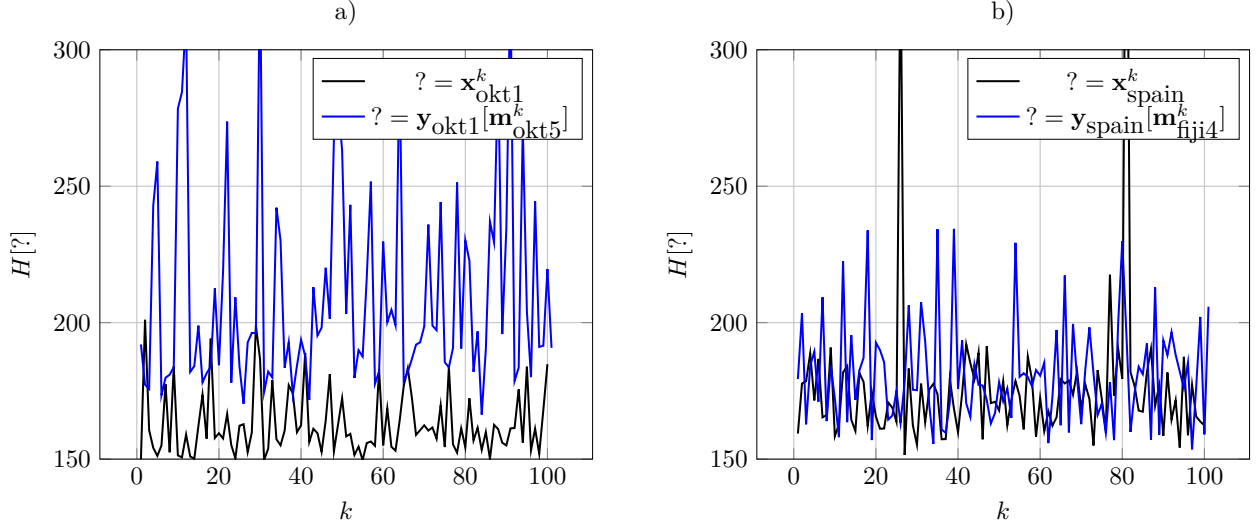


Figure 5: In the framework of source imaging, seismograms that exhibit higher H values (refer to eq. 19) are considered informative. a) Virtual seismograms of the Sea of Okhotsk earthquake (Mw 8.3, referred to as okt1) in blue, generated using the path effects of another Sea of Okhotsk earthquake (Mw 6.7, referred to as okt2), are found to be less informative than the actual recorded seismograms (black). This is because okt1 and okt2 have different source mechanisms; okt2 involves a single rupture, while okt1 consists of three distinct ruptures. b) Similar to (a) but displaying the H values for measured and virtual seismograms of the Spain earthquake (Mw 6.3, labeled spain). In this case, both the measured and virtual seismograms (using path effects from a Mw 6.8 earthquake of Fiji region) show comparable H values, indicating they are equally informative. This similarity suggests that both earthquakes share characteristics.

exhibit noise, meaning they are influenced by irrelevant variations. We achieved these results by training the SymAE on a collection of high-magnitude earthquake P-wave teleseismic displacement waveforms of earthquakes listed in Tab. 1. The specific details of this dataset and training process are presented later in the results section.

3.4 Redatuming and Conditional Generation of Virtual Seismograms

A key advantage of the latent space disentanglement achieved in SymAE is the ability to perform conditional generation of seismograms. In conditional generation, we can control the generation process by fixing the latent variables associated with the source information. By doing so, we can generate a collection of virtual seismograms (ensemble) that share the same underlying source characteristics while varying in their path effects. We denote virtual or synthesized seismograms $\mathbf{y}_j[\mathbf{m}_i^k]$, which is the mean of the Gaussian distribution, generated by the decoder \mathbf{f} of the trained SymAE model, given by $\mathbf{f}[\boldsymbol{\mu}_j, \mathbf{m}_i^k, \phi]$. Here, $\boldsymbol{\mu}_j$ represents the mean source information for the j th earthquake, obtained using the coherent encoder \mathbf{g}_s that yields $Q(\mathbf{s} | \mathbf{X}_j)$ using seismograms \mathbf{X}_j . The mean path information from \mathbf{x}_i^k , the k th seismogram of the i th earthquake, is denoted using \mathbf{m}_i^k . It is derived from the encoder \mathbf{g}_p which produces $Q(\mathbf{p}^k | \mathbf{x}_i^k)$ using \mathbf{x}_i^k . In seismic data processing, a key technique called redatuming aims to virtually reposition recorded seismic wavefields to a different reference level (datum). This is often done to simplify the interpretation of features within the Earth’s subsurface. In the context of SymAE, we can leverage the disentangled representation to perform a specific type of redatuming: swapping the path information between different seismograms to synthesize virtual seismograms.

Virtual seismograms \mathbf{y}_j generated by combining the source information of the j th earthquake with the path information of a different earthquake can provide insight into the role of path effects that obscure the interesting source information. For example, a clear seismogram (less affected by path effects and other noises) obtained from one earthquake could assist in producing a clear seismogram for another earthquake, potentially of smaller magnitude. However, it is crucial to assess their overall informativeness. To evaluate the informativeness of a virtual seismogram, we use the eq. 19 metric based on the KL divergence:

$$H[\mathbf{y}_j[\mathbf{m}_i^k]] = D_{\text{KL}}(Q(\mathbf{s} | \mathbf{X}_j) \parallel Q(\mathbf{s} | \mathbf{y}_j[\mathbf{m}_i^k])). \quad (20)$$

Here, $Q(\mathbf{s} | \mathbf{y}_j[\mathbf{m}_i^k])$ represents the source information in the virtual seismogram, $H[\mathbf{y}_j[\mathbf{m}_i^k]]$ quantifies the information loss when replacing the group of real seismograms \mathbf{X}_j with the single virtual seismogram $\mathbf{y}_j[\mathbf{m}_i^k]$ in terms of the source information. A lower KL divergence indicates a better preservation of source information in the virtual

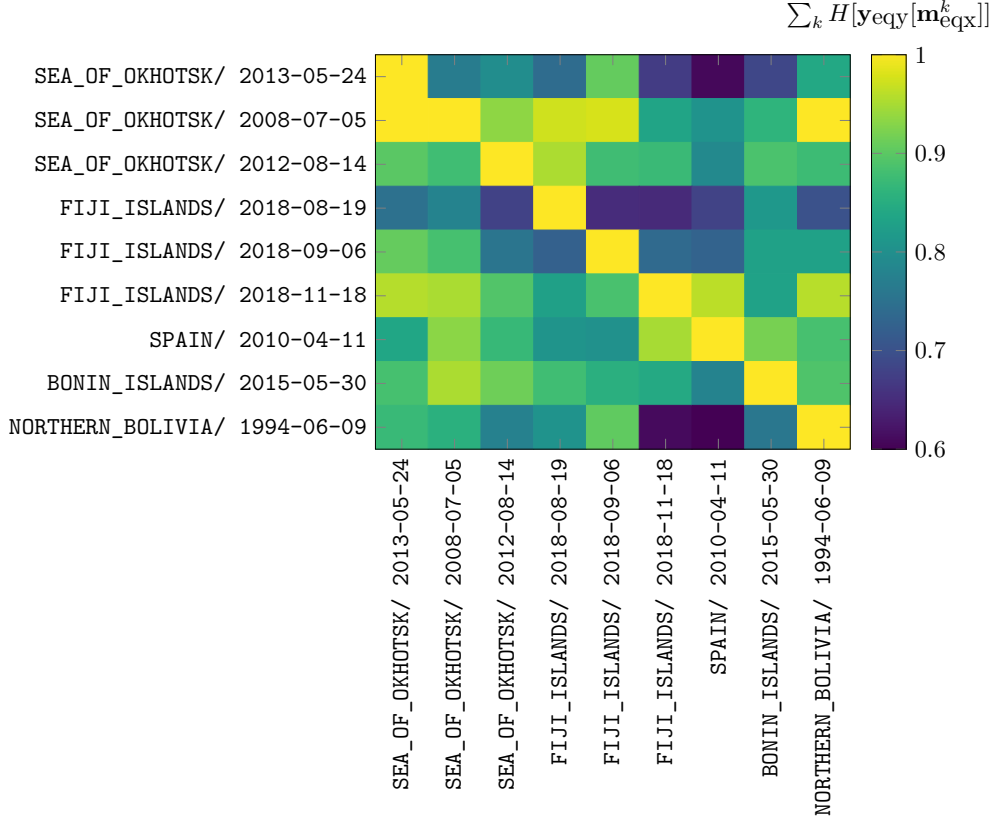


Figure 6: Heatmap illustrating the average information loss when generating virtual seismograms by combining source information from y-axis earthquakes with path information from x-axis earthquakes. Lighter colors indicate lower information loss, suggesting greater similarity between the source characteristics of the respective earthquakes.

seismogram. We observed a correlation between the informativeness of generated virtual seismograms and the similarity of source earthquakes. Redatuming between similar earthquakes generally results in lower information loss compared to redatuming between dissimilar events. This suggests that earthquakes with significantly different rupture complexities, such as those with varying numbers or characteristics of seismic subevents, may exhibit greater information loss when their path effects are swapped. Fig. 5 illustrates this phenomenon using two deep earthquakes as examples. Here, virtual seismograms generated by swapping path effects between two earthquakes even within the same tectonic setting (e.g., the Sea of Okhotsk), as in Fig. 5a, exhibited higher information loss. In the case of Fig. 5b with Spain and Fiji4 earthquakes in Tab. 1, the information loss was minimal, suggesting a resemblance between earthquakes, even if they originated from distinct tectonic settings.

To assess information loss across multiple earthquake pairs, we averaged the KL divergence $H[y_j[\mathbf{m}_i^k]]$ over all seismograms for each pair and visualized the results in a heatmap Fig. 6, where higher values after normalization indicate greater similarity between the source characteristics of the respective earthquakes. For instance, earthquakes fiji1 and fiji2, known for their complex rupture processes with multiple subevents [Jia et al., 2020], exhibited higher information loss when paired with other deep earthquakes in Tab. 1. In contrast, the earthquakes in Spain and Fiji4 showed a lower loss of information, suggesting potential similarities in their source characteristics. Similarly, our analysis revealed similarities between earthquakes okt1 and bon1, while demonstrating a clear dissimilarity between bon1 and Spain. We will further validate these observations by analyzing the extracted coherent source information in the results section.

3.5 Optimal Virtual Seismograms

We have shown that SymAE not only has the ability to generate virtual seismograms, but it can also effectively identify informative seismograms (eqs. 19 and 20). Taking advantage of this dual functionality, we aim to generate virtual seismograms that are enriched with task-relevant source information while simultaneously minimizing the influence

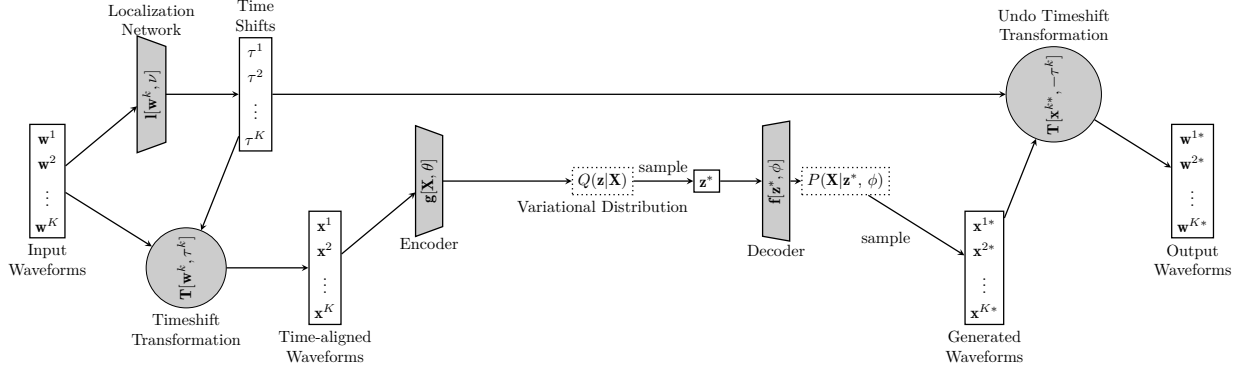


Figure 7: Architecture of the SymAE with time-shift transformer. The model incorporates a localization network to estimate optimal time shifts for input waveforms, followed by the core SymAE components for latent-space representation and reconstruction. An inverse time-shift operation is applied to the reconstructed waveforms to align them with the original waveforms.

of path factors. Essentially, we used a trained SymAE model to create a virtual seismogram that best represents the underlying source characteristics, effectively filtering out noise and irrelevant information that may be present in real-world data. The generation of this virtual seismogram translates into an optimization problem, where we search the space of possible nuisance codes (denoted by \mathbf{m}) to achieve this goal. This optimization problem can be formulated as follows:

$$\begin{aligned} \mathbf{m}_j &= \arg \min_{\mathbf{m}} D_{\text{KL}}(Q(\mathbf{s} | \mathbf{X}_j) \| Q(\mathbf{s} | \mathbf{y}_j[\mathbf{m}])) + \alpha \|\mathbf{y}_j[\mathbf{m}]\|_1, \\ &\text{subject to } \sum_{n=1}^D (\mathbf{y}_j[\mathbf{m}])_n = 0 \quad \text{and} \quad \frac{1}{D} \sum_{n=1}^D \left((\mathbf{y}_j[\mathbf{m}])_n - \frac{1}{D} \sum_n (\mathbf{y}_j[\mathbf{m}])_n \right)^2 = 1. \end{aligned} \quad (21)$$

The solution we seek is the optimal nuisance code, denoted by \mathbf{m} , corresponding to the j th earthquake. By manipulating this code within the optimization framework, we influence the characteristics of the generated virtual seismogram $\mathbf{y}_j[\mathbf{m}]$, which is the mean of decoder output $\mathbf{f}[\mu_j, \mathbf{m}]$. A key term in the objective function is the KL divergence, denoted by $D_{\text{KL}}(Q(\mathbf{s} | \mathbf{X}_j) \| Q(\mathbf{s} | \mathbf{y}_j[\mathbf{m}]))$. This term measures the difference between the information about the coherent source (represented by \mathbf{s}) extracted from all measured seismograms of the earthquake $Q(\mathbf{s} | \mathbf{X}_j)$ and the information obtained from the generated virtual seismogram $Q(\mathbf{s} | \mathbf{y}_j[\mathbf{m}])$. Minimizing this divergence ensures that the virtual seismogram retains the coherent information of the earthquake. Another crucial term promotes sparsity in the virtual seismogram. This term involves the ℓ^1 -norm of the predicted virtual seismogram. A higher ℓ^1 -norm penalty encourages a virtual seismogram with fewer non-zero values, potentially leading to a simpler representation with less influence from nuisance variations. The parameter α controls the weight given to this sparsity term, allowing us to adjust the trade-off between minimizing KL divergence and achieving a sparse virtual seismogram. Other relevant constraints, denoted by “subject to” in the formulation, include enforcing a zero mean for the virtual seismogram. This translates to requiring the sum of all elements in the D -dimensional vector $\mathbf{y} = (y_1, \dots, y_D)^T$ to equal zero. The second constraint ensures unit variance for the virtual seismogram. These constraints align the generated virtual seismogram with real seismograms, which exhibit zero mean and unit variance characteristics after preprocessing. In the results section, we show that generating a virtual seismogram for each earthquake effectively preserves important source information while reducing the impact of noise and other irrelevant variations.

4 Symmetric Autoencoder with Time-shift Transformer

Seismograms often exhibit variations in arrival times due to factors such as propagation path differences. This can hinder the effective application of the SymAE architecture discussed in the previous section, which has difficulty compressing time-shifted waveforms into a low-dimensional latent space. To address this challenge, we introduce a timeshift transformer module that aligns waveforms before encoding. We show that by learning optimal time shifts, the model effectively extracts underlying coherent information and generates accurate reconstructions.

The architecture of SymAE with time-shift transformer is illustrated in Fig. 7, while its corresponding generative model is shown in Fig. 2b, where time shifts are treated as deterministic variables. To account for potential time shifts in input

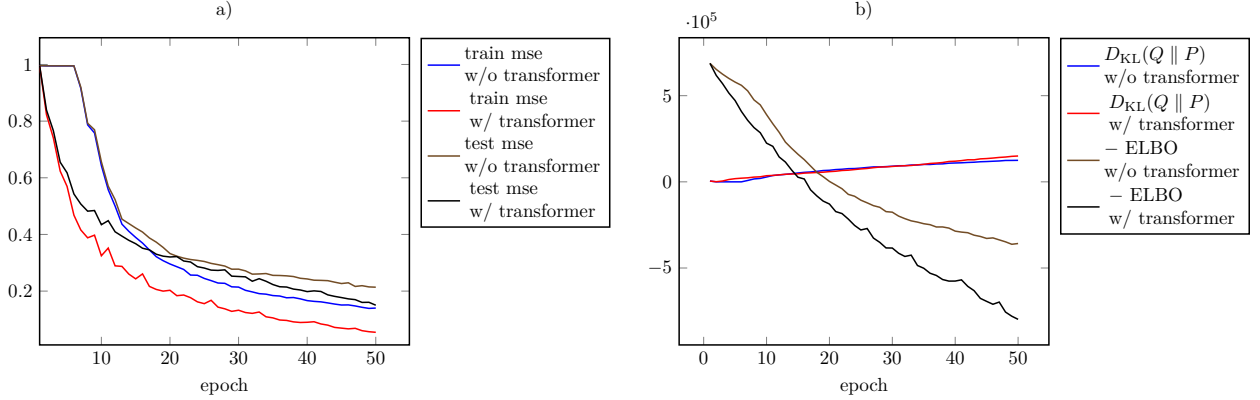


Figure 8: ELBO and KL divergence $D_{KL}(Q \parallel P)$ during SymAE training with and without time-shift transformer using noiseless seismograms (Figs. 1e and 1f). The time-shift transformer SymAE model exhibits lower reconstruction loss (mse).

waveforms, we introduce the notation $\mathbf{W}_j = \{\mathbf{w}_j^k\}$ to represent time-shifted seismograms. A localization network \mathbf{I} with parameters ν initially processes these waveforms to determine optimal time shifts $\tau_j^k = \mathbf{I}[\mathbf{w}_j^k, \nu]$, which are then applied to generate time-aligned versions $\mathbf{X}_j = \{\mathbf{x}_j^k\} = \{\mathbf{T}[\mathbf{w}_j^k, \tau_j^k]\}$. Here, \mathbf{T} denotes the time-shift operation, which we implement in the frequency domain using the Fourier-shift property. The core SymAE model subsequently operates on these aligned waveforms to learn a disentangled latent space representation. The loss function, adapted from the standard VAE formulation in Eq. 6, incorporates a reconstruction term, a KL divergence regularizer, and a time-shift regularization term:

$$\begin{aligned} \mathbb{L}(\theta, \phi, \nu) = & \sum_j \int Q(\mathbf{z} | \mathbf{W}_j, \theta) \log [P(\mathbf{W}_j | \mathbf{z}, \phi)] d\mathbf{z} - \sum_j D_{KL}(Q(\mathbf{z} | \mathbf{W}_j, \theta) \parallel P(\mathbf{z})) \\ & + \gamma \sum_j \sum_k (\tau_j^k)^2, \quad \text{where } \tau_j^k = \mathbf{I}[\mathbf{w}_j^k, \nu] \quad \text{and} \quad \mathbf{x}_j^k = \mathbf{T}[\mathbf{w}_j^k, \tau_j^k] \end{aligned} \quad (22)$$

An inverse time-shift operation using $-\tau_j^k$ is applied to the decoder-produced waveforms \mathbf{f} . The mean of the likelihood $P(\mathbf{W}_j | \mathbf{z}, \phi)$ results from this inverse time-shift operation on the decoder output $\mathbf{f}[\mathbf{z}, \phi]$. Whereas, the posterior $Q(\mathbf{z} | \mathbf{W}_j, \theta)$ is obtained by applying the encoder after time-shifting each input waveform using the corresponding τ_j^k . The time-shift regularization term penalizes excessive time shifts, ensuring stability and interpretability of the model. The loss function is minimized through backpropagation to optimize the model parameters ϕ, θ and ν .

5 Results

To evaluate the proposed SymAE model, synthetic seismograms incorporating noise and time shifts were generated (Fig. 1). The SymAE model was trained using the ADAM optimizer [Kingma and Ba, 2014]. Fig. 8 illustrates the ELBO loss during training for standard and time-shift transformer SymAE models using noiseless data. The latter exhibited lower reconstruction loss, and the extracted coherent source information (Figs. 9a and 9c) closely matched the ground-truth source signatures in both cases. However, in the presence of noise, the standard SymAE model performed poorly, as evidenced by the ELBO plot in Fig. 10. In contrast, the time-shift transformer SymAE demonstrated superior performance, effectively extracting coherent source information, as plotted in Fig. 9d.

Fig. 11 illustrates the effectiveness of the proposed SymAE method in extracting coherent source information of deep earthquakes (Tab. 1) compared to traditional deconvolution techniques [Vallée et al., 2011]. Here, SymAE is trained on roughly 5000 displacement seismograms (all components) for each earthquake. The SymAE-derived source time function exhibits a clearer and more detailed representation of the earthquake rupture process, capturing complex features due to multiple events that are often obscured and smoothed out in traditional deconvolution approaches. This improvement is particularly evident in the case of earthquakes like bon1, fiji4, and spain. The half-durations of the source functions extracted using SymAE closely align with those determined from the raw seismograms. Importantly, the extracted source-function similarities align closely with the relationships previously observed in the information loss heatmap in Fig. 6. This further validates the effectiveness of our proposed optimization approach in extracting the underlying characteristics of the earthquakes.

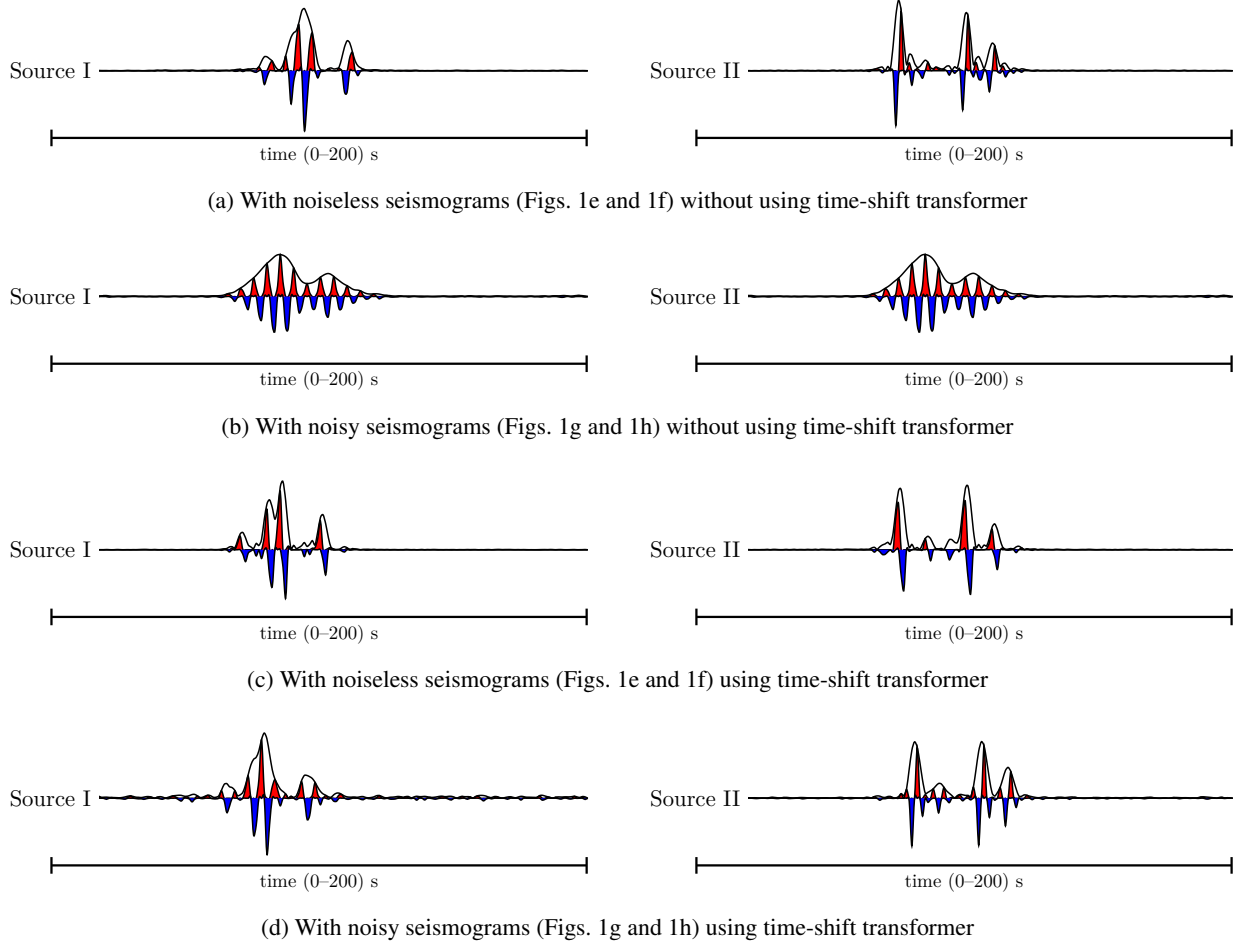


Figure 9: SymAE extracted coherent source information through latent space optimization for the synthetic experiment. True sources I and II are plotted in Fig. 1a and Fig. 1b, respectively.

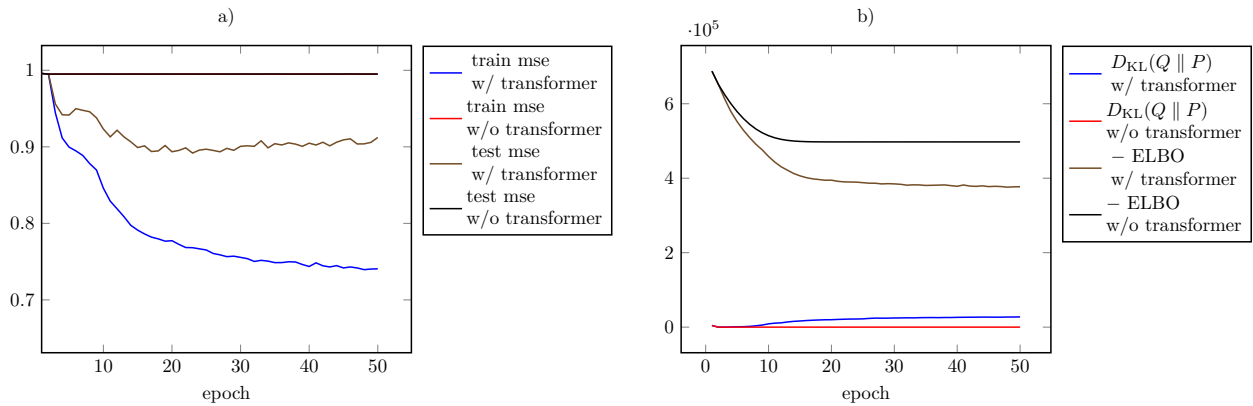


Figure 10: Same as Fig. 8, but using noisy seismograms in Figs. 1g and 1h.

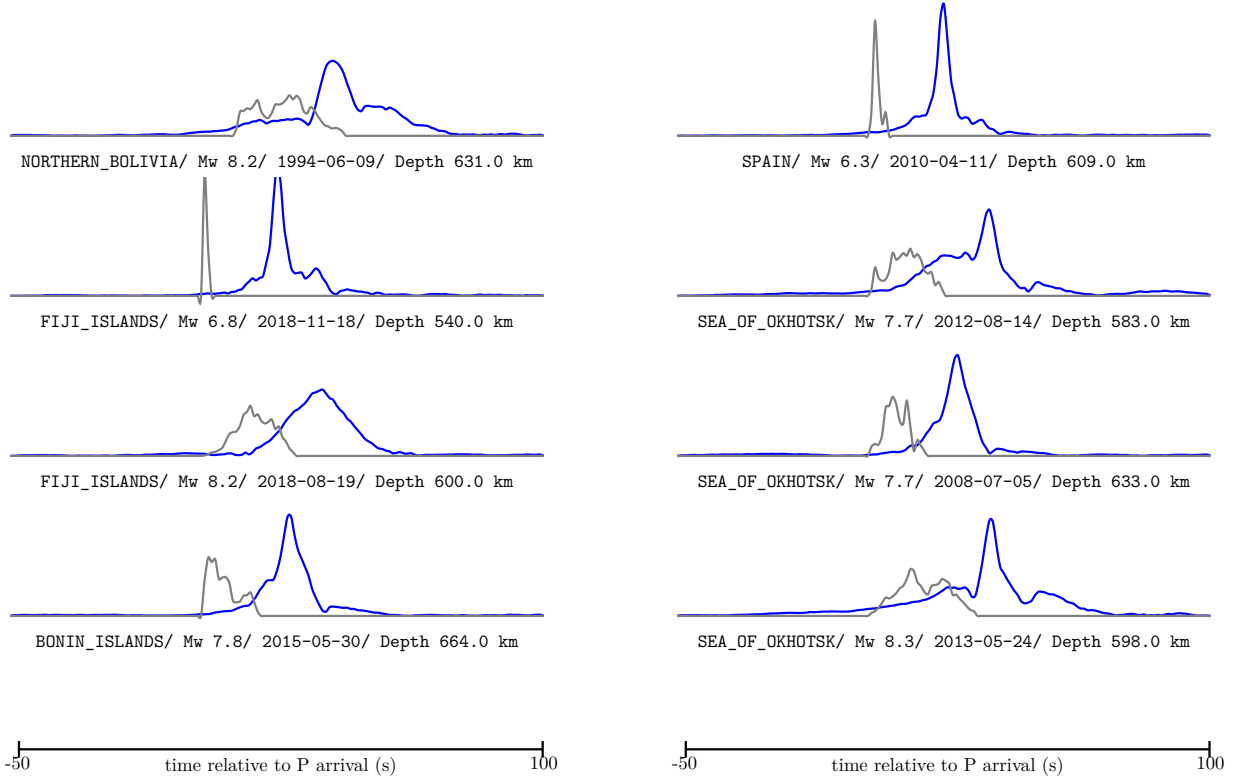


Figure 11: Comparison of coherent source information extracted using SymAE (blue) for complex deep earthquakes compared to traditional deconvolution techniques (gray).

6 Conclusions

This study introduces a novel approach to extracting coherent information from seismic wavefields using a Symmetric Autoencoder (SymAE) with a time-shift transformer. By disentangling the latent space into source and path components, the proposed method effectively isolates coherent seismic signals, reducing the impact of noise and other unwanted variations. The ability to quantify waveform informativeness is a key contribution, enabling the identification of data that significantly contributes to the overall coherent signal. Our experiments demonstrate the effectiveness of SymAE in extracting coherent source information, particularly in the presence of noise and complex seismic phenomena. The generated virtual seismograms provide valuable insights into the underlying earthquake processes. The proposed framework holds promise for a wide range of seismological applications that require the extraction of coherent signals from noisy data.

Future research could explore the application of SymAE to other seismic datasets, such as ambient noise tomography. Additionally, incorporating more complex physical models into the SymAE framework could lead to further improvements in the extraction of coherent information. We emphasize the importance of evaluating generative models, such as SymAE, within the context of their intended applications. We argue that a model’s performance should be judged on the basis of its ability to produce data that are physical and valuable for specific seismological tasks, as generic metrics may not capture the nuances required for these applications. Theis et al. [2016] recommended to use evaluation strategies that are tailored to the specific applications of generative models.

7 Acknowledgments

This work is funded by the Science and Engineering Research Board, Department of Science and Technology, India (Grant Number SRG/2021/000205). The authors would like to express their sincere gratitude to Matt Li and Laurent Demanet for their valuable insights on variational autoencoders, and to Sanket Bajad and Tiente Rengneichuonng for discussions. We also extend our thanks to Isha Lohan and Madhusudan Sharma for their contributions to earthquake data preprocessing. Our research was significantly aided by the utilization of the Flux.jl deep learning framework, which

enabled us to implement and train the SymAE model. We also acknowledge the utilization of the Julia programming language, specifically version 1.9.0, as well as Pluto notebooks, which played a pivotal role in the development of our numerical simulations and data analysis presented throughout this work.

8 Data Availability

All data and codes used in this study are open access. The seismological data is primarily sourced from the Incorporated Research Institutions for Seismology Data Management Center (IRIS-DMC), where the data was acquired from an array of institutions and seismic data centers, including AUPASS, BGR, EMSC, ETH, GEOFON, GEONET, GFZ, ICGC, IESDMC, INGV, IPGP, IRIS, IRISPH5, ISC, KNMI, KOERI, LMU, NCEDC, NIEP, NOA, ODC, ORFEUS, RASPISHAKE, RESIF, RESIFPH5, SCEDC, TEXNET, UIB-NORSAR, USGS, and USP.

References

- Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.
- Diederik P Kingma and Max Welling. An introduction to variational autoencoders. *Foundations and Trends in Machine Learning*, 12(4):307–392, 2019.
- Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.
- Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International conference on machine learning*, pages 1530–1538. PMLR, 2015.
- Suihong Song, Tapan Mukerji, and Jiagen Hou. Bridging the gap between geophysics and geology with generative adversarial networks. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–11, 2021.
- Yuan Xiao, Kewen Li, Yimin Dou, Wentao Li, Zhixuan Yang, and Xinyuan Zhu. Diffusion models for multidimensional seismic noise attenuation and super-resolution. *Geophysics*, 89(5):1–66, 2024.
- Daniele Trappolini, Laura Laurenti, Giulio Poggiali, Elisa Tinti, Fabio Galasso, Alberto Michellini, and Chris Marone. Cold diffusion model for seismic denoising. *Journal of Geophysical Research: Machine Learning and Computation*, 1(2):e2024JH000179, 2024.
- Pu Ren, Rie Nakata, Maxime Lacour, Ilan Naiman, Nori Nakata, Jialin Song, Zhengfa Bi, Osman Asif Malik, Dmitriy Morozov, Omri Azencot, et al. Learning physics for unveiling hidden earthquake ground motions via conditional generative modeling. *arXiv preprint arXiv:2407.15089*, 2024.
- Changxiao Sun, Alison Malcolm, Rajiv Kumar, and Weijian Mao. Enabling uncertainty quantification in a standard full waveform inversion method using normalizing flows. *Geophysics*, 89(5):1–69, 2024.
- Martin Schimmel, E Stutzmann, and J Gallart. Using instantaneous phase coherence for signal extraction from ambient noise data at a local to a global scale. *Geophysical Journal International*, 184(1):494–506, 2011.
- Gerrit Olivier, Florent Brenguier, Michel Campillo, Richard Lynch, and Philippe Roux. Body-wave reconstruction from ambient seismic noise correlations in an underground mine. *Geophysics*, 80(3):KS11–KS25, 2015.
- Claus-Peter Rückemann. Comparison of stacking methods regarding processing and computing of geoscientific depth data. In *Proceedings of the Fourth International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2012)*, volume 30, pages 35–40, 2012.
- Richard L Weaver and John Y Yoritomo. Temporally weighting a time varying noise field to improve green function retrieval. *The Journal of the Acoustical Society of America*, 143(6):3706–3719, 2018.
- Jinyun Xie, Yingjie Yang, and Yinhe Luo. Improving cross-correlations of ambient noise using an rms-ratio selection stacking method. *Geophysical Journal International*, 222(2):989–1002, 2020.
- Diane Bouchacourt, Ryota Tomioka, and Sebastian Nowozin. Multi-level variational autoencoder: Learning disentangled representations from grouped observations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.
- Pawan Bharadwaj, Matthew Li, and Laurent Demanet. Redatuming physical systems using symmetric autoencoders. *Physical Review Research*, 4(2):023118, 2022.
- Max Jaderberg, Karen Simonyan, Andrew Zisserman, and Koray Kavukcuoglu. Spatial transformer networks. In *Advances in neural information processing systems*, pages 2017–2025, 2015.

- SM Ali Eslami, Danilo Jimenez Rezende, Frederic Besse, Fabio Viola, Ari Morcos, Marta Garnelo, Avraham Ruderman, Andrei A Rusu, Ivo Danihelka, Karol Gregor, et al. Attend, infer, repeat: Fast scene understanding with generative models. In *Advances in neural information processing systems*, pages 3225–3233, 2016.
- Adam M Dziewonski, T-A Chou, and John H Woodhouse. Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *Journal of Geophysical Research: Solid Earth*, 86(B4): 2825–2852, 1981.
- Stuart A Sipkin. Estimation of earthquake source parameters by the inversion of waveform data: synthetic waveforms. *Physics of the Earth and Planetary interiors*, 30(2-3):242–259, 1982.
- Rachel E Abercrombie. Investigating uncertainties in empirical green’s function analysis of earthquake source parameters. *Journal of Geophysical Research: Solid Earth*, 120(6):4263–4277, 2015.
- G. Zandt and C. J. Ammon. Continental crust composition constrained by measurements of crustal poisson’s ratio. *Nature*, 374(6518):152–154, March 1995. doi:10.1038/374152a0.
- T. Zheng, Z. Ding, J. Ning, L. Chang, X. Wang, F. Kong, and et al. Crustal azimuthal anisotropy beneath the southeastern tibetan plateau and its geodynamic implications. *Journal of Geophysical Research: Solid Earth*, 123:9733–9749, 2018. doi:10.1029/2018JB015995.
- B. Dalai, P. Kumar, U. Srinu, and M. K. Sen. De-noising receiver function data using the unsupervised deep learning approach. *Geophysical Journal International*, 229(2):737–749, December 2021. doi:10.1093/gji/ggab494.
- W. Bloch, M. G. Bostock, and P. Audet. A cascadia slab model from receiver functions. *Geochemistry, Geophysics, Geosystems*, 24(10):e2023GC011088, 2023. doi:10.1029/2023GC011088.
- Jiafu Hu, Haiyan Yang, Guangquan Li, and Hengchu Peng. A review on the analysis of the crustal and upper mantle structure using receiver functions. *Journal of Asian Earth Sciences*, 111:589–603, 2015. ISSN 1367-9120. doi:10.1016/j.jseas.2015.06.007.
- RL Weaver. Lobkis 0 i. *Ultrasonic without a sources Thermal fluctuation correlations at MHz frequencies Physical Review Letters*, 87(13):134301, 2001.
- Richard L Weaver. Information from seismic noise. *Science*, 307(5715):1568–1569, 2005.
- Michel Campillo and Anne Paul. Long-range correlations in the diffuse seismic coda. *Science*, 299(5606):547–549, 2003.
- Kees Wapenaar. Retrieving the elastodynamic green’s function of an arbitrary inhomogeneous medium by cross correlation. *Physical review letters*, 93(25):254301, 2004.
- Roel Snieder. Extracting the green’s function from the correlation of coda waves: A derivation based on stationary phase. *Physical review E*, 69(4):046610, 2004.
- Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- Simon JD Prince. *Understanding Deep Learning*. MIT press, 2023.
- Christopher M. Bishop. *Deep Learning*. Springer, New York, NY, 2023. ISBN 978-3-030-12345-6.
- Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. In *International Conference on Learning Representations*, 2017.
- Christopher P Burgess, Irina Higgins, Arka Pal, Loic Matthey, Nick Watters, Guillaume Desjardins, and Alexander Lerchner. Understanding disentangling in beta-vae. *arXiv preprint arXiv:1804.03599*, 2018.
- Albert Tarantola. *Inverse problem theory and methods for model parameter estimation*. SIAM, 2005.
- Zhe Jia, Zhichao Shen, Zhongwen Zhan, Chenyu Li, Zhigang Peng, and Michael Gurnis. The 2018 fiji mw 8.2 and 7.9 deep earthquakes: One doublet in two slabs. *Earth and Planetary Science Letters*, 531:115997, 2020. ISSN 0012-821X. doi:https://doi.org/10.1016/j.epsl.2019.115997. URL https://www.sciencedirect.com/science/article/pii/S0012821X19306892.
- Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- Martin Vallée, Jean Charléty, Ana MG Ferreira, Bertrand Delouis, and Julien Vergoz. Scardec: a new technique for the rapid determination of seismic moment magnitude, focal mechanism and source time functions for large earthquakes using body-wave deconvolution. *Geophysical Journal International*, 184(1):338–358, 2011.
- Lucas Theis, Aäron van den Oord, and Matthias Bethge. A note on the evaluation of generative models. In *International Conference on Learning Representations (ICLR)*, 2016.