LEARNING EARTHQUAKE SOURCES USING SYMMETRIC AUTOENCODERS

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

Isha Lohan
Centre for Earth Sciences
Indian Institute of Science
Bengaluru
lohanisha@gmail.com

Madhusudan Sharma
Centre for Earth Sciences
Indian Institute of Science
Bengaluru
md24sharma@gmail.com

Abhinav Pratap Singh
Centre for Earth Sciences
Indian Institute of Science
Bengaluru
abhinavps@gatech.edu

November 14, 2023

ABSTRACT

We introduce Symmetric Autoencoder (SymAE), a neural-network architecture designed to automatically extract earthquake information from far-field seismic waves. SymAE represents the measured displacement field using a code that is partitioned into two interpretable components: source and path-scattering information. We achieve this source-path representation using the scale separation principle and stochastic regularization, which traditional autoencoding methods lack. According to the scale separation principle, the variations in far-field band-limited seismic measurements resulting from finite faulting occur across two spatial scales: a slower scale associated with the source processes and a faster scale corresponding to path effects. Once trained, SymAE enables the generation of virtual seismograms engineered to incorporate subsurface scattering effects from other seismograms. In this paper, we first validate the accuracy of these virtual seismograms generated by SymAE using both P and S waveforms. We then demonstrate the use of virtual seismograms for estimating the similarity between collocated earthquakes. Additionally, we introduce the concept of uniformly scattered virtual seismograms, where seismograms associated with different focal-sphere pixels are engineered to contain identical path effects. These uniformly scattered virtual seismograms can be analyzed to study source directivity effects with greater detail and accuracy compared to traditional envelope stacking approaches. SymAE is an unsupervised learning method that can efficiently scale with large amounts of seismic data and does not require labeled seismograms, making it the first framework that can learn from all available previous earthquakes to accurately characterize a given earthquake. The paper presents the analysis of nearly forty complex earthquake events, revealing differences between earthquakes in energy rise times, stopping phases, and providing insights into their rupture complexity.

1 Key Points

- Novel network-based approach for achieving a source-path representation of seismograms.

- Our network characterizes complex earthquakes by generating virtual seismograms.

- We present a method to estimate teleseismic similarity between earthquakes.

2 Plain Language Summary

Seismologists use seismograms, recordings of ground movements, to study earthquakes. However, these recordings can be distorted by factors like the Earth's composition. To truly understand earthquakes, we must distinguish the earthquake's own characteristics (source effects) from the impact of the ground it passed through (path effects).

While conventional methods typically employ convolutional models to differentiate source and path effects, our approach relies on Symmetric Autoencoding (SymAE). SymAE enables us to examine the spatial patterns in earthquake data and effectively separate slow-scale source effects from fast-scale path effects. This separation empowers us to accurately determine earthquake source parameters, free from the influence of path effects.

Our method not only provides insight into earthquake source parameters but also allows for the quantification of similarity between earthquakes originating from the same source region. Furthermore, our framework introduces concept of generating *virtual seismograms*, which greatly enhance our ability to understand earthquake source characteristics. This capability proves invaluable, especially in cases where traditional methods may fall short in delivering precise insights.

3 Introduction

In this paper, we restrict the analysis to the far-field seismic waves generated by earthquake sources, which are modeled as sudden slips on finite faults. Traditional earthquake quantification comprises of parameters whose inference is less affected by the subsurface heterogeneity. For example, 1. the polarity of the first arrival is used to infer the focal mechanism i.e., mean seismic moment density tensor [Dziewonski et al., 1981, Ekström et al., 2012]; 2. the low-frequency level of the displacement spectra is used to estimate the moment magnitude [Kanamori and Brodsky, 2004]; 3. the duration of P-wave arrivals to perform directivity analysis and infer rupture velocity [Park and Ishii, 2015, e.g.,]; 4. mean source spectra estimation by averaging over all the available stations [Prieto et al., 2004, e.g.,]. The advancement of our comprehension of earthquake sources heavily relies on methods capable of separating source and path effects. The inference of earthquake parameters is highly uncertain due to the complexity of disentangling source and path effects from seismic measurements. Notably, the source information of interest, especially at higher temporal frequencies, tends to be obscured as seismic waves propagate along the path between the earthquake source and the receivers. Our current ability to extract the earthquake source information from these waves is limited by the validity of a source-path convolution model [Langston and Helmberger, 1975, Andrews, 1986]

$$d(\mathbf{x}, \omega) \approx G(\mathbf{x}, \omega)\, s(\mathbf{x}, \omega), \tag{1}$$

where the measured displacement d at spatial location $\mathbf{x}$ is approximately given by a multiplication in the frequency domain (convolution in time) between an apparent source function s and a path-dependent Green's function G. We used ω to denote temporal frequency. The convolution model is limited; although it is conveniently assumed by almost all the existing methods to extract the source function via simpler deconvolution algorithms. This model ignores the sensitivity of the apparent source function to the type of seismic arrival, e.g., P, pP, PP, or sP [Bormann et al., 2013]. Contrarily, in fact a seismogram can be more accurately modeled after decomposing Green's function into e.g., $G \approx G_{\mathrm{P}} + G_{\mathrm{pP}} + G_{\mathrm{sP}}$, using the high-frequency approximation to write

$$d(\mathbf{x}, \omega) \approx G_{\mathrm{P}}(\mathbf{x}, \omega)\, s_{\mathrm{P}}(\mathbf{x}, \omega) + G_{\mathrm{pP}}(\mathbf{x}, \omega)\, s_{\mathrm{pP}}(\mathbf{x}, \omega) + G_{\mathrm{sP}}(\mathbf{x}, \omega)\, s_{\mathrm{sP}}(\mathbf{x}, \omega) + \cdots. \tag{2}$$

This leads to a convolution+mixer model, where several apparent source functions (s_{P}, s_{pP} and s_{sP}) are mixed after undergoing unique path-specific convolutions. For example, the source functions of the pP and sP phases interfere with that of the P phase as they arrive with a delay of less than $50\,\mathrm{s}$ at most of the receivers for sources shallower than $\approx 200\,\mathrm{km}$.

Furthermore, even in scenarios where the influence of the previously mentioned scattered arrivals can be neglected by isolating individual arrivals, the convolution model faces limitations. This is particularly evident in light of studies [Kaiser et al., 2017] that employ multiple ruptures occurring across different fault planes in the earthquake source region to model complex earthquakes. While the convolution model assumes that the seismic moment density is uniform (up to a scalar multiple) in the source region, a more accurate representation of a complex earthquake involves mixing rupture episodes with different source functions, denoted as s^1, s^2, s^3, and so on:

$$d(\mathbf{x}, \omega) \approx G^1(\mathbf{x}, \omega)\, s^1(\mathbf{x}, \omega) + G^2(\mathbf{x}, \omega)\, s^2(\mathbf{x}, \omega) + G^3(\mathbf{x}, \omega)\, s^3(\mathbf{x}, \omega) + \cdots. \tag{3}$$

Here, the convolution+mixer model acknowledges that the apparent source functions are convolved with distinct Green's functions (G^1, G^2, G^3, etc.) specific to the moment tensor of each rupture episode.

Lastly, even within the framework of the convolutional model, constructing deconvolution operators is not straightforward because both the source function and Green's function are typically unknown. This challenge has prompted
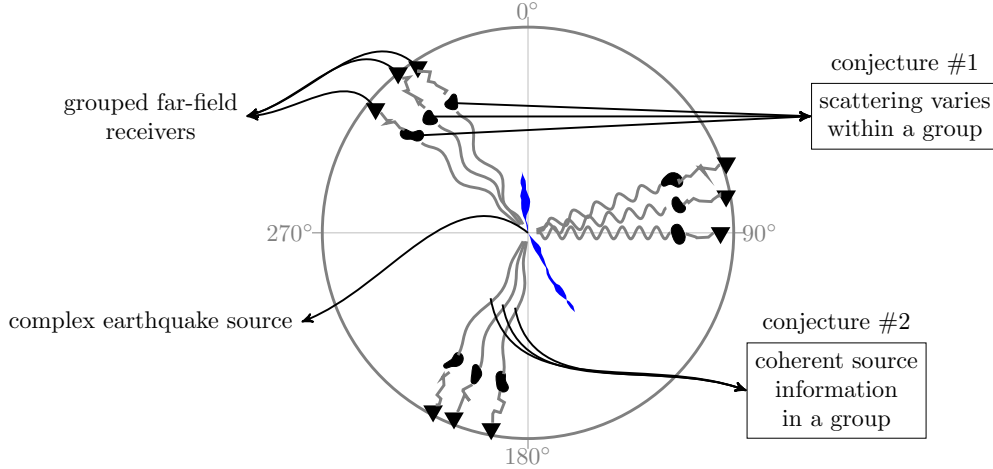
Figure 1: Symmetric autoencoder assumes a separation between the source and the path scales i.e., the earthquake source information (depicted using the raypath pattern) varies on a much slower spatial scale than the path effects (depicted using the blob shape). Here, for simplicity, only the azimuthal angle θ is considered as spatial coordinate, and ray bending is ignored. Each triangular marker represents a receiver.

the development of blind deconvolution strategies in seismology [Kaaresen and Taxt, 1998, Bharadwaj et al., 2020]. It's important to note that deconvolution algorithms overcome the mixing issue by isolating or windowing individual arrivals and rupture episodes within seismograms. However, it's essential to acknowledge that the success of isolating a specific arrival or rupture episode is influenced by various factors. These factors include the earthquake's depth, its complexity, and the duration of the source time function. As a result, the process of isolation can be subjective and may introduce biases into the analysis. This limitation becomes particularly evident when studying source functions and stopping phases of shallow earthquakes with larger durations, especially those associated with a high moment magnitude (Mw), as compared to deep earthquakes [Houston, 2015]. It also becomes pronounced when conducting comparative analyses involving a large number of earthquakes. Additionally, this limitation affects directivity analysis and the estimation of rupture velocities, especially for complex ruptures.

The Empirical Green's Function (EGF) method proposed by Hartzell [1978] was a breakthrough towards the investigation of earthquake source dynamics [Irikura, 1986, Plourde and Bostock, 2017, Wu et al., 2019, Lanza et al., 1999, Hutchings and Viegas, 2012, Bezada and Humphreys, 2012, Zhan et al.]. EGFs are seismograms that have a nearly impulsive source function and similar path effects as the main earthquake. Traditional methods for obtaining EGFs involve using low-magnitude earthquakes from the same source region, but this approach has several limitations. For example, 1. the focal mechanism of the low-magnitude earthquake may not be similar to that of the main earthquake, leading to ambiguity [Alessio et al., 1990]; 2. the EGF seismogram may have low signal-to-noise ratios when the magnitude is too low or a complex source function when an intermediate magnitude earthquake is chosen; 3. main earthquake with a low number of aftershocks cannot be analyzed using this approach, and 4. the selected EGF earthquake may not radiate the same apparent source function at all receivers, which can complicate the directivity analysis of the main earthquake. Synthetic seismograms generated with known Earth models were also used as EGF [Vallée et al., 2011, Vallée and Douet, 2016]. But synthetic seismograms have inaccurate propagation effects (geometrical spreading, scattering, anelastic attenuation). Nevertheless, the major issue with these methods is that they are not array-based i.e., they operate individually on seismograms. While backprojection [Ishii et al., 2005, Kiser and Ishii, 2012] methods utilize the spatial coherence of the wavefield recorded by a seismic network, they can be affected by the 3D velocity structure-related uncertainties, requiring empirical travel-time calibration to refine rupture imaging, as reported by Zeng et al. [2022].

This paper challenges the conventional assumptions inherent in convolutional models of seismograms, aiming to characterize earthquakes using a novel deep-learning approach. It presents a neural-network architecture, termed symmetric autoencoder [Bharadwaj et al., 2022, SymAE], that can be trained to isolate earthquake information from far-field seismograms. The design of our architecture relies on the fundamental assumption of scale separation, associating *slow* and *fast* spatial variations with source and path phenomena, respectively. It is crucial to clarify that the terms slow and fast do not pertain to time but rather to spatial characteristics. Specifically, slow denotes spatial variations with lower rapidity, while fast corresponds to spatial variations with higher rapidity. This framework can be seen as a vast generalization of multichannel blind deconvolution, where neural networks replace the convolution signal model.

SymAE learns to represent each seismogram using separate latent codes correlated to the source and path parameters. It constitutes of two encoders to generate source- and path-specific latent codes for each seismogram, and a decoder to reconstruct the seismogram in a near-lossless fashion using both codes. Some of the striking characteristics of our framework are given below.

1. A trained SymAE facilitates the generation of virtual seismograms by swapping the source or path information between any two seismograms of a given earthquake. This process is termed *redatuming*. Redatuming is a common technique in exploration seismology, often employed through representation theorems Wapenaar and Fokkema [2006]. However, our work introduces a novel application of redatuming in earthquake seismology, with a significant departure from traditional approaches. Instead of relying on representation theorems, we adopt an innovative autoencoding method for redatuming. This work underscores the importance of virtual seismograms. By deliberately incorporating consistent subsurface scattering effects into seismograms, we can extract interesting earthquake source information that is otherwise obscured in seismograms.

2. Traditional autoencoders cannot be trained to disentangle source information from seismograms. Therefore, SymAE network design and training are guided by the idea of *scale separation*, which is depicted in Fig. 1. Scale separation states that the variations in the far-field seismic measurements due to finite faulting occur across two different spatial scales — a slower scale in which the source processes occur and another faster scale corresponding to the path effects. Scale separation leads to an approximate symmetry: the source information is invariant w.r.t. receiver permutations within a group of *closely-spaced* receivers. We validate scale separation using the representation theorem and the Fraunhofer approximation for finite fault sources.

3. The training is unsupervised, therefore, our approach doesn't require traveltime or source-duration picking, empirical Green's function, knowledge on the subsurface scattering, modelling of source physics and path effects etc. Notably, SymAE can be trained on all the available far-field seismograms of multiple earthquakes. In this paper, we used displacement seismograms from ≈ 40 earthquakes at a variety of depths and epicentres. A trained SymAE can be reused or transferred subsequently when new earthquakes are detected e.g., we refer to Weiss et al. [2016] for a review on transfer learning.

4. SymAE associates the spatial coherence in the measured displacement field with the earthquake sources. Therefore, our approach scales well with data and benefits from improvements in the spatial coverage of the seismic networks. This approach offers a valuable middle ground between single-station analysis and array-based methods.

After training, to validate whether SymAE successfully disentangles source and path effects, we perform several sanity checks in this paper, including comparisons between P and S analysis. We then utilize the trained models in two different ways to characterize earthquakes:

- In alignment with the empirical Green's function approach, we develop strategies to quantify similarity between earthquakes originating from the same source region. Our analysis of collocated earthquakes from the Ridgecrest (California), Sea of Okhotsk, and Sumatra regions reveals intriguing associations among earthquakes. Of particular interest is the finding that a Sea of Okhotsk deep earthquake, previously analyzed and concluded to be supershear by Zhan et al., exhibits the least similarity compared to the rest of the Okhotsk region earthquakes. Another notable discovery, consistent with Yue et al. [2017], is the lack of correlation in similarity scores for Sumatra region earthquakes between P and S windows, unlike in other regions we have analyzed.

- We introduce the concept of uniformly generated virtual seismograms (USVS), which aid in directivity analysis of individual earthquakes. We demonstrate that USVS can offer superior resolution and accuracy compared to traditional envelope stacking methods. The most remarkable discovery using USVS, attributed to their high resolution, is the presence of complex stopping phases in earthquakes that would otherwise be smoothed out during the envelope averaging process due to time misalignments.

The remainder of the paper is organized as follows: we begin by introducing the concept of scale separation in the next section, followed by a section outlining the preparation of the training dataset. We then delve into the architecture of symmetric autoencoders, and the final sections present the application of the trained network on earthquake data.

4 Scale Separation

In this section, we use the Fraunhofer approximation to demonstrate that the spatial variation of the displacement field caused by finite faulting can be divided into two scales: a slow scale related to the source effects and a fast scale associated with subsurface scattering. This process, called scale separation, enables us to interpret the spatial coherence
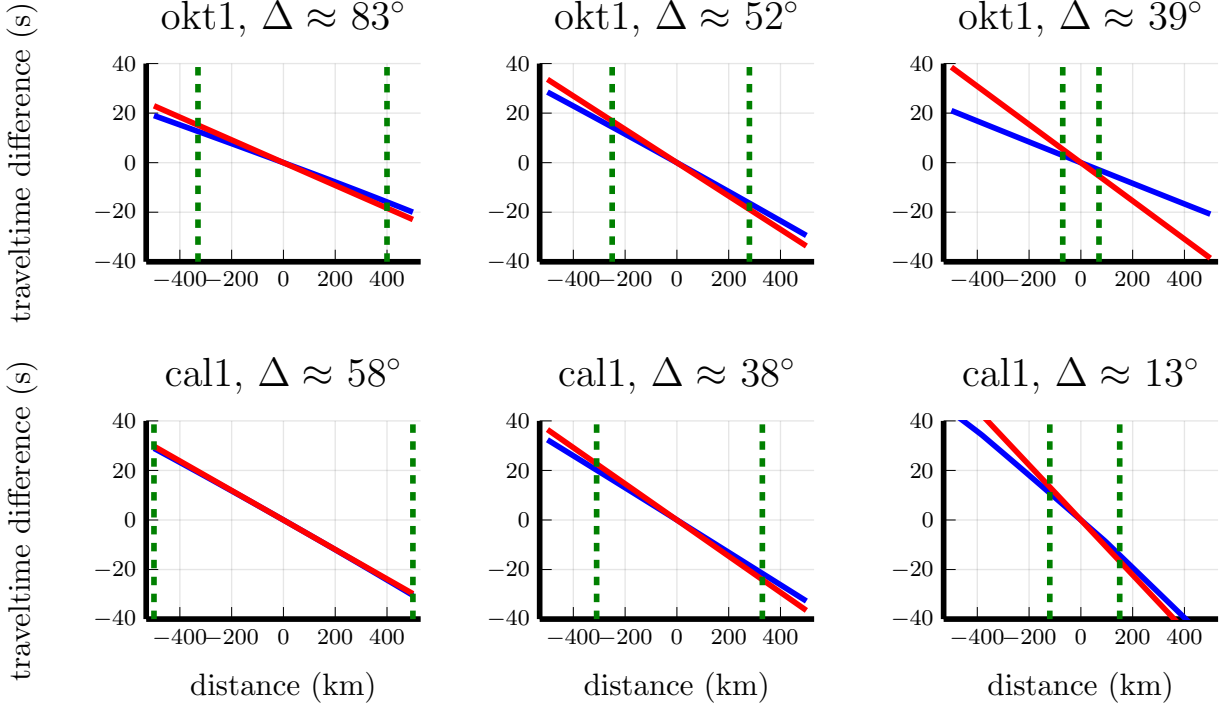
Figure 2: In this figure, red Fraunhofer-approximated traveltimes are compared to the actual PREM model traveltimes in blue, computed from different locations in the source region to a fixed receiver at a given epicentral distance. We considered a line source oriented in the direction of the ray leaving the focus. The approximation error increases as the epicentral distance decreases. The maximum fault length that can be analyzed based on the error in the Fraunhofer approximation is marked by dashed green lines, indicating the point where the error exceeds a quarter of the minimum period of 10 s considered in our analysis. In our work, we only used receivers with a minimum epicentral distance of 15°. Here, cal1 is a shallow crustal earthquake, while okt1 is a deep focus earthquake.

in the displacement field as a characteristic of the seismic source. Our starting point is the representation theorem [Aki and Richards, 2002, Chapter 3]. We consider a displacement discontinuity at $\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3)$ in the fault region. The ith component of the far-field displacement u^i thus generated at $\mathbf{x} = (x_1, x_2, x_3)$ is given in the frequency ω domain as:

$$u^i(\mathbf{x}, \omega; \boldsymbol{\xi}) = \sum_{j,k=1}^{3} \mathcal{G}^{ij,k}(\mathbf{x}, \omega; \boldsymbol{\xi}) \, m^{jk}(\omega; \boldsymbol{\xi}). \tag{4}$$

Here, m^{jk} denotes the (j, k)th component of the moment density tensor; and $\mathcal{G}^{ij,k}$ denotes the kth spatial derivative of the (i, j)th component of the elastodynamic Green's tensor. We first decompose Green's tensor into components associated with individual ray paths utilizing the high-frequency approximation: $\mathcal{G} = \mathcal{G}_1 + \mathcal{G}_2 + \cdots$, where the subscript denotes the ray path index. Then, we only consider the ray paths having wave velocity α in the source region, for instance, P, pP, PP rays. Finally, we rewrite Green's term in the Eq. (4) after employing Fraunhofer approximation as

$$\mathcal{G}^{ij,k}(\mathbf{x}, \omega; \boldsymbol{\xi}) \approx \sum_{l} \mathcal{G}_l^{ij,k}(\mathbf{x}, \omega; \boldsymbol{\xi}_0) \exp\left[-\imath\omega(\boldsymbol{\xi} \cdot \gamma_l)/\alpha\right], \tag{5}$$

where γ_l is the unique angle with which lth ray leaves the focus denoted by $\boldsymbol{\xi}_0$. Note that a homogeneous P-wave velocity is assumed in the source region. The Fraunhofer approximation simplifies the full-wave solution for the displacement field caused by finite faulting by assuming that the fault is far enough from the receiver. This approximation accounts only for the far-field phase correction, or travel-time difference, between the source region locations $\boldsymbol{\xi}$ and $\boldsymbol{\xi}_0$ and the receiver. The validation of the Fraunhofer approximation, which is integral to the concept of scale separation, was assessed through ray tracing, albeit recognizing it as a somewhat weaker form of validation. In Figure 2, we compared travel times calculated using the Fraunhofer approximation with those obtained by tracing rays from various source locations to a receiver within the PREM model. Multiple receivers at epicentral distances exceeding 10 degrees were considered. The accuracy of the Fraunhofer approximation is contingent upon the introduced treveltime error,

which should be no greater than one-fourth of the minimum period. Consequently, the validity of this approximation is frequency-dependent. In our study, we analyzed a wavefield with limited bandwidth and restricted frequencies to a maximum of $0.1\,\text{Hz}$. This permitted the application of the Fraunhofer approximation for dominant source region lengths up to $300\,\text{km}$, as depicted in Figure 2. However, it should be noted that higher frequencies can still be analyzed when dealing with low-magnitude earthquakes possessing shorter dominant source region lengths. While the validity of the Fraunhofer approximation imposes constraints on the highest possible frequency, conversely, the scale separation principle imposes restrictions on the lowest feasible frequency for analysis, as we will elaborate on in subsequent sections of this paper.

We can express the total displacement at $\mathbf{x}$ by integrating u^i over the source region using the principle of superposition and rewriting eqs. (4) and (5), as follows:

$$d^i(\mathbf{x}, \omega) = \int_\Xi u^i(\mathbf{x}, \omega; \boldsymbol{\xi})\, \mathrm{d}\Xi = \sum_l \sum_{j,k=1}^3 \left(\mathcal{G}_l^{ij,k}(\mathbf{x}, \omega; \boldsymbol{\xi}_0) \underbrace{\int_\Xi m^{jk}(\omega; \boldsymbol{\xi})\, \exp\left[-\imath \omega (\boldsymbol{\xi} \cdot \boldsymbol{\gamma}_l)/c\right] \mathrm{d}\Xi}_{\text{apparent source function } s_l^{jk}(\omega; \gamma_l)} \right). \tag{6}$$

Here, $\mathrm{d}\Xi$ represents an infinitesimal element at $\boldsymbol{\xi}$ in the source region. It is clear that the apparent source function depends on γ_l, which is unique to a particular seismic arrival, making it impossible to simplify this equation further unless we ignore multipathing. The simplified signal model in Eq. (3) can be derived for a single seismic arrival, which still contains a mixture of rupture episodes. Similarly, assuming the invariance of the moment density tensor with ξ, we can separate the contribution of different arrivals and derive a simpler model, as shown in Eq. (2). The use of simplified models may limit the accuracy of the obtained source information from the seismic wavefield. Our approach, on the other hand, does not impose restrictions on the source model; instead, we rely on the principle of scale separation described below. It is important to note that the accuracy of source information is particularly important when estimating the kinematic rupture properties and directivity effects, described in Madariaga [2015].

4.1 Source Scale

The extent of the fault region determines the spatial scale at which the source information varies. A point source located at the focus, represented by $m^{jk}(\omega; \boldsymbol{\xi}) = M^{jk}(\omega)\delta(\boldsymbol{\xi})$, where M^{jk} is a function of angular frequency ω and δ is the Dirac delta function, produces a source term in Eq. (6) that is independent of the ray's takeoff angle γ_l. Here, we first observe that the form of the source integral is the spatial Fourier transform of the moment distribution in the source region [Demanet, 2012]. Next, we apply the convolution theorem to determine the source scale. For convenience, we assume a line source along the ξ_1 dimension and also use a single arrival for simplicity by dropping the subscript l. Rewriting the source term using these simplifications leads to:

$$s^{jk}(\omega; k_1) = \int_{\xi_1} \Pi\left(\frac{2\xi_1}{L_1}\right) m^{jk}(\omega; \xi_1) \exp\left[-\imath\, k_1\, \xi_1\right] \mathrm{d}\xi_1, \tag{7}$$

where $L_1 > 0$ is the length of the fault, Π denotes a unit rectangular function and $k_1 = \omega \cos(\psi)/c$ is the transform variable (wavenumber) dependent on the angle ψ between the line source (ξ_1 axis) and the ray takeoff direction (unit vector γ). Here, the multiplication with Π constrains the spatial extent of the moment rate function m^{jk}. According to the convolution theorem, this multiplication is equivalent to a convolution in the wavenumber domain with a sinc function

$$\mathcal{F}\left(\Pi\left(\frac{2\xi_1}{L_1}\right)\right) = \frac{L_1}{2\sqrt{(2\pi)}} \operatorname{sinc}\left(\frac{L_1 k_1}{4}\right), \tag{8}$$

where $\mathcal{F}$ denotes Fourier transform. This convolution smears the source information along the spatial dimension of the measurements — it leads to a reduction in the resolution proportional to the reciprocal of the fault length L_1 [Harris, 1978]. The resolution along k_1, which can be approximated by the width of the sinc function in Eq. (8), is given by

$$\Delta k_1 = \frac{4\pi}{L_1}. \tag{9}$$

Finally, a linear approximation of $k_1 = \omega \cos(\psi)/c$, as given by

$$\mathrm{d}k_1 = \frac{-\mathrm{d}\psi\, \omega\, \sin(\psi)}{c}, \tag{10}$$

with perturbations $\mathrm{d}k_1$ and $\mathrm{d}\psi$, can be used to estimate the corresponding angular resolution

$$\Delta\psi = L_1^{-1} \frac{4\,\pi\,c}{\omega\,\sin(\psi)}. \tag{11}$$

dominant fault length (km) $L_{\max}$	angular resolution (degree) $\Delta\psi$ for P waves	angular resolution (degree) $\Delta\psi$ for S waves
100.00	103.1	51.6
200.00	51.5	25.8
300.00	34.4	17.2
400.00	25.8	13.1

Table 1: The angular resolution of the source information is dependent on the dominant length of the source region, and in our study, we used the Healpix spherical tessellation of the focal sphere to group the receivers. The approximate angular resolution of this tessellation was $14°$.

The angular resolution $\Delta\psi$ for a three-dimensional source region depends on the length of the dominant dimension $L_{\max} = \max\{L_1, L_2, L_3\}$. For large-magnitude earthquakes, assuming large fault lengths, the source information varies at a faster spatial scale. For a given fault length, the source information in S waves (high c) at the source region exhibits variations at a faster scale than P waves (low c) at the source region. We analyze frequencies up to 0.1 Hz and use Tab. 1 to determine expected angular resolutions for various fault lengths for both P and S waves. For instance, if the dominant length of the source region is 300 km, the expected angular resolution of S-wave source information is roughly $17°$. Our next objective is to show that the path information varies on a faster scale compared to this source scale as depicted in Fig. 1.

4.2 Path Scale

Given the smallest angular resolution of the source information as presented in Tab. 1, we opted for the Hierarchical Equal Area isoLatitude Pixelation method [Gorski et al., 2005, HEALPIX]. HEALPIX divides the earthquake focal sphere into pixels, each possessing equal surface area, providing an approximate angular resolution of $\Delta\psi = 14°$. Fig. 3a illustrates the resulting pixelated focal sphere. Within this pixelation scheme, all traced rays intersecting a particular focal-sphere pixel are assumed to carry identical source information. In other words, the source effects are considered uniform across all far-field receivers connected by rays intersecting a given pixel.

To establish scale separation, our objective is to demonstrate significant variations in path effects among a group of receivers associated with any given pixel. An important observation is that the receivers within each group are distributed across a wide range of epicentral distances. The coloring of the receiver groups in Fig. 3b reveals that lower-indexed pixels exhibit receivers spread across a broader range of epicentral distances in comparison to higher-indexed pixels. We selected a pixel and computed the raypaths connecting the source focus to the receivers while intersecting the pixel. Subsequently, we determined the corresponding travel times along these raypaths, and the results are presented in Fig. 4. Our analysis focused on examining the variation in traveltime delay of scattered arrivals relative to the P arrival within the receiver group. Similar analysis can be performed for waves recorded in the S window. This variation in traveltime delay among the receivers serves as an indicator of the scale at which scattering effects fluctuate. Significant variations in time delay among the receivers suggest that path effects are changing at a faster rate than source effects, which remain uniform for the selected pixel.

Arrivals with a time-delay variation of less than a quarter of the period within the group of receivers are classified as coherently scattered arrivals. This category typically includes pP and sP arrivals, as shown in Fig. 4. It can be noticed that the pP and sP arrivals arrive after a particular time delay following the P arrival irrespective of the epicentral distance, and their traveltime curves are roughly parallel to that of the P arrival. On the other hand, arrivals with traveltime delays of more than a quarter of the period within the group of receivers are classified as incoherent scattering. This category includes crustal phases, PP, PcP and scattering due to random inhomogeneities when considering the P window. Incoherent scattering adheres to the scale separation conjecture.

Following the observation of coherent scattering, we want to emphasize the nuanced use of the terms *source effects* and *path effects* which we will consistently use in the remainder of this article. While our goal is to disentangle source effects from receiver-specific path variations, we are, in reality, more accurately isolating coherent features from within receiver groups that may encompass coherent path information, e.g. surface-reflected arrivals pP and sP, source-side reverberations, and coherent attenuation of high-frequencies, in addition to the primary source information. The coherent scattering effects due to intrinsic attenuation in the Earth is a subject that we plan to explore in future research. Note that the coherent scattering is path independent and cannot be effectively disentangled from the source information using our methods due to the violation of scale separation (Fig. 1). We synonymously use the term *path*
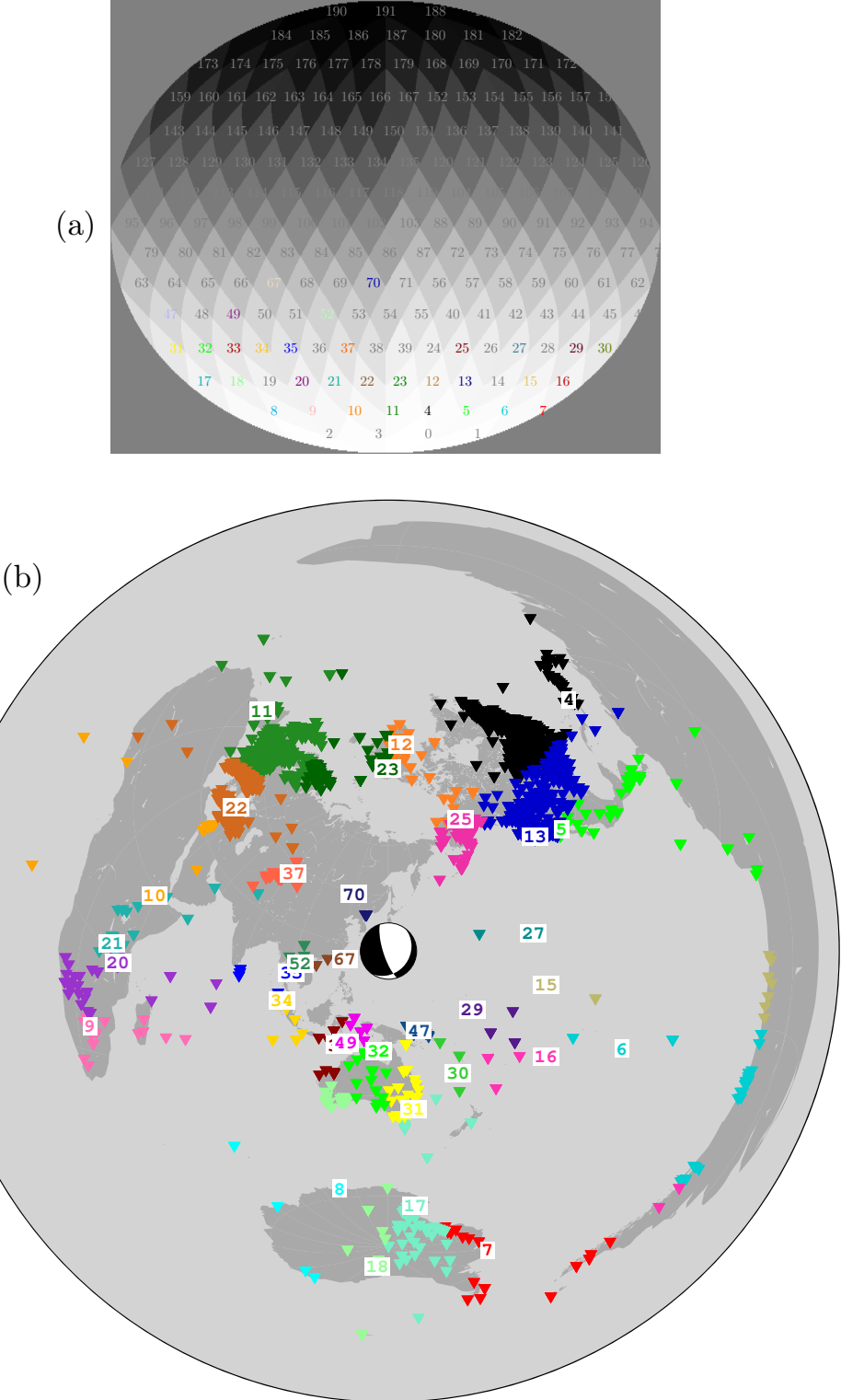
Figure 3: (a) Hierarchical Equal Area isoLatitude Pixelation [Gorski et al., 2005, HEALPix] of the earthquake focal sphere, where areas of all the pixels are identical, as the name suggests. The pixels are arranged in order of increasing azimuthal angle while maintaining constant polar angles, e.g., it can be observed that pixels 4–11 have the same polar angle. (b) Grouping of receivers: each numbered receiver group (represented by a distinct color) corresponds to a pixel on the focal sphere. The pixels on the focal sphere are also color-coded in the same unique colors to indicate their association with specific receiver groups. Here, we present an example of the Bonin earthquake (Mw 7.8) that occurred on May 30, 2015.
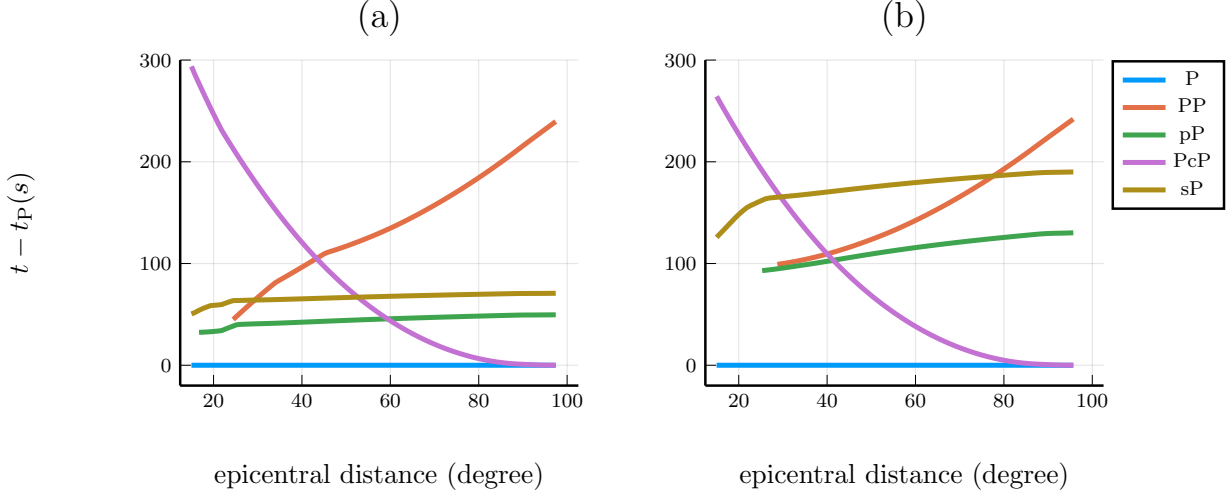
Figure 4: Traveltime delays relative to the P arrival for certain scattered arrivals in the P-window of seismograms are depicted for both (a) 200 km and (b) 600 km deep earthquakes. Here, the arrivals such as PP and PcP are indicative of incoherent scattering, as their delays exhibit variations exceeding a quarter of the maximum period ($100\,\mathrm{s}$), which corresponds to the low-frequency limit examined in our study. Conversely, the surface-reflected arrivals pP and sP exhibit coherence, as their delays show minimal dependence on the epicentral distance. Incoherent scattering aligns with the scale-separation conjecture (Fig. 1). On the other hand, surface-reflected coherent scattering poses a challenge as it cannot be disentangled from the source effects within the framework we have presented.

effects to loosely refer to incoherent scattering, where the traveltime delay (relative to either P or S) varies more than a quarter of the maximum period corresponding to a minimum frequency of $0.01\,\mathrm{Hz}$ considered in this study.

5 Training Data

This section outlines the methodology used to construct the dataset for training SymAE to learn a source-path representation. Broadband displacement seismograms of selected earthquakes (listed in Tab. 2) were retrieved from the IRIS data center using the ObspyDMT package [Hosseini and Sigloch, 2017]. The seismograms were then instrument-corrected and manually inspected to remove outliers. Subsequently, standard pre-processing techniques, such as band-pass filtering, resampling, and standardization, were applied to the remaining seismograms. During this study, our analysis focused on the time window of $\pm 300\,\mathrm{s}$ centered around the PREM P or S arrivals of the seismograms. The length of the time window is not critical as long as it is sufficient to capture the source effects. Furthermore, the same analysis can be applied to other time windows by identifying the relevant arrivals.

In order to utilize SymAE for each earthquake, the seismograms were divided into multiple groups based on the idea that the receivers in each group capture identical apparent source function, which is sensitive to the direction (represented by γ) in which the waves depart the source region. To aid in this analysis, we used a pixelated focal sphere, as implemented in HEALPIX, which divided the spherical surface into pixels of equal surface area with an angular resolution of $14°$. As depicted in Fig. 3, the pixels were numbered and arranged in order of increasing azimuthal angle while maintaining constant polar angles, with higher pixel indices corresponding to higher values of the polar angle ϕ. This distribution allowed us to easily examine the azimuthal dependence of the apparent source function by selecting pixels with constant polar angles. The seismograms were grouped by tracing rays from the source region to all the receivers using the PREM model, and assigning each receiver to a pixel index based on the intersecting pixel of the emanating ray from the source. For instance, Fig. 3 displays the grouped receivers for the Bonin earthquake (with code bon1 in Tab. 2). This grouping strategy was applied to all earthquakes in the dataset, ensuring that seismograms from different earthquakes were grouped separately. On average, 10 groups per earthquake were obtained, although this depended on data availability.

Once the seismograms are grouped based on the pixel indices as described earlier, the number of receivers in each group can vary depending on the distribution of receivers for the specific earthquake under analysis. To ensure that the coherent information present in each group of seismograms is solely attributed to the earthquake source, it is crucial that path factors are dissimilar among the receivers within a group. If a group contains a low number of receivers that are clustered closely together in space, local geological structures or other nearby seismic sources could have

#	Code	Date	Time	Mw	Depth (km)	Half Duration (s)	Region
1	okt1	2013-05-24	05:44:48	8.30	598.00	35.7	Sea of Okhotsk
2	cal1	2019-07-04	17:33:49	6.40	10.00	4.1	Central California
3	cal2	2019-07-06	03:19:53	7.10	8.00	8.0	Central California
4	nic	2015-11-08	16:47:02	6.60	10.00	4.6	Nicobar Island
5	chl1	2010-02-27	06:34:11	8.80	23.00	60.0	Near Coast of Central Chile
6	kam	2018-11-14	21:21:50	6.10	49.00	2.9	East Coast of Kamchatka
7	mnd	2019-05-31	10:12:32	6.10	99.00	2.9	Mindanao Philippines
8	crt	2019-11-27	07:23:42	6.00	69.00	2.6	Crete Greece
9	cps	2015-12-17	19:49:53	6.60	85.00	4.7	Near Coast of Chiapas
10	csbg	2019-04-29	14:19:52	6.30	10.00	3.5	Carlsberg Ridge
11	okt2	2013-05-24	14:56:31	6.70	624.00	5.6	Sea of Okhotsk
12	okt3	2008-07-05	02:12:04	7.70	633.00	17.3	Sea of Okhotsk
13	okt4	2008-11-24	09:02:58	7.30	492.00	10.9	Sea of Okhotsk
14	okt5	2013-10-01	03:38:21	6.70	573.00	5.6	Sea of Okhotsk
15	okt6	2012-08-14	02:59:38	7.70	583.00	17.8	Sea of Okhotsk
16	chl2	2010-03-05	11:47:06	6.60	18.00	4.9	Near Coast of Central Chile
17	chl3	2010-03-16	02:21:57	6.70	18.00	5.1	Near Coast of Central Chile
18	cal3	2019-07-06	03:47:53	5.50	5.00	-	Central California
19	cal4	2020-06-04	01:32:11	5.51	8.00	1.5	Central California
20	pb2	2015-11-24	22:45:38	7.60	606.00	16.6	Peru Brazil
21	dnl	2002-11-03	22:12:41	7.90	4.20	23.5	Central Alaska
22	fij1	2018-08-19	00:19:40	8.20	600.00	30.8	Fiji Islands
23	fij2	2018-09-06	15:49:18	7.90	670.00	21.4	Fiji Islands
24	fij3	2014-11-01	18:57:22	7.10	434.00	8.7	Fiji Islands
25	fij4	2018-11-18	20:25:46	6.80	540.00	5.8	Fiji Islands
26	fij5	2021-04-24	00:23:38	6.50	301.00	4.5	Fiji Islands
27	hnd1	2015-10-26	09:09:42	7.50	231.00	13.7	Hindu Kush Region
28	rat	2014-06-23	20:53:09	7.90	109.00	22.4	Rat Islands
29	bon1	2015-05-30	11:23:02	7.80	664.00	20.7	Bonin Island Japan Reg
30	van1	2014-03-05	09:56:57	6.30	638.00	3.8	Vanuatu Islands
31	spn	2010-04-11	22:08:12	6.30	609.00	3.6	Spain
32	nbl	1994-06-09	00:33:16	8.20	631.00	20.0	Northern Bolivia
33	nzd	2016-11-13	11:02:56	7.80	15.11	1.8	South Island New Zealand
34	toh1	2011-32-11	05:46:24	9.10	29.0	70.0	East Coast of Honshu
35	sum1	2012-04-11	08:38:36	8.60	20.0	32.2	Sumatra
36	tur1	2023-02-06	01:17:35	7.80	8.0	18.9	Turkey
37	okt7	2002-11-17	04:53:53	7.30	459.00	11.4	Sea of Okhotsk
38	sum2	2012-01-10	18:36:59	7.2	19.00	9.6	Sumatra
39	sum3	2006-04-19	20:36:46	6.2	17.00	3.1	Sumatra
40	sum4	2007-10-04	12:40:31	6.2	35.00	3.2	Sumatra

Table 2: The SymAE was trained using displacement seismograms of earthquakes with both shallow and deep events. The earthquake code names used throughout the paper are shown in the table. Following training, SymAE synthesized Uniformly Scattered Virtual Seismograms (USVS) for each earthquake.

effects that are not sufficiently dissimilar among the receivers within the group. Consequently, we excluded groups with fewer than 20 seismograms (for any displacement component) or a maximum inter-receiver distance below a specified threshold value. This approach helps to enhance the accuracy and dependability of the analysis. Additionally, it is worth mentioning that during grouping and subsequent analysis, all available components of the displacement field were processed independently.

Our framework allows for the training and analysis of a large number of earthquakes, utilizing all available grouped seismograms to form training datapoints $\{X_i\}_{i=1,\cdots,n_X}$. Each datapoint X_i is filled with time-windowed seismograms corresponding to a label, which contains the earthquake code and the pixel index of the focal sphere. The seismograms in each datapoint are indexed as $X_i[j]$ for $j = 1, \cdots, n_p$, where $X_i = [X_i[1]; \cdots; X_i[n_p]]$. In our notation, $[Y_1; Y_2]$ denotes a vertical concatenation of two vectors Y_1 and Y_2. Each seismogram $X_i[j]$ represents the measured displacement over time, without regard to the component. To generate labels for each datapoint, we randomly sample an earthquake and its focal pixel with replacement, with probabilities of selection proportional to the number of receivers in the group. We use the notation

$$\underset{\text{eq}\rightsquigarrow\text{A}}{\overset{\text{eq\_}k}{D}}, \tag{12}$$

to denote a pre-processed displacement seismogram in this paper. The superscript denotes the source information of the earthquake (eq) recorded at receivers associated with kth pixel, and the subscript indicates the effects caused by the path connecting the source to the receiver. The usefulness of this unconventional notation will become apparent when we introduce redatuming and the generation of virtual seismograms later. Using this notation, we can represent a data point corresponding to an earthquake code eq for receivers A, B, and C associated with a given focal pixel as follows: For instance, if we use this notation, we can represent a datapoint corresponding to an earthquake eq for receivers A, B, and C associated with a given focal pixel as follows:

$$X_i = \left[\underset{\text{eq}\rightsquigarrow\text{A}}{\overset{\text{eq\_}k}{D}}; \underset{\text{eq}\rightsquigarrow\text{B}}{\overset{\text{eq\_}k}{D}}; \underset{\text{eq}\rightsquigarrow\text{C}}{\overset{\text{eq\_}k}{D}}; \cdots \right]. \tag{13}$$

As discussed earlier, the seismograms within a datapoint vary due to path differences and contain identical source information. We will next demonstrate that having access to sufficiently dissimilar receivers allows for disentangling source and path information, without relying on the underlying convolutional model.

6 Symmetric Autoencoder

To better understand the displacement field caused by earthquakes, it is necessary to convert the representation of the measured displacement from its original time-position coordinates to source-path coordinates. This conversion process involves non-linear transformations, which can be learned in an unsupervised manner through representation learning. Representation learning is a method of discovering the underlying structure of data and finding useful representations for various tasks. Autoencoders, which are comprised of two components, are a popular method for representation learning. In our context, the first component is an encoder Enc that transforms seismograms in each datapoint X_i into latent code $H_i = \text{Enc}(X_i)$. The second component is a decoder Dec that reconstructs X_i from the latent code. Both the encoder and decoder functions are non-linear and are determined by minimizing the reconstruction loss,

$$\text{Enc, Dec} = \underset{\text{Enc, Dec}}{\arg\min} \sum_i \|X_i - \text{Dec}(\text{Enc}(X_i))\|^2, \tag{14}$$

over the training dataset $\{X_i\}$. If the encoder and decoder functions are constrained to be linear operators, this program reduces to principal component analysis (PCA). In this case, the output of the encoder (i.e., the latent code) is analogous to the singular values, while the encoder corresponds to the left singular vectors of the training dataset. A traditional deterministic autoencoding strategy [Doersch, 2016] is not sufficient for obtaining the source-path representation of seismograms. While traditional autoencoders efficiently compress seismograms to the latent space [Dai et al., 2018], their interpretation is not straightforward because the latent representation no longer describes the geometry of the dataset using linear subspaces [Klys et al., 2018]. Therefore, we use symmetric autoencoder [Bharadwaj et al., 2022], which utilizes the scale separation principle (described in Section 4) and stochastic regularization to disentangle the source and path effects. The network architecture, as illustrated in Fig. 5, is specifically designed for training with the grouped seismograms discussed in the previous section.

The symmetric autoencoder is a variant of non-linear autoencoders that includes a unique encoder structure that can be mathematically described as follows:

$$\text{Enc}(X_i) = [\text{SEnc}(X_i); \text{PEnc}(X_i[1]); \cdots ; \text{PEnc}(X_i[n_p])]. \tag{15}$$
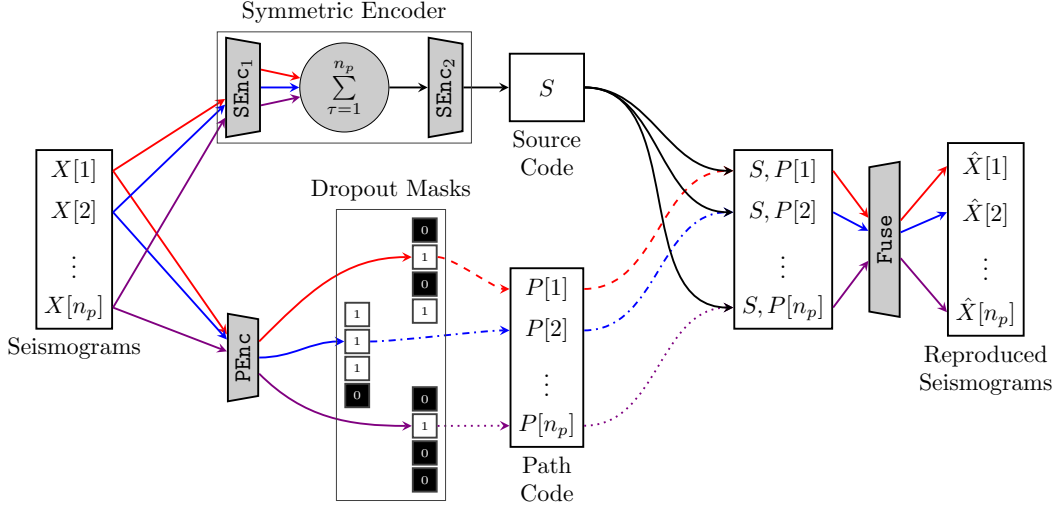
Figure 5: The architecture of a symmetric autoencoder disentangles the source information from the path information in the latent space. The earthquake source information is coherent across the grouped seismograms in a datapoint X. It can only propagate through the network via solid arrows because dropout masks prevent its propagation. Arrow colours indicate the seismogram-specific path effects — notice that a symmetric function, i.e., symmetric w.r.t. the order of the receivers, prevents the propagation of path information.

Here, the output of the encoder is partitioned into the interpretable source and path components. Specifically, each datapoint, $X_i = [X_i[1]; \cdots ; X_i[n_p]]$, is represented as a structured latent code,

$$H_i = [S_i; P_i[1]; \cdots ; P_i[n_p]], \tag{16}$$

where the sub-component $S_i = \texttt{SEnc}(X_i)$ contains the source information from X_i, while the remaining sub-components $P_i[j] = \texttt{PEnc}(X_i[j])$ encode the seismogram-specific path information. The dimensions l and r of the latent codes $S_i \in \mathbb{R}^l$ and $P_i[\cdot] \in \mathbb{R}^r$ are user-defined hyperparameters and need not be identical. The length of S_i can be extended as desired, while the length of P_i should be small to restrict the expressiveness of the path code as discussed later. To determine the optimal value of r, it is essential to tune it on a validation dataset. The decoder function, $\texttt{Dec}$, uses these latent codes to reconstruct the original seismograms in the datapoint, one at a time. In other words, it constitutes a function $\texttt{Fuse}$ that non-linearly combines the source code S_i with each seismogram-specific path code $P_i[\cdot]$ to reconstruct the original datapoint:

$$\begin{aligned} \hat{X}_i &= \texttt{Dec}(H_i) = \texttt{Dec}([S_i; P_i[1]; \cdots ; P_i[n_p]]) \\ &= [\texttt{Fuse}([S_i; P_i[1]]); \cdots ; \texttt{Fuse}([S_i; P_i[n_p]])]. \end{aligned} \tag{17}$$

SymAE's architecture does not impose any constraints on the parametrization of the functions $\texttt{PEnc}$ and $\texttt{Fuse}$. We have employed standard convolutional layers as fundamental building blocks in our deep learning model to parameterize the functions $\texttt{PEnc}$ and $\texttt{Dec}$. However, the function $\texttt{SEnc}$ is specifically designed to capture the coherency or similarities among the seismograms within a datapoint X_i. Permutation invariance is imposed on the function, guaranteeing that the output remains the same regardless of the order of the seismograms within the datapoint. Mathematically, this means that

$$S_i = \texttt{SEnc}(X_i) = \texttt{SEnc}(X_i[\Pi(1{:}n_p)]) \tag{18}$$

for all permutations Π of the seismograms. By enforcing this permutation invariance, we ensure that path information cannot be encoded using $\texttt{SEnc}$ without significant loss of information, allowing only coherent source information to be encoded in S_i. Since the autoencoder aims to reconstruct the input data in a lossless manner, the path information is expected to be encoded by $\texttt{PEnc}$.

The property, known as symmetry, is a characteristic of permutation-invariant network architectures [Zaheer et al., 2017, Ilse et al., 2018]. These architectures make use of pooling functions, such as the mean or the max, across the seismograms. In our case, the seismograms input to the symmetric function $\texttt{SEnc}$ are first transformed using $\texttt{SEnc}_1$ and then summed together. This output is then processed by $\texttt{SEnc}_2$, resulting in the equation:

$$S_i = \texttt{SEnc}_2 \left(\frac{1}{n_p} \sum_{j=1}^{n_p} \texttt{SEnc}_1(X_i[j]). \right). \tag{19}$$

Intuitively, this network architecture of SEnc ensures that the source information extracted by SEnc_1 from each of the seismograms constructively interferes during the summation, and then is further compressed using SEnc_2. The key observation is that this summation is symmetric with respect to the ordering of seismograms. This ensures that the desired permutation invariance of SEnc (Eq. (18)) is achieved. Again, the functions SEnc_1 and SEnc_2 are parametrized using standard deep-learning blocks, i.e., by compositions of fully connected layers or convolutional layers.

The unconstrained path encoder PEnc has the capability to capture the path information and seismogram-specific details in a datapoint. However, it can also encode the source information, which goes against the objective of achieving disentanglement of source and path information. In other words, it is crucial that the decoder Fuse does not ignore the S_i component in favor of using solely $P_i[\cdot]$ information for lossless reconstruction. Hence, similar to how we constrained SEnc via symmetry, we limit the expressiveness of the path encoder PEnc using stochastic regularization [Wang and Manning, 2013, Kingma et al., 2015]. We can limit the expressiveness of the path encoder PEnc by decreasing the length of the path code r, but this may hinder the goal of lossless reconstruction. As an alternative approach we apply *Bernoulli dropout* [Srivastava et al., 2014] to the output of PEnc during training, resulting in an obfuscated path code

$$P_i[j] = \text{Dropout}(\text{PEnc}(X_i[j])). \tag{20}$$

By introducing dropout with high probability q to obfuscate each element P_i, artificial dissimilarities are induced among the seismograms. This incentivizes the decoder to rely on the code S, which is held constant for each seismogram as shown in Eq. (17), to reconstruct the source information. At test time, the unaltered and unobfuscated P_i code is fed into the decoder. The strength of the dropout, proportional to q, along with the path-code length r, is a hyperparameter that must be tuned based on an external validation set. During the tuning process, it is essential to strike a balance between q and r to ensure that each P_i can still encode sufficient path information, while the dropout is strong enough to obfuscate the source information. In summary, SymAE achieves lossless reconstruction by relying on the symbiotic relationship between its source and path encoders, each with its own limitations, complementing each other to reconstruct the data accurately without loss. The components PEnc, SEnc, and Fuse of SymAE are trained concurrently by minimizing Eq. (14) with the regularization mechanism just described, ensuring the joint optimization of the network for both source and path information. The following sections showcase the use of these components after training.

7 Redatuming

SymAE learns a representation of seismograms with disentangled earthquake source and path information, which is highly valuable in seismology. This representation allows us to manipulate the latent representation of seismograms in order to synthesize virtual or hybrid seismograms, a process known as redatuming. With this ability, we can transfer either the source or path information from one seismogram to another, providing new insights and understanding of the underlying seismic data. For instance, we can use SymAE to swap source and path information from two different displacement seismograms,

$$\underset{\text{chl2}\rightsquigarrow\text{A}}{\overset{\text{chl2\_}k}{D}} \quad \text{and} \quad \underset{\text{chl3}\rightsquigarrow\text{B}}{\overset{\text{chl3\_}k}{D}}, \tag{21}$$

to create two new virtual seismograms. The first virtual seismogram,

$$\underset{\text{chl3}\rightsquigarrow\text{B}}{\overset{\text{chl2\_}k}{D}} = \text{Fuse}\left(\left[\text{SEnc}\left(\left[\underset{\text{chl2}\rightsquigarrow\text{A1}}{\overset{\text{chl2\_}k}{D}}; \underset{\text{chl2}\rightsquigarrow\text{A2}}{\overset{\text{chl2\_}k}{D}}; \underset{\text{chl2}\rightsquigarrow\text{A3}}{\overset{\text{chl2\_}k}{D}}; \cdots\right]\right); \text{PEnc}\left(\underset{\text{chl3}\rightsquigarrow\text{B}}{\overset{\text{chl3\_}k}{D}}\right)\right]\right), \tag{22}$$

is generated by combining the source information from the kth pixel of the chl2 earthquake with the path effects from the second chl3 seismogram. Similarly, the second virtual seismogram,

$$\underset{\text{chl2}\rightsquigarrow\text{A}}{\overset{\text{chl3\_}k}{D}} = \text{Fuse}\left(\left[\text{SEnc}\left(\left[\underset{\text{chl3}\rightsquigarrow\text{B1}}{\overset{\text{chl3\_}k}{D}}; \underset{\text{chl3}\rightsquigarrow\text{B2}}{\overset{\text{chl3\_}k}{D}}; \underset{\text{chl3}\rightsquigarrow\text{B3}}{\overset{\text{chl3\_}k}{D}}; \cdots\right]\right); \text{PEnc}\left(\underset{\text{chl2}\rightsquigarrow\text{A}}{\overset{\text{chl2\_}k}{D}}\right)\right]\right), \tag{23}$$

is generated by combining the source information from the kth pixel of chl3 with the path effects from the first chl2 seismogram. Again, recall that SEnc is trained to extract the source information from seismograms, PEnc extracts the path information, and Fuse fuses the source and path information to generate a virtual seismogram. In our notation, the superscript denotes the source information contained in the seismogram, while the subscript denotes the path that connects the source to the receiver. In the subsequent discussion, we will elucidate the importance of these virtual seismograms in two distinct scenarios. Further applications of redatuming will be explored in future research.

7.1 Similarity Between Collocated Earthquakes — Generalizing Empirical Green's Functions

In this subsection, we follow a methodology closely aligned with the empirical Green's function (EGF) approach, focusing on earthquakes originating from the same source region. Specifically, our analysis centers on redatuming path

(a) P Waveform Redatuming
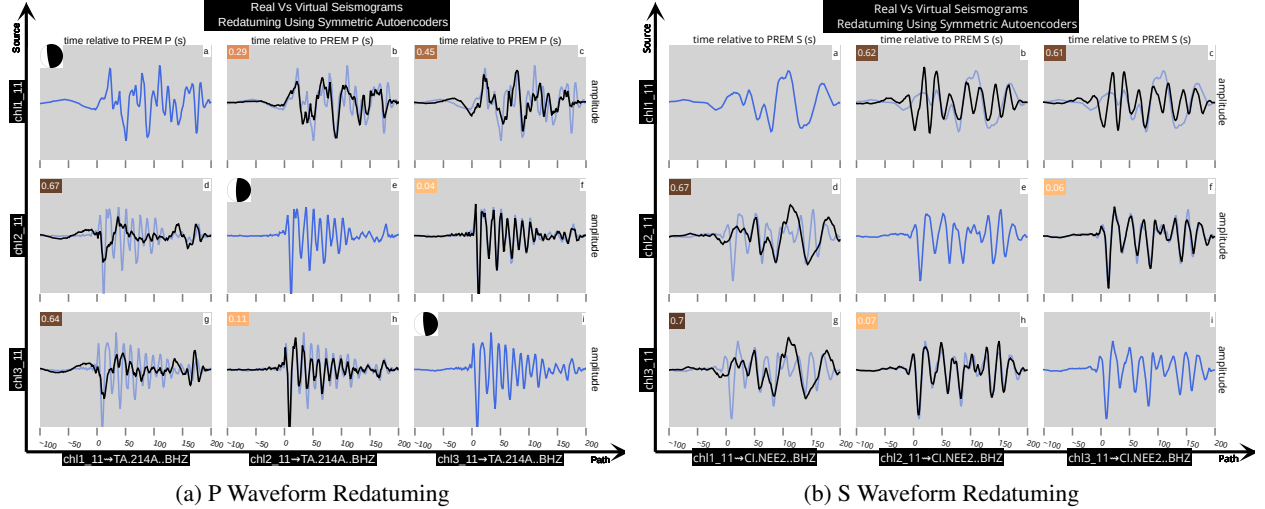
(b) S Waveform Redatuming

Figure 6: Displacement seismograms are embedded in the source-path latent space of SymAE. The blue time series on the diagonal represent the measured seismograms, while the black ones are virtual seismograms generated through redatuming, a process involving the exchange of source and path information between the measured seismograms. Since the earthquakes are selected from the same source region, if they share similarity, the virtual seismograms should exhibit strong correlation with the measured seismograms of a particular row. In this context, the Pearson correlation distance between the measured and virtual seismograms is depicted. (a) Demonstrates an example of P-windowed seismograms from earthquakes along the central coast of Chile. (b) Replicates the process in a), but with S-windowed seismograms. In both scenarios, earthquakes chl2 and chl3 exhibit substantial similarity, whereas the presence of higher correlation distances implies a distinct source mechanism for chl1 when compared to chl2 and chl3. The agreement between the P and S redatuming results substantiates the effectiveness of the SymAE redatuming.

effects among earthquakes stemming from a common source region. To facilitate this, we selected stations that have recorded all the earthquakes in question, denoting their set using $\{Q1, \cdots, QN\}$. For instance, we start by examining three earthquakes: chl1 (Mw 8.8), chl2 (Mw 6.6), and chl3 (Mw 6.7), all originating from a source region located near the central coast of Chile (see Tab. 2). Following the EGF assumption, we presume that the path effects remain consistent across the three seismograms recorded at each station. Take, for example, the first station, Q1, where we denote the seismograms as

$$\underset{\text{chl1}\rightsquigarrow Q1}{\overset{\text{chl1\_}k}{D}}, \quad \underset{\text{chl2}\rightsquigarrow Q1}{\overset{\text{chl2\_}k}{D}} \quad \text{and} \quad \underset{\text{chl3}\rightsquigarrow Q1}{\overset{\text{chl3\_}k}{D}}. \tag{24}$$

Here, note that since the earthquakes are collocated, station Q1 is associated with the same focal-sphere pixel, indexed k, for all three earthquakes. Next, using Eq. 22, we created six virtual seismograms by exchanging the path effects between each pair of the three seismograms. For instance, when considering the chl2 and chl3 pair, the resulting virtual seismograms associated with the first station are given by:

$$\underset{\text{chl3}\rightsquigarrow Q1}{\overset{\text{chl2\_}k}{D}} \quad \text{and} \quad \underset{\text{chl2}\rightsquigarrow Q1}{\overset{\text{chl3\_}k}{D}}. \tag{25}$$

All the virtual seismograms are embedded into the SymAE's latent space in Fig. 6a — in this source path representation, the seismograms plotted on the diagonal using blue color are measured, while the off-diagonal seismograms plotted in black are virtual seismograms. We now perform a sanity check to ensure that redatuming is working properly and producing accurate results.

Sanity check: comparison between low-magnitude earthquakes with similar focal mechanism. To confirm the effectiveness of SymAE redatuming, we compared virtual seismograms with measured seismograms from low-magnitude earthquakes that share similar focal mechanisms and might have simpler source mechanisms, despite occurring at different times. Given that these earthquakes originate from the same source region, redatuming, involving the interchange of path effects between earthquakes, should mostly leave the seismograms unchanged. We quantify the changes post-redatuming, by computing the Pearson correlation distance between the measured and virtual seismograms with the same source information, for example, between

$$\left(\underset{\text{chl3}\rightsquigarrow Q1}{\overset{\text{chl2\_}k}{D}} \quad \text{and} \quad \underset{\text{chl2}\rightsquigarrow Q1}{\overset{\text{chl2\_}k}{D}} \right) \quad \text{or} \quad \left(\underset{\text{chl3}\rightsquigarrow Q1}{\overset{\text{chl3\_}k}{D}} \quad \text{and} \quad \underset{\text{chl2}\rightsquigarrow Q1}{\overset{\text{chl3\_}k}{D}} \right). \tag{26}$$

(a) Chile P   (b) Chile S   (c) Sumatra P   (d) Sumatra S

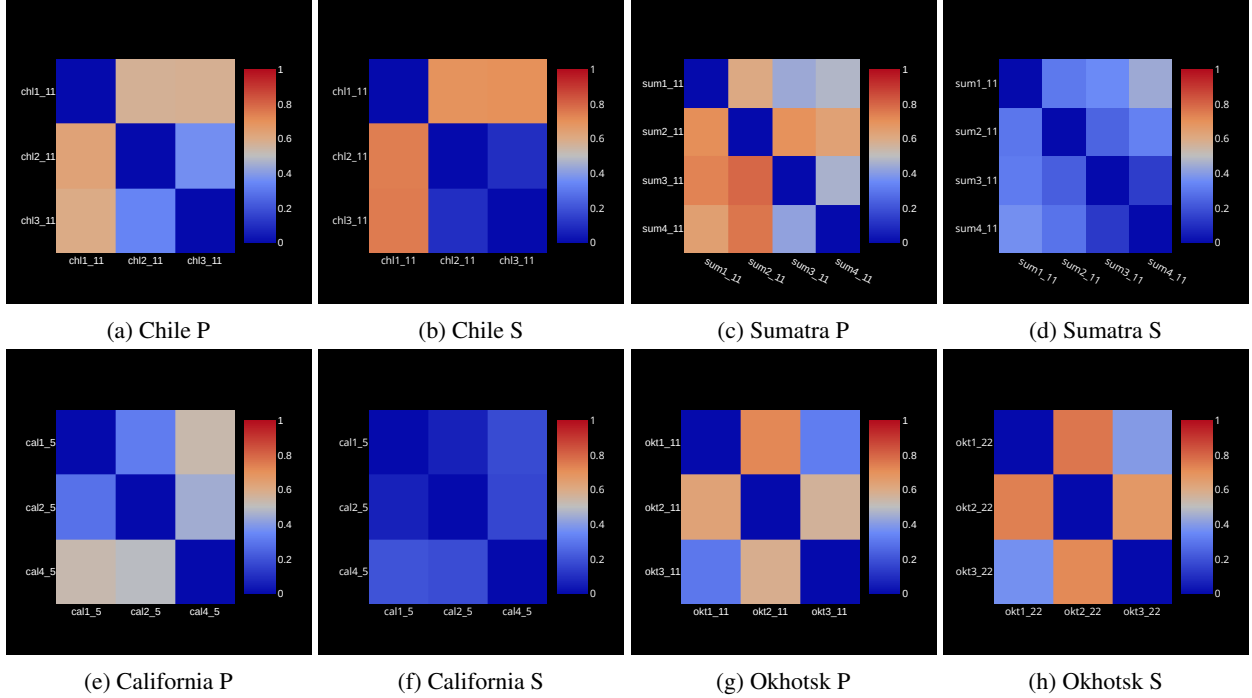(e) California P   (f) California S   (g) Okhotsk P   (h) Okhotsk S

Figure 7: The similarity $\mathcal{C}$ measured using Eq. 27 after redatuming path effects from one earthquake to another within the same source region. The similarity patterns are consistent for both P- and S-window seismogram training. The earthquake codes listed in Tab. 2.

As depicted in Fig. 6a, a lower correlation distance between these seismograms of chl2 and chl3 earthquakes corroborates the precision and dependability of the SymAE model in redatuming.

In line with Eq. 3, the path information contained in the seismograms of complex earthquakes, featuring multiple episodes, inherently differs statistically from that of simpler earthquakes characterized by a single rupture episode. Consequently, we anticipate a perturbed virtual seismogram, compared to the measured seismogram, when exchanging path effects between a simpler and a complex earthquake scenario. Interestingly, we observed a higher correlation distance when redatuming involved at least one earthquake with a distinct source mechanism. It is evident in Fig. 6a that the virtual seismograms generated using the path effects from the chl1 earthquake do not correlate with their respective measured seismograms. This discrepancy can be attributed to the fact that chl1 features more rupture episodes compared to the other two earthquakes [Yue et al., 2014, Delouis et al., 2010]. Towards the goal of quantifying the degree of similarity between two almost collocated earthquakes, we averaged the correlation distance over all the available shared stations:

$$\mathcal{C}(\text{chl2\_}k, \text{chl3\_}k) = \frac{1}{N} \sum_{Q \in \{Q1, \cdots, QN\}} \text{correlation\_distance} \left(\underset{\text{chl2}\rightsquigarrow Q}{\overset{\text{chl2\_}k}{D}}, \underset{\text{chl3}\rightsquigarrow Q}{\overset{\text{chl2\_}k}{D}} \right). \tag{27}$$

Note that $\mathcal{C}(\text{chl2\_}k, \text{chl3\_}k)$ may not necessarily be equal to $\mathcal{C}(\text{chl3\_}k, \text{chl2\_}k)$. We postulate that a lower value of $\mathcal{C}$ between two earthquakes implies a similarity in their source mechanism, indicating that these earthquakes conform to the EGF assumption. The values of $\mathcal{C}$ for various pairs of Chile earthquakes are depicted in Fig. 7a, illustrating the similarity between chl2 and chl3, with a distinct source mechanism for chl1 compared to chl2 and chl3.

Sanity check: comparison between P and S seismograms. We further validated this similarity quantification approach (Eq. 27) by utilizing SymAE models independently trained on S-windowed seismograms. The level of similarity between chl2 and chl3 can be more confidently confirmed by examining Figs. 6b and 7b. This independent analysis using S waves enhances our confidence in the accuracy and reliability of the redatuming approach.

During our analysis of Indian Ocean earthquakes, including sum1, sum2, sum3, and sum4 (Tab. 2), we observed a notable disparity in the similarity patterns between P and S waves. Previous studies by various authors [Fan and Shearer, 2016, Yue et al., 2017, Fan and Shearer, 2018] have examined the waveforms of these earthquakes. Our similarity metric $\mathcal{C}$ when applied to S waves showed lower values when the differences in half-duration between corresponding earthquakes were small, as indicated in Fig. 7d. However, an interesting difference arose in the case of

P-wave redatuming. Firstly, the $\mathcal{C}$ values involving the sum1 and sum2 earthquakes were unusually high compared to S-wave redatuming. Secondly, the $\mathcal{C}$ matrix, which represents the $\mathcal{C}$ values for each pair of earthquakes, exhibited an asymmetry, as shown in Fig. 7c. Specifically, the $\mathcal{C}$ values for pairs of earthquakes, especially involving sum1 and sum2, did not demonstrate uniformity when interchanged. For instance, $\mathcal{C}(\text{sum1\_11}, \text{sum4\_11})$ was greater than $\mathcal{C}(\text{sum4\_11}, \text{sum1\_11})$. This discrepancy between P and S analyses is attributed to the potential excitation of localized water reverberation phases by the high bathymetry gradient near the trench, examined by Yue et al. [2017]. Further research is needed to gain a comprehensive understanding of the reasons behind the observed disparities in P and S similarity patterns for these earthquakes.

Furthermore, our analysis extended to include the 2019 Ridgecrest earthquakes [Liu et al., 2019, Chen et al., 2020]. These earthquakes, predominantly characterized by strike-slip faulting, were labeled cal1, cal2, and cal4, as listed in Tab. 2. Notably, a reasonable degree of similarity was observed between the 6.5 Mw foreshock (cal2), which occurred on 4 July 2019 and ruptured along a NE-SW trending fault, and the 7.1 Mw mainshock (cal1), which predominantly ruptured along an NW-SE trending fault. In contrast, both of them displayed significant dissimilarity compared to another aftershock (cal4) with a magnitude of 5.5 (the largest aftershock in the sequence) that occurred eleven months later to the east of Ridgecrest. These observations are evident in the similarity matrices, visualized in Figs. 7e and 7f, which are generated using P and S waveforms, respectively.

Finally, we examined the similarity among three deep earthquakes that occurred in the subducting slab beneath the Sea of Okhotsk, specifically okt1, okt2, and okt3, as detailed in Tab. 2. Among these earthquakes, a previous study by Zhan et al. analyzed the waveforms of okt2, a Mw 6.7 earthquake, and suggested that it may be a supershear earthquake based on the empirical Green's function approach. However, to date, no definitive evidence has been obtained for supershear rupture in deep earthquakes. Our analysis reveals that okt1 and okt3 earthquakes exhibit a reasonable degree of similarity (see Figs. 7g and 7h), which is also evident from the raw seismograms, showing that these earthquakes are composed of at least three rupture episodes. In contrast, we observed that the supposedly supershear deep earthquake, okt2, is dissimilar to most of the other Okhotsk region earthquakes. In the following subsection, we provide a more detailed analysis of the okt2 earthquake and conclude that it has a complex source mechanism with a distinct stopping phase. In conclusion, this work holds the potential to substantially enhance our understanding of teleseismic similarity between earthquake sources. Furthermore, the similarity analysis facilitated by SymAE introduces a fundamentally different approach to studying earthquake self-similarity.

7.2 Uniformly Scattered Virtual Seismograms

Accurate directivity analysis of earthquakes often relies on seismograms free from scattering effects or those where azimuthal scattering dependencies are eliminated. Envelope stacking, a conventional technique, is employed to mitigate incoherent scattering in seismograms and facilitate azimuthal variation-based directivity analysis [Fan and Shearer, 2015, e.g.,]. Here, multiple seismograms associated with a given focal-sphere pixel can be summed based on their envelope functions (computated via the Hilbert transform), effectively reducing incoherent noise. Envelope stacking is vulnerable to travel time perturbations as it relies on the precise initial alignment of seismograms to enhance coherent source signatures. Travel time perturbations exceeding a quarter of the period can introduce phase misalignments, leading to signal interference during stacking and potentially causing a loss of resolution. Additionally, variations in the number of receivers associated with each focal-sphere pixel can lead to azimuthal differences in the envelope stacks, potentially resulting in misinterpretations and incorrect associations with directivity effects.

SymAE facilitates the generation of uniformly scattered virtual seismograms (USVS), providing a seismogram for each focal-sphere pixel. USVS are consistently influenced or perturbed by a specific path effect, therefore offer a superior alternative to envelope stacking for enhancing the quality of source signals and improving our ability to extract valuable directivity information. We represent the collection of USVS for pixels 1 to K of the okt1 earthquake by the vector

$$\underset{\text{okt1}\rightsquigarrow\text{Q}}{\overset{\text{okt1}}{U}} = \left[\underset{\text{okt1}\rightsquigarrow\text{Q}}{\overset{\text{okt1\_1}}{D}} ; \underset{\text{okt1}\rightsquigarrow\text{Q}}{\overset{\text{okt1\_2}}{D}} ; \underset{\text{okt1}\rightsquigarrow\text{Q}}{\overset{\text{okt1\_3}}{D}} ; \cdots ; \underset{\text{okt1}\rightsquigarrow\text{Q}}{\overset{\text{okt1\_K}}{D}} \right], \tag{28}$$

where each virtual seismogram is influenced by scattering effects associated with the common path okt1 $\rightsquigarrow$ Q. Similar to redatuming in Eq. 22, the virtual seismogram corresponding to the kth pixel is generated as follows

$$\underset{\text{okt1}\rightsquigarrow\text{Q}}{\overset{\text{okt1\_k}}{D}} = \text{Fuse}\left(\left[\text{SEnc}\left(\left[\underset{\text{okt1}\rightsquigarrow\text{B1}}{\overset{\text{okt1\_k}}{D}} ; \underset{\text{okt1}\rightsquigarrow\text{B2}}{\overset{\text{okt1\_k}}{D}} ; \cdots \underset{\text{okt1}\rightsquigarrow\text{BN}}{\overset{\text{okt1\_k}}{D}} \right] \right) ; \text{PEnc}\left(\underset{\text{okt1}\rightsquigarrow\text{Q}}{\overset{\text{okt1\_i}}{D}} \right) \right] \right). \tag{29}$$

In this context, the receivers $\{B1, \cdots, BN\}$ are linked to the kth pixel of the earthquake's focal sphere. However, it's important to note that the path effects may be selected from a receiver associated with any other pixel, such as i, and not necessarily the kth pixel. USVS for various earthquakes, whether using P or S seismograms, are illustrated in Figs. 8–11. It's important to note that while USVS are displacement seismograms, including both positive and

negative amplitudes, we have only plotted the envelopes of USVS in this paper to facilitate their comparison with the positive-only stacked envelopes. In our study, we generated stacked envelopes by aligning seismograms using the PREM (Preliminary Reference Earth Model) model. We now perform the following sanity checks using several examples to validate the generation of USVS using SymAE.

Sanity check: comparison with envelope stacking. To ensure the accuracy of USVS, we conducted a comparison with stacked envelopes across all our results presented in Figs. 8–11. In most cases, especially when analyzing P seismograms, it is apparent that the salient features of USVS align with those of stacked envelopes, providing further confirmation of the accuracy of USVS generation. For instance, consider the P-windowed stacked envelopes from the okt1 earthquake, as displayed in Fig. 8a. In Fig. 8b, it is evident that the USVS for the okt1 earthquake exhibit three distinct peaks between -10 to 30 seconds, corresponding to the three rupture episodes observed by Zhang et al. [2021]. These peaks are also identifiable in the case of the stacked envelopes. The study by Wei et al. [2013] provides insights into the complex nature of the okt1 earthquake. Additionally, we regenerated USVS using a different common path in Fig. 8c. The source features in the regenerated USVS are consistent with those of the previous USVS and the features of the stacked envelopes. We have noted that the source features remain consistent even when selecting the common path from different seismographic networks, as long as the noise in the selected seismogram is minimal. This consistency further underscores the reliability of the USVS approach.

Sanity check: comparison between P and S seismograms. Following a similar procedure as in the previous discussion, we independently trained SymAE using S-windowed seismograms, generated USVS, and observed that the source features are similar between P- and S-wave USVS. As expected, the S-wave USVS have lower resolution compared to the P-wave USVS. In the previous example of okt1, it's evident that the S-wave USVS (Figs. 8e and 8f) exhibit the three rupture episodes of the okt1 earthquake. This comparison further confirms the reliability of USVS.

Sanity check: consistency among pixels. The principle of scale separation discussed in Section 4 suggests that source features should exhibit smooth variations across pixels. Given that path effects are common among USVS seismograms and the variations in USVS are primarily due to source and directivity effects, we assessed whether the scale separation principle holds by examining the similarity among pixels in USVS. When employing different reference path effects, we observed a consistent pattern of similarity among virtual seismograms for all pixels across various earthquakes, as shown in Figs. 8–11. To illustrate this, we examined the case of another deep-focus earthquake, bon1, using both its P and S seismograms in Fig. 9. Notably, the stacked envelopes exhibited significant dissimilarities between pixels due to the challenges of incomplete removal of incoherent noise during the summation of independent receiver groups. This phenomenon highlights one of the limitations of traditional envelope stacking methods in contrast to the improved consistency observed in USVS. In to case of bon1 earthquake, the USVS, whether for P or S seismograms, consistently displayed source features such as the rapid rise of rupture, as previously identified by Ye et al. [2016]. This level of consistency further supports that the features learned by SymAE's symmetric encoder are more closely associated with the earthquake source or coherent scattering, rather than any incoherent path effects.

We conclude this section with a discussion of some of the intriguing earthquakes we have examined in this work. In Fig. 10, we present both P- and S-wave USVS for the supposed supershear deep event, okt2. USVS, with their superior resolution compared to envelope stacks, reveal a complex rise-and-fall pattern for both P- and S-wave scenarios. This observation aligns with our findings when analyzing the similarity among earthquakes in the Sea of Okhotsk region, as shown in Figs. 7g and 7h. The intricate fall pattern suggests the occurrence of multiple rupture episodes following the primary event. A notable feature is evident at $t = 0$, where a sharp decline is observed for most of the common path effects. This intriguing behavior warrants further investigation.

We analyzed two intermediate-depth events: 1) Rat Islands (rat) earthquake, known for its complex rise pattern with a Mw of 7.9, and 2) a lower magnitude earthquake (Mindanao; mnd) with a Mw of 6.1 in the Philippines. In both cases, the SCARDEC source time functions (plotted in Figs. 11a and 11d) appear relatively simple, where the mnd earthquake exhibits a single impulsive rupture with an approximate half-duration of 3 seconds. Our analysis of the rat earthquake (Figs. 11b–11c) reveals its complex rise pattern, characterized by a relatively weak seismic wave during the first 15 seconds, followed by short periods of intense energy release. This pattern aligns with findings in Ye et al. [2014]. For the mnd earthquake (Figs. 11e–11f), we observed a steep rise, along with a complicated stopping phase, particularly evident in the USVS during the time interval between the P and pP arrivals. These features are also observable in the envelope stack but are notably enhanced when using USVS. Both examples highlight the enhanced resolution provided by USVS, even when surface-related arrivals closely interfere with the source features, in contrast to deep earthquakes analyzed previously.

Although SymAE effectively uniformizes path effects across various pixels, it's crucial to note that path effects are not entirely eliminated but rather equalized among them. Therefore, the interpretation of features within USVS should be done cautiously, considering that they still bear the imprint of common path effects. However, differences observed among pixels can be largely attributed to source effects. In cases where specific features are consistently present in

USVS regardless of the common path used for their generation, it's a strong indication that these features are associated with the source. This principle guided our identification of source features discussed in this section, and that's why we often present USVS generated with two different independent common paths. Our analysis underscores the practicality of USVS in identifying rupture episodes in earthquakes occurring shortly after the initial energy release. This capability enhances their usefulness in comprehending complex earthquake sources, including their stopping phases.

USVS significantly enhance signal quality, offering superior resolution when compared to conventional methods like envelope stacking. This heightened resolution enables the precise extraction of rupture directivity information, especially for deep and intermediate-depth earthquakes with limited near-field observations. Specifically, USVS have the potential to serve as valuable input for time reversal methods, where the virtual seismograms can be simply back propagated in a homogeneous medium to image the source. In other words, a key advantage of USVS is that they do not require empirical time calibration, as they are inherently calibrated based on the common path chosen for their generation. Notably, USVS, generated using an ensemble of common path effects, offer a means for conducting uncertainty analysis when extracting source information. However, the full exploration of this application using USVS remains a promising area for future research endeavors.

8 Discussion

The capability to transfer path effects from one seismogram to another, as demonstrated in Eq. (22), is indeed a potent tool. It may be tempting to redatum source and path effects between distinct earthquakes, such as moving them from a shallow earthquake to a deep one. However, caution is warranted as the interpretation of redatumed seismograms in this context can be challenging. As previously discussed, surface-reflected phases and other nearby source-related reverberations cannot be effectively separated from the source information by the symmetric autoencoder. Therefore, in essence, while it is feasible to redatum coherent effects from a deep to a shallow earthquake using SymAE, the interpretation of the generated virtual seismogram can be complex due to the presence of ambiguities related to surface-reflected phases. In this paper, we have limited our experiments to scenarios where path effects are exchanged between two closely located earthquakes or between stations within the same earthquake.

In this study, we have demonstrated the applicability of SymAE to both P-wave and S-wave seismograms. Additionally, we propose the future generation of uniformly scattered virtual seismograms for various time windows of far-field seismograms, offering the possibility of obtaining additional constraints on rupture parameters through comparisons across different time windows. The challenge lies in developing novel imaging algorithms that can effectively leverage these high-resolution uniformly scattered virtual seismograms, especially considering that coherent scattering is not easily distinguishable from the source information. As discussed by Langston [1978], challenges arise when dealing with source characterization in the presence of surface-reflected arrivals. However, the analysis of USVS can significantly enhance our understanding of earthquake sources — they highlight the promising use of deep learning techniques in far-field seismic source imaging.

Our approach does have some limitations. It's crucial to carefully tune and validate the regularization strength of the path encoder, especially on a test dataset, where the sanity checks discussed in this paper provide valuable guidance during tuning. The development of additional sanity checks could further aid in the interpretation of results during redatuming processes. Furthermore, our approach relies on the scale separation between source and path effects. This implies that there are limitations on the maximum fault dimensions and the range of wave frequencies that can be effectively analyzed. In cases of earthquakes with large fault dimensions, with higher frequencies exceeding $0.1\,\mathrm{Hz}$ and near-field receiver locations, the Fraunhofer approximation is no longer valid. To address these scenarios, it is necessary to develop more advanced architectures for source imaging.

8.1 Conclusions

In this paper, we tackled the challenge of earthquake source characterization using far-field seismograms. Unlike conventional methods relying on the source-path convolutional model, we introduced a novel approaches by generating virtual seismograms through redatuming with a trained Symmetric Autoencoder (SymAE) network. Our results demonstrate that virtual seismograms can unveil valuable earthquake source information that might be concealed in traditional seismograms due to subsurface scattering effects. The SymAE network is specifically designed to disentangle source information from seismograms, leveraging the concept of scale separation, where source processes operate at a slower scale than path effects. Notably, the training process is unsupervised and doesn't demand labeled data, seismic arrival picking, or prior knowledge of subsurface scattering or source physics. Moreover, our approach is highly versatile, allowing training on all available far-field seismograms (both P- and S-wave) from multiple earthquakes and easily scaling with the amount of data. After validating the generation of virtual seismograms, we have highlighted two significant advantages of them in earthquake source characterization: 1. These virtual seismograms enable us to

(a) P raw envelopes with SCARDEC.

(b) USVS with common path I.

(c) USVS with common path II.

(d) S raw envelopes.

(e) USVS with common path I.
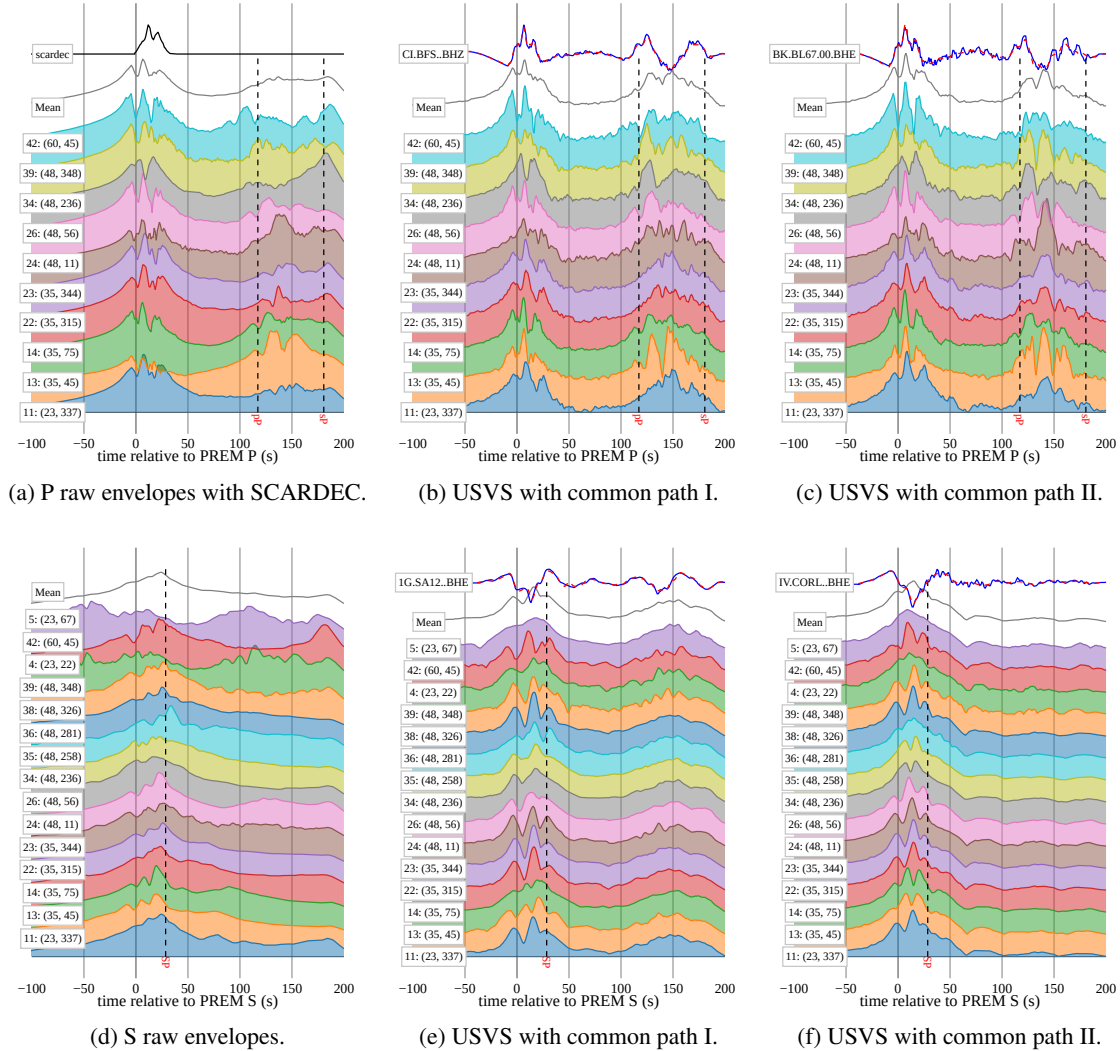
(f) USVS with common path II.

Figure 8: Time series data is presented for each pixel on the focal sphere of the Sea of Okhotsk earthquake, okt1, with labels indicating their respective pixel index, polar, and azimuthal angles. The subfigures depict: (a) and (d) P- and S-windowed stacked envelopes, respectively, after alignment with respect to the PREM model. (b) and (c) P uniformly scattered virtual seismograms (USVS) generated using Eq. 28 after extracting common path effects from two P seismograms. (e) and (f) USVS for S seismograms. Coherently scattered arrivals (pP, sP, SP) are roughly marked. In each plot, the seismograms from which common path effects are extracted are represented in blue, while their SymAE reconstructions are depicted using dashed red lines.

(a) P raw envelopes with SCARDEC.

(b) USVS with common path I.

(c) USVS with common path II.

(d) S raw envelopes.

(e) USVS with common path I.
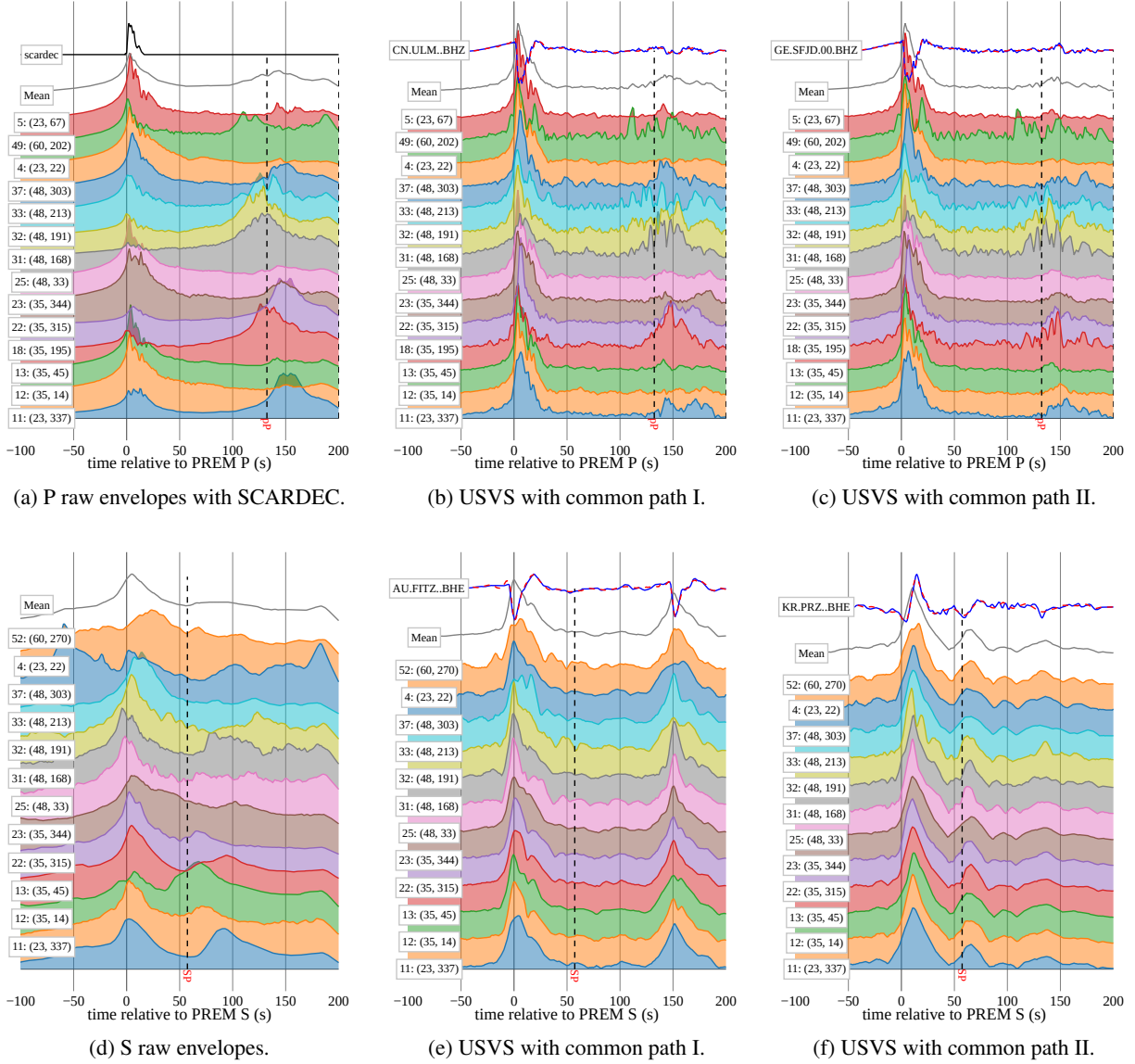
(f) USVS with common path II.

Figure 9: Same as Fig. 8 but for the Bonin (bon1) deep-focus earthquake. While traditional stacked envelopes exhibit considerable dissimilarities between pixels due to issues in eliminating incoherent noise, USVS maintain a remarkable consistency.
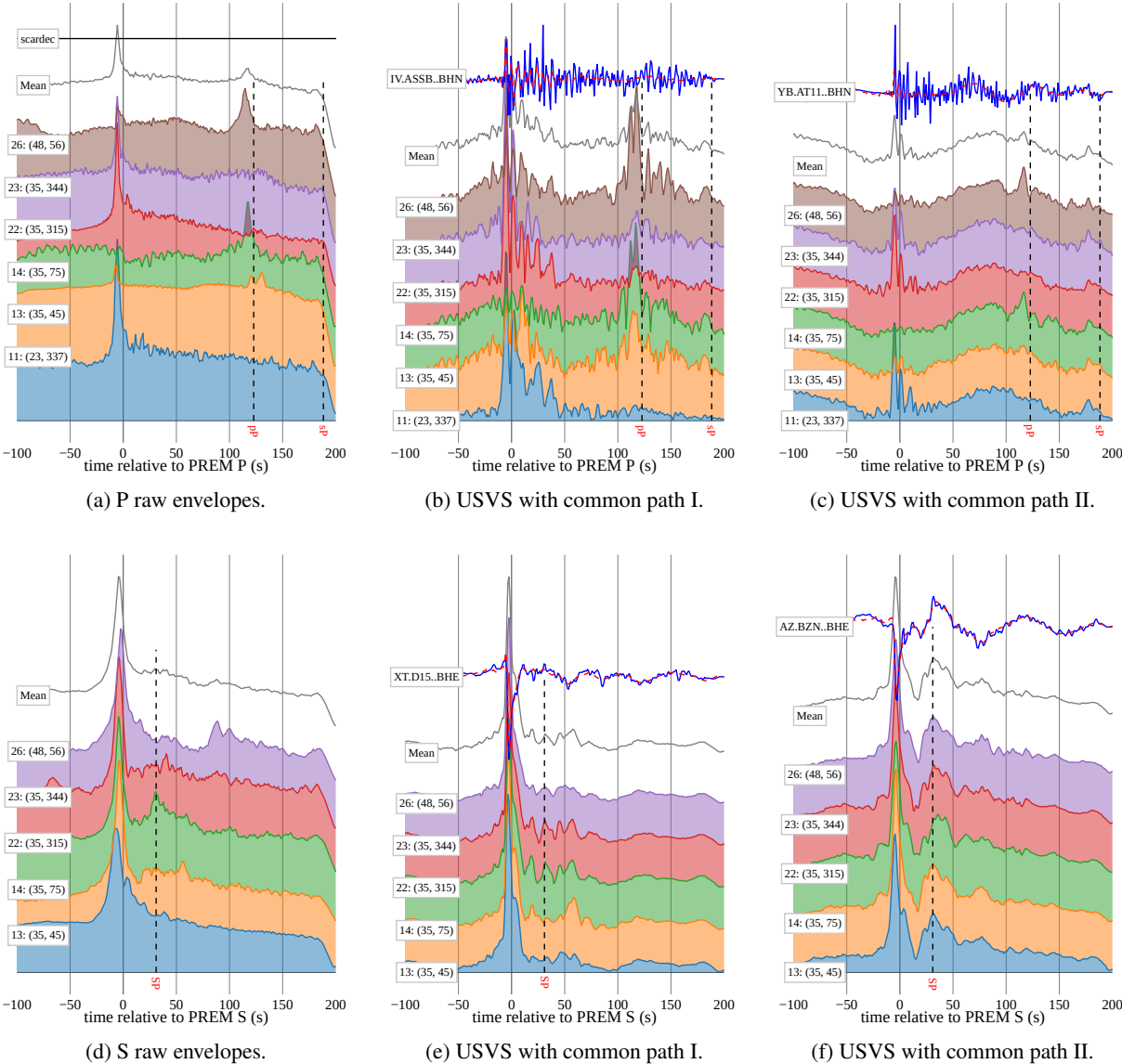
(a) P raw envelopes.

(b) USVS with common path I.

(c) USVS with common path II.

(d) S raw envelopes.

(e) USVS with common path I.

(f) USVS with common path II.

Figure 10: Similar to Fig. 8, this figure showcases virtual seismograms (USVS), focusing on another Sea of Okhotsk deep-focus earthquake (okt2). A complex rise-and-fall pattern, characteristic of multiple rupture episodes following the primary event, is clearly visible. USVS, with their high resolution, emphasize intricate source features, notably a sharp decline at zero relative time.

(a) P raw envelopes with SCARDEC.

(b) USVS with common path I.

(c) USVS with common path II.

(d) P raw envelopes with SCARDEC.

(e) USVS with common path I.
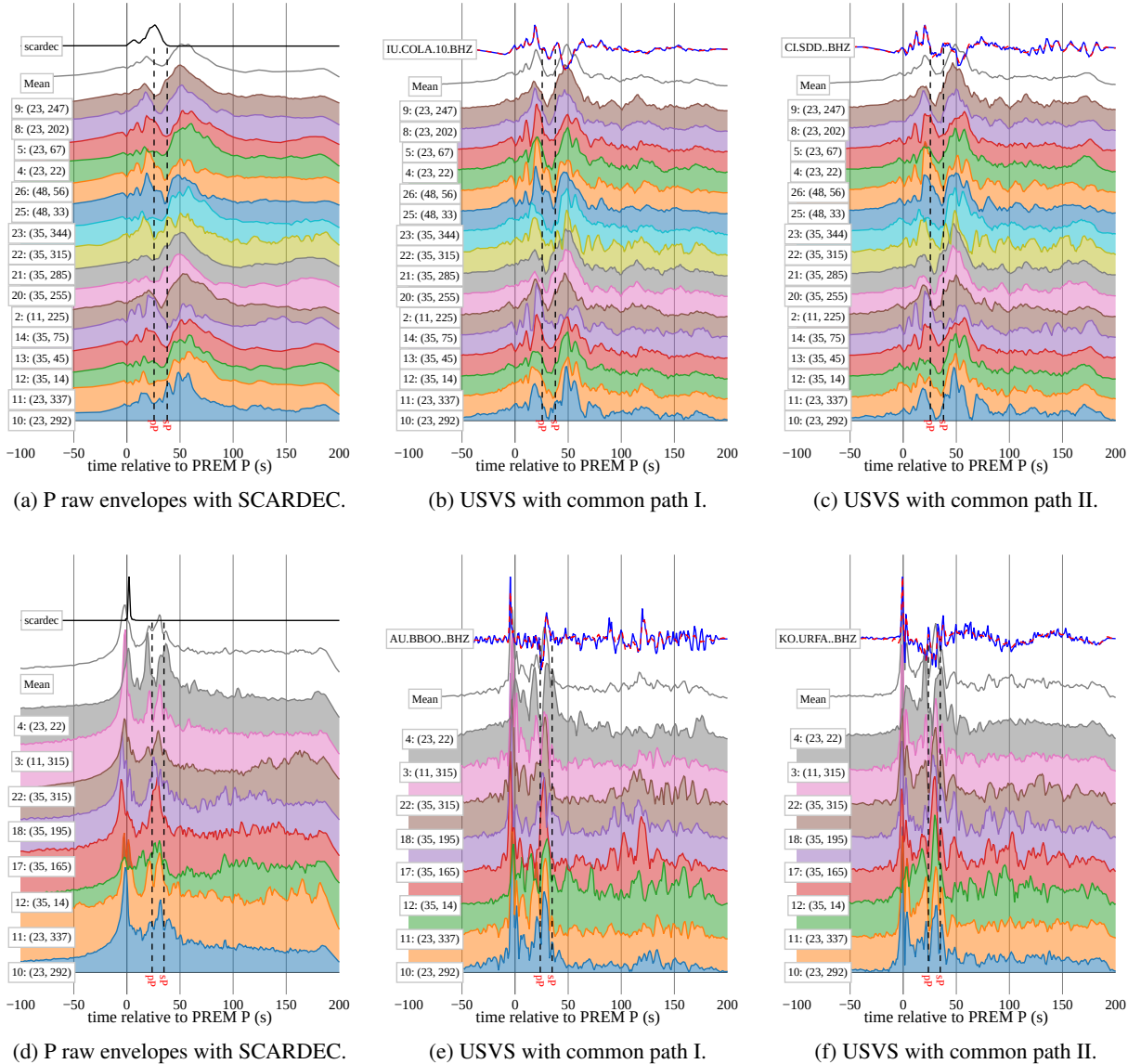
(f) USVS with common path II.

Figure 11: USVS analysis for intermediate-depth earthquakes, as shown in a manner similar to Fig. 8, allows us to uncover intriguing source features made more apparent by the improved resolution of USVS. (a)–(c) Rat Island earthquake (rat). (d)–(f) Mindanao, Philippines earthquake (mnd).

explore the similarity between earthquakes originating from the same source regions, shedding light on the relationships between seismic events. 2. Virtual seismograms, associated with different rays emanating from the source regions in various directions but engineered to possess identical path effects, offer substantial improvements in directivity analysis, can enhance the reliability of time reversal imaging, and provide better resolution while characterizing source rise and stopping phases. Overall, our innovative approach offers a promising avenue for advancing earthquake source characterization and opens up new possibilities for enhancing our understanding of seismic events.

9 Data and Resources

The data and resources utilized in this study encompassed seismograms obtained from multiple sources, which were primarily sourced from the Incorporated Research Institutions for Seismology Data Management Center (IRIS-DMC). Seismographic data was meticulously acquired from an array of institutions and seismic data centers, including AUSPASS, 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. Each of these sources played an integral role in assembling the comprehensive dataset necessary for training symmetric autoencoders. We would like to extend our appreciation to the International Federation of Digital Seismograph Networks (FDSN) and IRIS Data Services for facilitating access to the earthquake waveform data. 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. We would like to acknowledge the use of Plotly.js, a valuable tool that greatly facilitated the generation of the high-quality graphics and visualizations presented in this paper. We would also like to express our appreciation for PyGMT (GMT - Generic Mapping Tools). The half duration values in Tab. 2 and the beachball plots in Fig. 6a are taken from GCMT (Global Centroid Moment Tensor) catalog. Furthermore, we would like to acknowledge the use of HealPix, a versatile and widely adopted pixelization scheme for mapping the spherical surface of the Earth.

10 Acknowlegements

References

A. M. Dziewonski, T.-A. Chou, and J. 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. doi:https://doi.org/10.1029/JB086iB04p02825. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JB086iB04p02825`.

G. Ekström, M. Nettles, and A.M. Dziewoński. The global cmt project 2004–2010: Centroid-moment tensors for 13,017 earthquakes. *Physics of the Earth and Planetary Interiors*, 200-201:1–9, 2012. ISSN 0031-9201. doi:https://doi.org/10.1016/j.pepi.2012.04.002. URL `https://www.sciencedirect.com/science/article/pii/S0031920112000696`.

Hiroo Kanamori and Emily E Brodsky. The physics of earthquakes. *Reports on Progress in Physics*, 67(8):1429, jul 2004. doi:10.1088/0034-4885/67/8/R03. URL `https://dx.doi.org/10.1088/0034-4885/67/8/R03`.

Sunyoung Park and Miaki Ishii. Inversion for rupture properties based upon 3-D directivity effect and application to deep earthquakes in the Sea of Okhotsk region. *Geophysical Journal International*, 203(2):1011–1025, 2015.

Germán A. Prieto, Peter M. Shearer, Frank L. Vernon, and Debi Kilb. Earthquake source scaling and self-similarity estimation from stacking P and S spectra. *Journal of Geophysical Research: Solid Earth*, 109(B8), 2004. doi:https://doi.org/10.1029/2004JB003084. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2004JB003084`.

Charles A Langston and Donald V Helmberger. A procedure for modelling shallow dislocation sources. *Geophysical Journal International*, 42(1):117–130, 1975.

DJ Andrews. Objective determination of source parameters and similarity of earthquakes of different size. *Earthquake source mechanics*, 37:259–267, 1986.

Peter Bormann, Dmitry A Storchak, and Johannes Schweitzer. The IASPEI standard nomenclature of seismic phases. In *New Manual of Seismological Observatory Practice 2 (NMSOP-2)*, pages 1–20. Deutsches GeoForschungsZentrum GFZ, 2013.

A. Kaiser, N. Balfour, B. Fry, C. Holden, N. Litchfield, M. Gerstenberger, E. D'Anastasio, N. Horspool, G. McVerry, J. Ristau, S. Bannister, A. Christophersen, K. Clark, W. Power, D. Rhoades, C. Massey, I. Hamling, L. Wallace, J. Mountjoy, Y. Kaneko, R. Benites, C. Van Houtte, S. Dellow, L. Wotherspoon, K. Elwood, and K. Gledhill. The 2016 Kaikōura, New Zealand, Earthquake: Preliminary Seismological Report. *Seismological Research Letters*, 88(3): 727–739, 04 2017. ISSN 0895-0695. doi:10.1785/0220170018. URL `https://doi.org/10.1785/0220170018`.

Kjetil F Kaaresen and Tofinn Taxt. Multichannel blind deconvolution of seismic signals. *Geophysics*, 63(6):2093–2107, 1998.

Pawan Bharadwaj, Chunfang Meng, Aimé Fournier, Laurent Demanet, and Mike Fehler. Redshift of earthquakes via focused blind deconvolution of teleseisms. *Geophysical Journal International*, 223(3):1864–1878, 2020.

Heidi Houston. *Deep Earthquakes*, pages 329–354. 12 2015. ISBN 9780444538031. doi:10.1016/B978-0-444-53802-4.00079-8.

Stephen H Hartzell. Earthquake aftershocks as Green's functions. *Geophysical Research Letters*, 5(1):1–4, 1978.

Kojiro Irikura. Prediction of strong acceleration motion using empirical Green's function. In *Proc. 7th Japan Earthq. Eng. Symp*, volume 151, pages 151–156, 1986.

Alexandre P. Plourde and Michael G. Bostock. Multichannel Deconvolution for Earthquake Apparent Source Time Functions. *Bulletin of the Seismological Society of America*, page ssabull;0120170015v1, June 2017. ISSN 0037-1106, 1943-3573. doi:10.1785/0120170015.

Qimin Wu, Xiaowei Chen, and Rachel E. Abercrombie. Source Complexity of the 2015 Mw 4.0 Guthrie, Oklahoma Earthquake. *Geophysical Research Letters*, 46(9):4674–4684, 2019. ISSN 1944-8007. doi:10.1029/2019GL082690.

V. Lanza, D. Spallarossa, M. Cattaneo, D. Bindi, and P. Augliera. Source parameters of small events using constrained deconvolution with empirical Green's functions. *Geophysical Journal International*, 137(3):651–662, June 1999. ISSN 0956-540X. doi:10.1046/j.1365-246x.1999.00809.x.

Lawrence Hutchings and Gisela Viegas. Application of empirical Green's functions in earthquake source, wave propagation and strong ground motion studies. *Earthquake Research and Analysis-New Frontiers in Seismology*, pages 87–140, 2012.

Maximiliano Bezada and Eugene Humphreys. Contrasting rupture processes during the April 11, 2010 deep-focus earthquake beneath Granada, Spain. *Earth and Planetary Science Letters*, 353–354:38–46, 11 2012. doi:10.1016/j.epsl.2012.08.001.

Zhongwen Zhan, Donald V Helmberger, Hiroo Kanamori, and Peter M Shearer. Supershear rupture in a Mw 6.7 aftershock of the 2013 Sea of Okhotsk earthquake. page 5.

Giuliana Alessio, Cataldo Godano, and Antonella Gorini. A low magnitude seismic sequence near Isernia (Molise, Central Italy) in January 1986. *Pure and Applied Geophysics*, 134(2):243–260, Sep 1990. ISSN 1420-9136. doi:10.1007/BF00877000. URL `https://doi.org/10.1007/BF00877000`.

M. Vallée, J. Charléty, A. M. G. Ferreira, B. Delouis, and J. 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: Wave deconvolution and earthquake parameters. *Geophysical Journal International*, 184 (1):338–358, January 2011. ISSN 0956540X. doi:10.1111/j.1365-246X.2010.04836.x.

Martin Vallée and Vincent Douet. A new database of source time functions (STFs) extracted from the SCARDEC method. *Physics of the Earth and Planetary Interiors*, 257:149–157, August 2016. ISSN 00319201. doi:10.1016/j.pepi.2016.05.012.

Miaki Ishii, Peter M. Shearer, Heidi Houston, and John E. Vidale. Extent, duration and speed of the 2004 sumatra–andaman earthquake imaged by the hi-net array. *Nature*, 435(7044):933–936, June 2005. doi:10.1038/nature03675. URL `https://doi.org/10.1038/nature03675`.

Eric Kiser and Miaki Ishii. Combining seismic arrays to image the high-frequency characteristics of large earthquakes. *Geophysical Journal International*, 188(3):1117–1128, January 2012. doi:10.1111/j.1365-246x.2011.05299.x. URL `https://doi.org/10.1111/j.1365-246x.2011.05299.x`.

Hongyu Zeng, Shengji Wei, and Ares Rosakis. A travel-time path calibration strategy for back-projection of large earthquakes and its application and validation through the segmented super-shear rupture imaging of the 2002 Mw 7.9 Denali earthquake. *Journal of Geophysical Research: Solid Earth*, 127(6):e2022JB024359, 2022.

Pawan Bharadwaj, Matthew Li, and Laurent Demanet. Redatuming physical systems using symmetric autoencoders. *Physical Review Research*, 4(2):023118, 2022.

Kees Wapenaar and Jacob Fokkema. Green's function representations for seismic interferometry. *Geophysics*, 71(4): SI33–SI46, 2006.

Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. A survey of transfer learning. *Journal of Big data*, 3(1):1–40, 2016.

Han Yue, Jorge C. Castellanos, Chunquan Yu, Lingsen Meng, and Zhongwen Zhan. Localized water reverberation phases and its impact on backprojection images. *Geophysical Research Letters*, 44(19):9573–9580, 2017. doi:https://doi.org/10.1002/2017GL073254. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL073254`.

Keiiti Aki and Paul G Richards. *Quantitative Seismology*. 2002.

R. Madariaga. Seismic Source Theory. In *Treatise on Geophysics*, pages 51–71. Elsevier, 2015. ISBN 978-0-444-53803-1. doi:10.1016/B978-0-444-53802-4.00070-1.

Laurent Demanet. Waves and Imaging. `https://math.mit.edu/icg/resources/notes367.pdf`, 2012. Accessed: 2012-12-01.

Fredric J Harris. On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, 66(1):51–83, 1978.

Krzysztof M Gorski, Eric Hivon, Anthony J Banday, Benjamin D Wandelt, Frode K Hansen, Mstvos Reinecke, and Matthia Bartelmann. Healpix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere. *The Astrophysical Journal*, 622(2):759, 2005.

Kasra Hosseini and Karin Sigloch. obspyDMT: A Python toolbox for retrieving and processing of large seismological datasets. 2017.

Carl Doersch. Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*, 2016.

Bin Dai, Yu Wang, John Aston, Gang Hua, and David Wipf. Connections with robust PCA and the role of emergent sparsity in variational autoencoder models. *Journal of Machine Learning Research*, 19(41):1–42, 2018. URL `http://jmlr.org/papers/v19/17-704.html`.

Jack Klys, Jake Snell, and Richard Zemel. Learning latent subspaces in variational autoencoders. *arXiv preprint arXiv:1812.06190*, 2018.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbhakhsh, Barnabás Póczos, Ruslan Salakhutdinov, and Alexander J. Smola. Deep sets. *Advances in Neural Information Processing Systems*, (ii):3392–3402, apr 2017. ISSN 10495258. URL `http://arxiv.org/abs/1703.06114`.

Maximilian Ilse, Jakub M. Tomczak, and Max Welling. Attention-based deep multiple instance learning. *35th International Conference on Machine Learning, ICML 2018*, 5(Mil):3376–3391, 2018.

Sida Wang and Christopher Manning. Fast dropout training. In *international conference on machine learning*, pages 118–126. PMLR, 2013.

Durk P Kingma, Tim Salimans, and Max Welling. Variational dropout and the local reparameterization trick. *Advances in neural information processing systems*, 28:2575–2583, 2015.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

Han Yue, Thorne Lay, Luis Rivera, Chao An, Christophe Vigny, Xiaopeng Tong, and Juan Carlos Báez Soto. Localized fault slip to the trench in the 2010 Maule, Chile Mw = 8.8 earthquake from joint inversion of high-rate GPS, teleseismic body waves, InSAR, campaign GPS, and tsunami observations. *Journal of Geophysical Research: Solid Earth*, 119(10):7786–7804, 2014. doi:https://doi.org/10.1002/2014JB011340. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014JB011340`.

Bertrand Delouis, Jean-Mathieu Nocquet, and Martin Vallée. Slip distribution of the february 27, 2010 Mw = 8.8 Maule earthquake, central Chile, from static and high-rate GPS, InSAR, and broadband teleseismic data. *Geophysical Research Letters*, 37(17), 2010. doi:https://doi.org/10.1029/2010GL043899. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010GL043899`.

Wenyuan Fan and Peter M. Shearer. Fault interactions and triggering during the 10 January 2012 Mw 7.2 Sumatra earthquake. *Geophysical Research Letters*, 43(5):1934–1942, 2016. doi:https://doi.org/10.1002/2016GL067785. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016GL067785`.

Wenyuan Fan and Peter M. Shearer. Coherent seismic arrivals in the P wave coda of the 2012 Mw 7.2 Sumatra earthquake: Water reverberations or an early aftershock? *Journal of Geophysical Research: Solid Earth*, 123(4): 3147–3159, 2018. doi:https://doi.org/10.1002/2018JB015573. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2018JB015573`.

Chengli Liu, Thorne Lay, Emily E. Brodsky, Kelian Dascher-Cousineau, and Xiong Xiong. Coseismic rupture process of the large 2019 Ridgecrest earthquakes from joint inversion of geodetic and seismological observations. *Geophysical Research Letters*, 46(21):11820–11829, 2019. doi:https://doi.org/10.1029/2019GL084949. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL084949`.

Kejie Chen, Jean-Philippe Avouac, Saif Aati, Chris Milliner, Fu Zheng, and Chuang Shi. Cascading and pulse-like ruptures during the 2019 Ridgecrest earthquakes in the eastern california shear zone. *Nature Communications*, 11(1): 22, 2020. doi:10.1038/s41467-019-13750-w. URL `https://doi.org/10.1038/s41467-019-13750-w`.

Wenyuan Fan and Peter M. Shearer. Detailed rupture imaging of the 25 April 2015 Nepal earthquake using teleseismic P waves. *Geophysical Research Letters*, 42(14):5744–5752, 2015. doi:https://doi.org/10.1002/2015GL064587. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015GL064587`.

Hao Zhang, Suzan van der Lee, Craig R. Bina, and Zengxi Ge. Deep dehydration as a plausible mechanism of the 2013 Mw 8.3 Sea of Okhotsk deep-focus earthquake. *Frontiers in Earth Science*, 9, 2021. ISSN 2296-6463. doi:10.3389/feart.2021.521220. URL `https://www.frontiersin.org/articles/10.3389/feart.2021.521220`.

Shengji Wei, Don Helmberger, Zhongwen Zhan, and Robert Graves. Rupture complexity of the Mw 8.3 Sea of Okhotsk earthquake: Rapid triggering of complementary earthquakes? *Geophysical Research Letters*, 40(19):5034–5039, 2013. doi:https://doi.org/10.1002/grl.50977. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/grl.50977`.

Lingling Ye, Thorne Lay, Zhongwen Zhan, Hiroo Kanamori, and Jin-Lai Hao. The isolated 680 km deep 30 May 2015 Mw 7.9 Ogasawara (Bonin) islands earthquake. *Earth and Planetary Science Letters*, 433:169–179, 01 2016. doi:10.1016/j.epsl.2015.10.049.

Lingling Ye, Thorne Lay, and Hiroo Kanamori. The 23 June 2014 Mw 7.9 Rat islands archipelago, Alaska, intermediate depth earthquake. *Geophysical Research Letters*, 41(18):6389–6395, 2014. doi:https://doi.org/10.1002/2014GL061153. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014GL061153`.

Charles A Langston. Moments, corner frequencies, and the free surface. *Journal of Geophysical Research: Solid Earth*, 83(B7):3422–3426, 1978.