Teleseismic Rayleigh wave Helmholtz tomography in northeastern Tibet using physics-informed neural networks

Yunpeng Chen[1], Sjoerd A.L. de Ridder[2*], Sebastian Rost[2], Zhen Guo[3], Yongshun Chen[3†], and Lin Gan[1,4‡]

[1]Hetao Institute of Mathematics and Interdisciplinary Sciences, Shenzhen 518017, China
[2]School of Earth and Environment, University of Leeds, LS29JT, Leeds, UK
[3]Department of Ocean Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China
[4]Yau Mathematical Sciences Center, Tsinghua University, Beijing, 100084, China

Key Points:

- PINN-based teleseismic Helmholtz tomography recovers reliable phase velocity from far fewer events, validated on a synthetic model.
- Phase velocity maps support major structural contrasts beneath northeastern Tibet, with lateral heterogeneity diminishing at longer periods.
- Results agree well with conventional Helmholtz and ambient noise tomography, with complementary depth sensitivity across periods.

*Sjoerd A.L. de Ridder: s.deridder@leeds.ac.uk
†Yongshun Chen: johnyc@sustech.edu.cn
‡Lin Gan: lingan@tsinghua.edu.cn

Abstract

In this study, we apply a physics-informed neural network (PINN)-based teleseismic Rayleigh wave Helmholtz tomography (pinnTHT) to investigate the crustal and upper mantle structure beneath the northeastern Tibetan Plateau. The method incorporates the Helmholtz equation directly into the loss function, allowing phase velocity, amplitude, and traveltime fields to be jointly constrained. We represent the physical fields as continuous, grid-free functions whose spatial derivatives are evaluated by automatic differentiation rather than by finite differences on a fixed grid. We apply this approach to teleseismic surface wave data recorded by a dense seismic array and analyze Rayleigh wave phase velocities at periods of 20-80 s. Our results reveal significant lateral heterogeneity in velocity structure that correlates well with major tectonic features across the study area. The method achieves reliable resolution using only approximately 20% of the observational data compared to traditional approaches, and a synthetic ground-truth test confirms that pinnTHT recovers a known velocity model more accurately than conventional Helmholtz tomography applied to the full dataset. Comparisons with ambient noise tomography results at overlapping periods show consistent velocity anomaly patterns, indicating that the combination of earthquake and ambient noise data provides a systematic understanding of phase velocity structure across multiple periods. These results demonstrate that PINN-based teleseismic Helmholtz tomography offers a memory efficient and robust approach for investigating deep Earth structure, particularly beneficial in regions with sparse or unevenly distributed data.

Plain Language Summary

Teleseismic surface wave tomography is a vital tool for studying the Earth's interior using seismic records from earthquakes. In this study, we develop a new method to obtain subsurface images by combining deep learning with physics principles. Traditional methods require large amounts of earthquake data to produce clear images, but our approach, which we call pinnTHT, can achieve similar quality results with much less data. We applied this method to the northeastern Tibet. Our results reveal velocity variations in the crust and upper mantle, which help us understand this region's tectonic evolution and dynamics. When comparing our results with those from other methods, we find a good consistency about the structure beneath Tibet. This method could be particularly useful in regions where earthquake data are limited, helping us better understand Earth's structure in previously challenging areas.

1 Introduction

Unraveling the structure of the crust and upper mantle is crucial for understanding a wide range of the Earth's geodynamic processes, including plate tectonics, mantle convection, and earthquake generation. Among various geophysical techniques, seismic surface wave tomography has emerged as a pivotal tool for obtaining the subsurface velocity structure of the crust and upper mantle (Woodhouse & Dziewonski, 1984; Anderson, 1987; Forsyth et al., 2005; Sabra et al., 2005; Shapiro et al., 2005; Nishida et al., 2009; Schaeffer & Lebedev, 2013). Surface waves are particularly sensitive to shear wave velocity variations within the crust and upper mantle, due to their dispersive nature and depth-dependent sampling properties. These properties allow surface waves with different periods to provide valuable constraints on the velocity structure at different depths, enabling higher vertical resolution in the crust and upper mantle compared to body wave tomography (Golos et al., 2018). Surface wave tomography can generate phase and/or group velocity models using either earthquake data (Zhang et al., 2011) or ambient noise (Li et al., 2014). Tele-

seismic surface wave tomography leverages the long-period surface waves generated by global earthquakes. As these waves propagate through the Earth's interior, they carry information about the subsurface velocity structure over extended distances. In contrast to ambient noise tomography, which relies on shorter-period surface waves, teleseismic tomography can reveal deeper structures into the upper mantle due to the longer wavelengths involved, making it a powerful tool for investigating deeper geological features and geodynamic processes.

With the development of dense arrays, a large number of high-resolution imaging methods using surface wave have been developed, including the two-station method (Sato, 1955; Yao et al., 2005, 2006), two-plane-wave surface wave tomography (Forsyth et al., 1998, 2005; Li & Li, 2015), and eikonal and Helmholtz tomography (Lin et al., 2009; Lin & Ritzwoller, 2011; Jin & Gaherty, 2015), which are widely used in analyzing surface wave signals generated by earthquakes. Among these methods, eikonal and Helmholtz tomography has emerged as particularly effective approaches. These methods involve wavefront tracking and gradient calculation based on the eikonal and Helmholtz equations, enabling them to directly obtain surface wave phase/group velocity without conventional forward and inverse processes (Lin et al., 2009). Traditional eikonal tomography allows direct measurement of phase velocity but neglects the amplitude Laplacian term. Thus, the eikonal equation can be regarded as the Helmholtz formulation with the frequency-independent scattering term omitted, rather than strictly as a high-frequency approximation (Lehujeur & Chevrot, 2020). To correct for this bias, Helmholtz tomography introduces an amplitude correction term, theoretically providing more accurate phase velocity estimates. The theoretical properties of surface wave phase velocities and amplitudes in generic velocity structures have been quantitatively investigated in previous studies (Tsai & Atiganyanun, 2014). Recent developments have further improved the numerical implementation of Helmholtz tomography through iterative strategies and smooth amplitude–phase field representations (Sergeant et al., 2025). Our framework is conceptually similar in minimizing a loss function that combines a data misfit with physical-equation constraints, but relies on the Helmholtz equation alone rather than additionally exploiting the transport equation (Sergeant et al., 2025). However, in practical implementation, reconstruction of the traveltime field relies on simple interpolation schemes, which introduce errors particularly when station spacing is irregular or station coverage is sparse (L. Li et al., 2023). More critically, the Helmholtz equation requires computing the Laplacian of the amplitude field. When applied to discrete and noisy station data, standard finite-difference approximations for this second-order derivative are numerically unstable and highly susceptible to computational artifacts. Such computational challenges have prompted the exploration of alternative methodologies capable of directly learning complex relationships from data while respecting the underlying physical principles.

The application of deep learning has provided new opportunities for development in seismology. Deep neural networks (DNNs), composed of multiple layers of interconnected neurons, are able to learn to represent and extract complex patterns from massive amounts of seismic data. For instance, DNNs have been widely used in seismic arrival time picking (Zhu & Beroza, 2019; Mousavi et al., 2020; Zhu et al., 2023), earthquake and noise classification (Seydoux et al., 2020; Maguire et al., 2024; Zhu et al., 2024) and earthquake location (Zhang et al., 2020; Elsayed et al., 2023; Zhang et al., 2025). Beyond these applications, the powerful nonlinear fitting and regression abilities of DNNs have led to substantial advancements in seismic tomography, enhancing both efficiency and precision (Araya-Polo et al., 2018; Bianco et al., 2019; Yang & Ma, 2019; Fu et al., 2021; Liu et al., 2021; Jo & Ha, 2023; Muller et al., 2023; Mousavi et al., 2024; Xue et al., 2025). A key feature of DNNs is their use of automatic differentiation, which eliminates the need for discretization (Baydin

et al., 2018). Moreover, deep learning offers flexibility to operate without relying on mesh grids. Neural networks (NNs) function as continuous, grid-free representations, capable of evaluating on new data points regardless of their distribution after training (Grossmann et al., 2024). These capabilities enable DNNs to handle the partial differential equations (PDEs) in tomographic problems more effectively. Compared to traditional grid-based methods, deep learning provides a more adaptable and robust approach to solving the eikonal equation.

Physics-informed neural networks (PINNs) represent a novel framework for solving various PDEs using deep learning techniques (Raissi et al., 2019). This approach combines the strengths of data-driven and theory-based modeling, enabling the training of accurate and generalized models even with limited data. It offers several advantages over purely data-driven models, including improved robustness to observation errors and enhanced generalization beyond the training data. In seismology, PINNs have demonstrated remarkable potential in various applications, such as wave equation simulations (Karimpouli & Tahmasebi, 2020; Moseley et al., 2020; bin Waheed et al., 2021; Song et al., 2021; Huang & Alkhalifah, 2022a, 2022b; Song & Wang, 2023; Wu et al., 2023; Chai et al., 2024; Ren et al., 2024), traveltime tomography (Agata et al., 2023; Gou et al., 2023; Chen et al., 2022; Chen et al., 2023; Li et al., 2025; Song et al., 2025), wavefield reconstruction inversion (Song & Alkhalifah, 2021) and full waveform inversion (Rasht-Behesht et al., 2022; Yang & Ma, 2023; Lu et al., 2024), which include forward and inverse problems in isotropic and anisotropic media and in time and frequency-domain. During these applications, the framework and architecture of PINNs have also been improved using Bayesian theory (Agata et al., 2023; Gou et al., 2023), Fourier features (Song & Wang, 2023; Chai et al., 2024), recurrent neural networks (Lu et al., 2024) and generative adversarial networks (Yang & Ma, 2023). Particularly, Chen et al. (2022) and Chen et al. (2023) were the first to apply PINNs to dense arrays using ambient noise eikonal tomography for both the isotropic and anisotropic case, demonstrating their practicality on large-scale, real-world datasets. However, PINN-based tomography to analyze long-period surface waves and earthquake data in complex tectonic regions such as Tibet remains unexplored. The continuous, grid-free nature of PINNs, combined with automatic differentiation, offers a natural way to mitigate the Laplacian instability of Helmholtz tomography. By representing the amplitude field as a continuous NN, PINNs allow for the analytical evaluation of spatial derivatives ($\nabla^2 A$), thereby bypassing the numerical singularities and discretization errors inherent in traditional grid-based methods. This continuous parameterization eliminates discretization and interpolation artifacts associated with mesh-based representations. However, it does not improve the intrinsic spatial resolution, which is fundamentally limited by the seismic wavelength and the spatial sampling of the data. The effective resolution is of the order of the average wavelength of the analyzed frequency band (roughly 65–90 km at 20–25 s, increasing to 250–340 km at 60–80 s), which at the shortest periods is comparable to the 40–70 km station spacing.

In this study, we introduce a PINN-based teleseismic Helmholtz tomography method (pinnTHT) to simultaneously retrieve Rayleigh wave phase velocities across multiple frequencies. We apply this teleseismic Helmholtz tomography technique to data collected from a dense network of stations on the northeastern Tibetan Plateau. To assess the performance and physical behavior of the proposed method, we systematically compare the pinnTHT with those obtained from conventional Helmholtz tomography, a PINN-based eikonal formulation and ambient noise tomography derived from independent datasets. These comparisons provide a comprehensive framework for evaluating the robustness, consistency and complementary strengths of different surface wave tomography approaches.

2 Data

The seismic records were collected from ChinArray-Himalaya II (S. Li et al., 2022; Wu et al., 2023), a dense seismic array consisting of 676 stations deployed across the northeastern Tibetan Plateau and its surrounding regions. These stations were deployed between September 2013 and March 2016, with inter-station spacing from 40 to 70 km (Figure 1). The earthquake events were selected based on the following criteria: epicentral distances between 20° and 160°, source depths less than 50 km, and magnitude greater than Ms 5.8. Data preprocessing of the Z-component included cutting into 1-day segments, resampling to 1 Hz and removing mean value, trends and instrument responses. Following these criteria and preprocessing steps, we identified 119 teleseismic events suitable for surface wave tomography analysis. The spatial distribution of these earthquakes (Figure 2) shows generally good backazimuthal coverage, though with a higher concentration of events from the northeast directions and relatively fewer from the northwest. Notably, the pinnTHT results were obtained utilizing only approximately one-fifth of the total teleseismic events, yet they achieve reasonable and consistent phase velocity across multiple periods (see Figure S1 for details of event number per period). This efficiency in data utilization has been noted before by Chen et al. (2023), whose convergence test showed that the recovered maps stabilize beyond a modest number of sources. The event subset is not selected at random. For each period, we retain all available stations but only those events recorded by at least 400 stations and providing a relatively uniform backazimuthal coverage (Figure S1). For each selected event, the traveltime and amplitude constraints are taken from the same set of available stations, and this available station set may itself vary between events and periods.

Gee and Jordan (1992) proposed a generalized seismological data functional (GSDF) that employs cross-correlations between observed and synthetic waveforms to extract phase and amplitude from individual seismograms. A five-parameter Gaussian wavelet was then applied to the narrow-band filtered cross-correlation functions to extract a series of dispersion parameters. This method has been widely used to construct models of Earth's crust and upper mantle (Gaherty & Jordan, 1995; Gaherty et al., 1996; Gaherty, 2001, 2004; P. Chen et al., 2007). Jin and Gaherty (2015) extended the GSDF method by applying cross-correlation functions between neighboring stations to derive dispersion parameters from observed waveforms rather than synthetic waveforms, and then applied it to multi-channel analysis. We use this automated surface wave measuring system (ASWMS) to measure phase delays between all nearby station pairs. The five-parameter Gaussian wavelet used in this approach is defined as (Gee & Jordan, 1992):

$$F_i * W_c C(t) \approx A G_a[\sigma(t - t_g)] cos[w(t - t_p)], \tag{1}$$

where $F_i * W_c C(t)$ represents the filtered correlograms obtained by convolving each narrow-band filter F_i (centered at frequency ω_i) with the windowed cross-correlogram $W_c C(t)$, W_c is a window function, $C(t) = S_1 * W_s S_2$ represents the cross-correlation function between record S_1 and windowed record $W_s S_2$, G_a is the Gaussian function $e^{-x^2/2}$, five parameters are A, σ, ω, t_g and t_p, which respectively represent the amplitude factor, the half-bandwidth, the center frequency, the frequency-dependent group delay time and the phase delay time. These parameters are estimated by fitting Equation 1 to the filtered and windowed cross-correlogram using a weighted least-squares method. The phase delay t_p is further corrected for cycle-skipping and bias caused by the window function W_s (Gee & Jordan, 1992; Jin & Gaherty, 2015):

$$\delta\tau_p = t_p - \widetilde{t}_p, \tag{2}$$

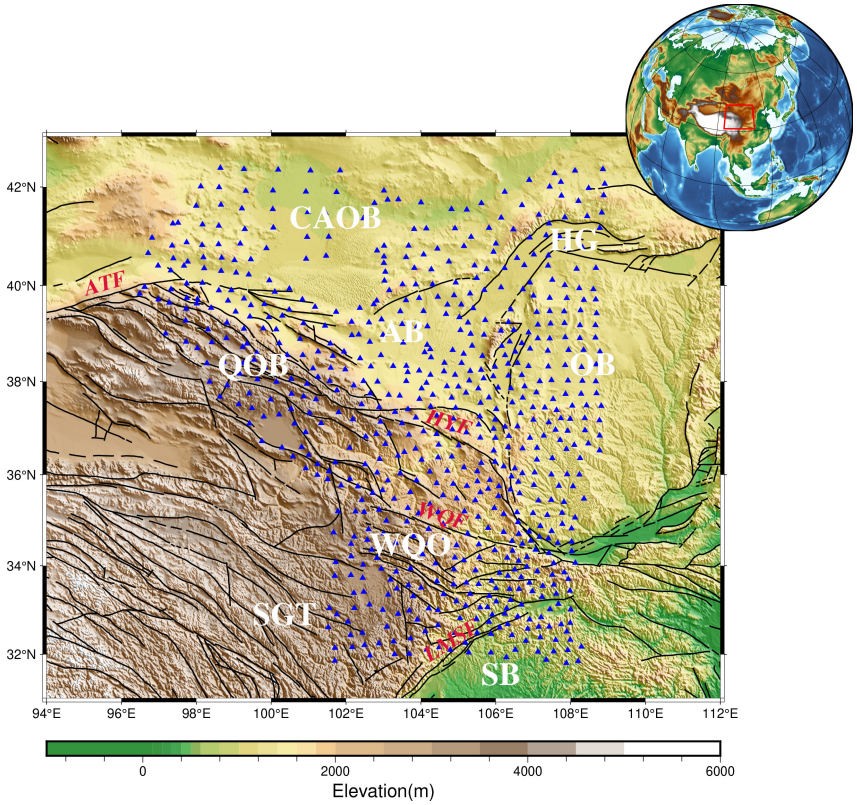


Figure 1. Seismic stations geometry and geological background in northeastern Tibetan Plateau. Black lines indicate the main faults, blue triangles indicate stations. The main faults are labeled as: ATF = Altyn-Tagh Fault; HYF = Haiyuan Fault; KLF = Kunlun Fault; LMSF = Longmenshan Fault; WQF = West-Qinling Fault. The main tectonic blocks are labeled as AB = Alxa Block; CAOB = Central Asian Orogenic Belt; HG = Hetao Graben; OB = Ordos Block; QOB = Qilian Orogenic Belt; SB = Sichuan Basin; SGT = Songpan-Ganzi Terrane; WQO = Western Qinling Orogen.

where $\delta\tau_p$ is the bias-corrected phase delay time, $\widetilde{t}_p$ is obtained by fitting the five-parameter Gaussian wavelet to the filtered, windowed cross-correlation function $\widetilde{C}(t) = S_2 * W_s S_2$ between record S_2 and windowed record $W_s S_2$. We calculated the phase delay time between a given station and several nearby stations, with the inter-station distance usually within a radius of 200 km. These inter-station phase delays represent local, relative traveltime information and do not directly form a continuous traveltime surface. Therefore, following standard practice in teleseismic surface wave studies, the measured phase delays were first used to construct smooth traveltime surfaces through eikonal tomography using the ASWMS framework. The reconstructed traveltime surfaces are subsequently used to provide spatially continuous wavefield representations. Traveltime values are sampled from these surfaces at station locations, while amplitude values are obtained directly from the independently measured amplitude field. These combined phase and amplitude constraints serve as inputs to the Helmholtz tomography.

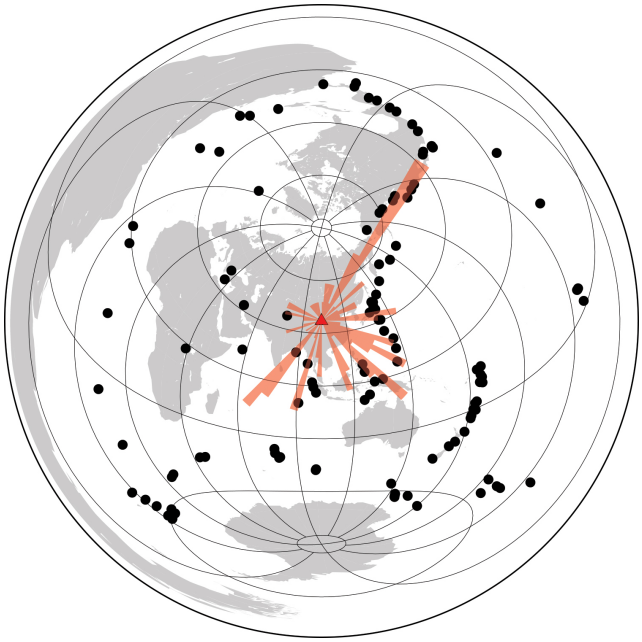


Figure 2. Distribution of the earthquakes after data processing. Black dots indicate teleseismic events, red triangle indicates the center of the seismic array, and the rose map indicates the backazimuth distribution of those events.

3 Methodology

3.1 Physics-informed neural network

PINNs leverage the powerful capabilities of DNNs as universal function approximators to solve problems governed by physical laws (Hornik et al., 1989). The fundamental principle of PINNs is to embed PDEs as soft constraints in the training process, allowing the network to learn solutions that naturally satisfy the underlying physical constraints. This is achieved by adding a PDE-based penalty term to the network's loss function (Raissi et al., 2019). Through automatic differentiation, the network learns to approximate functions that map input coordinates to solutions while respecting the governing physical equations.

PINNs provide a flexible framework for solving diverse types of PDEs, generally expressed as (Raissi et al., 2017):

$$u_t + \mathcal{F}[u] = 0, \quad x \in \Omega, \quad t \in [0, T], \tag{3}$$

where $u = u(\mathbf{x}, t)$ represents the solution to be determined, $\mathbf{x}$ and t denote spatial and temporal coordinates, respectively, u_t represents the partial derivative of u with respect to t, and Ω is a subset of $\mathbb{R}^d$. The differential operator $\mathcal{F}[\cdot]$ encapsulates the mathematical structure of the PDE, which can accommodate both linear and nonlinear systems, as well as time-dependent and time-independent problems. Given initial and boundary conditions, the physical field $u(\mathbf{x}, t)$ can be solved over the spatiotemporal domain.

The PINN loss function typically consists of two terms: one term ensuring adherence to the governing equations, and another term enforcing consistency with observed measurements. The total loss function in PINNs is expressed as:

$$\mathcal{L} = \sum_{i}^{N_d} ||N_u(\theta; \mathbf{x}_i, t_i) - u(\mathbf{x}_i, t_i)||^2 + \sum_{j}^{N_p} ||L(N_u(\theta; \mathbf{x}_j, t_j))||^2, \tag{4}$$

where N_d is the total number of data points, $N_u(\theta; \mathbf{x}_i, t_i)$ refers to the NN's prediction at points $(\mathbf{x}_i, t_i)$, θ denotes the network's parameters that are specific to each NN, $u(\mathbf{x}_i, t_i)$ represents the observed data value, N_p is the number of collocation points for enforcing physical laws, L represents the residual constructed based on Equation 3. The first term (denoted the data term) on the right side of Equation 4 measures the discrepancy between the network's output and the known values and the second term (denoted the physical term) measures how well the network satisfies the PDE residual at collocation points. In practice, a weighting factor can be introduced to balance these two terms based on the specific problem. This combined loss guides the NN to find solutions that satisfy both the data and the governing equations, forming a robust framework for solving complex physics-based problems. In the context of seismic tomography, this framework can be adapted to solve the inverse problem of determining subsurface velocity structure from observed seismic waveforms.

Each of the physical fields in our framework is represented by a standard fully connected feed-forward multilayer perceptron (MLP), in which every hidden layer applies an affine transformation followed by a smooth nonlinear activation. For a network N_q with L_q layers, the forward pass is

$$N_q(\mathbf{z}_q; \theta_q) = \mathbf{W}_q^{(L_q)} \mathbf{h}_q^{(L_q-1)} + \mathbf{b}_q^{(L_q)}, \qquad \mathbf{h}_q^{(l)} = \sigma\Big(\mathbf{W}_q^{(l)} \mathbf{h}_q^{(l-1)} + \mathbf{b}_q^{(l)}\Big), \tag{5}$$

where $q \in \{\tau, A, c\}$ indexes the traveltime, amplitude, and velocity networks, $\mathbf{z}_q$ is the network input, $\mathbf{h}_q^{(l)}$ is the output of hidden layer l, and $\sigma(s) = \arctan(s)$ is the smooth activation function. The parameter set $\theta_q = \{\mathbf{W}_q^{(l)}, \mathbf{b}_q^{(l)}\}_{l=1}^{L_q}$ comprises all trainable weights and biases of network q and has no direct physical interpretation, as it is determined solely through minimization of the loss function. The smooth activation makes the first and second spatial derivatives required by the Helmholtz constraint available through automatic differentiation.

3.2 Teleseismic Helmholtz tomography using PINNs

In seismic tomography problems we are interested in solving for the medium parameters (e.g. velocity) governing wave propagation. To achieve this within a deep learning framework, Figure 3 presents a schematic diagram of PINN architecture for teleseismic Helmholtz tomography. The pinnTHT is composed of three interconnected networks (N_τ, N_c and N_A), which approximate the traveltime, phase velocity, and amplitude, respectively. These networks work in conjunction with physical constraints governed by the Helmholtz equation (Equation 6). Specifically, the spatial variations of predicted phase velocity are related to the observed phase traveltimes and an amplitude-dependent correction through this PDE. The automatic differentiation capability enables efficient computation of derivatives needed for the physical constraints, making the system both physically plausible and computationally tractable.

Before detailing the mathematical formulation of our PINN approach, it is essential to understand the physical foundation of Helmholtz tomography and how it improves upon traditional methods. Traditionally, seismic tomography involves solving a global inverse problem where observed traveltimes are related to velocity structure through ray path integrals. Consequently, each observation constrains the average phase velocity along the entire ray path rather than at specific points, limiting the resolution for features smaller than the seismic wavelength. To overcome

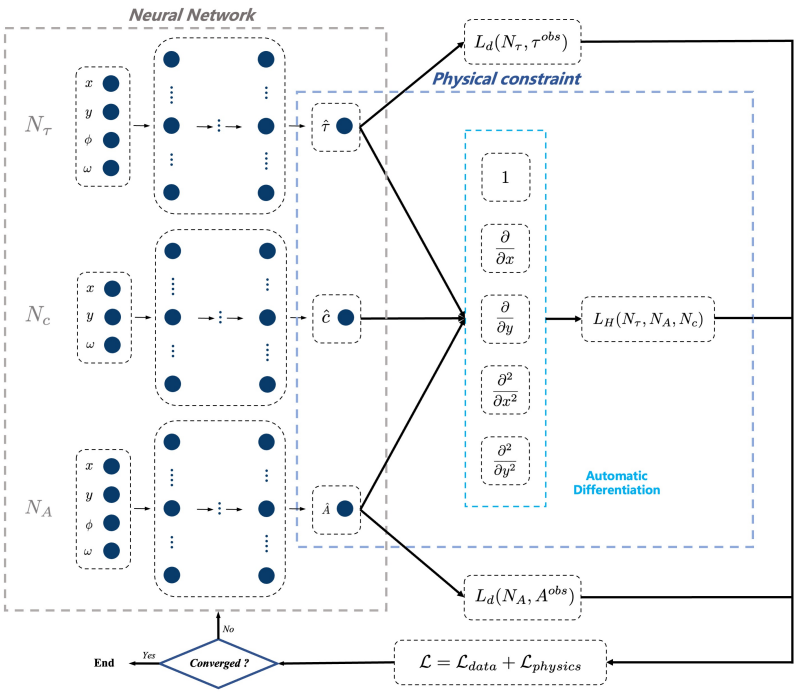


Figure 3. Schematic of PINN for teleseismic Helmholtz tomography. N_τ, N_c and N_A are the traveltime, velocity and amplitude NNs, respectively, x, y represent the spatial coordinates of seismic receivers in the study area, ϕ denotes the measured backazimuth, ω represents the angular frequency of the Rayleigh waves being analyzed, $\hat{\tau}$, $\hat{c}$ and $\hat{A}$ are the predicted traveltime, velocity and amplitude field from the trained N_τ, N_c and N_A. The total loss function ($\mathcal{L}$) consists of two main terms: $\mathcal{L} = \mathcal{L}_{data} + \mathcal{L}_{physics}$, where $\mathcal{L}_{data}$ ($L_d(N_\tau, \tau^{\text{obs}})$, $L_d(N_A, A^{\text{obs}})$) represents the data misfit between predicted and observed Rayleigh wave traveltimes and amplitudes, $\mathcal{L}_{physics}$ ($L_H(N_\tau, N_A, N_c)$) enforces the Helmholtz equation constraint. The automatic differentiation of NNs calculates spatial derivatives $\frac{\partial}{\partial x}$, $\frac{\partial}{\partial y}$ $\frac{\partial^2}{\partial x^2}$ and $\frac{\partial^2}{\partial y^2}$, which relates the traveltime gradient to the local velocity structure with a correction term that accounts for amplitude variations. The neural network part on the left is used to approximate the solution, while the data and physical constraints part on the right is used to optimize the network parameters θ. The training process should terminate when the model is considered converged.

these limitations, gradient-based methods like eikonal tomography were developed to directly determine local phase velocities from the spatial gradients of traveltime surfaces. However, the eikonal equation neglects amplitude information, which may introduce errors in the phase velocity measurement when multi-path propagation of waves occurs (L. Li et al., 2023). The Helmholtz equation, a second-order nonlinear PDE, further corrects the influence of amplitude on the basis of the Eikonal equation (Wielandt, 1993; Lin & Ritzwoller, 2011):

$$\frac{1}{c^2(\mathbf{x})} = |\nabla\tau(\mathbf{x})|^2 - \frac{\nabla^2 A(\mathbf{x})}{A(\mathbf{x})\omega^2}, \tag{6}$$

where ∇^2 denotes the Laplacian operator, $\tau(\mathbf{x})$ represents the phase traveltime from an effective source to a receiver at location $\mathbf{x} = (x, y)$, $c(\mathbf{x})$ denotes the local phase velocity at $\mathbf{x}$, $A(\mathbf{x})$ refers to the amplitude at the same receiver location, ω

is the angular frequency. Note that τ, c, and A are all frequency-dependent quantities associated with the analyzed Rayleigh-wave frequency. Both τ and c vary with frequency because of surface-wave dispersion, while A is measured independently at each analyzed frequency band. By incorporating the amplitude correction, the Helmholtz equation improves the accuracy of phase velocity estimates especially at longer periods.

Based on Equation 6, the physical term in the loss function is defined as:

$$L_H(N_\tau, N_A, N_c) = \sum_{k}^{N_{freq}} \sum_{j}^{N_{src}} \sum_{i}^{N_{rcv}} \left(|\nabla N_\tau|^2 - \frac{1}{N_c^2} - \frac{\nabla^2 N_A}{N_A\, \omega_k^2} \right)^2, \tag{7}$$

where L_H is the loss function of the Helmholtz constraint, $N_\tau(\mathbf{x}, \phi, \omega; \theta_\tau)$, $N_A(\mathbf{x}, \phi, \omega; \theta_A)$, and $N_c(\mathbf{x}, \omega; \theta_c)$ denote the NN approximations of phase traveltime, amplitude, and phase velocity, respectively, with trainable parameter sets θ_τ, θ_A, and θ_c. Here $\mathbf{x}$ is the receiver location, ϕ is the measured event backazimuth, and ω is the angular frequency.

Critically, this study employs array-derived measured backazimuths (ϕ in Figure 3) rather than theoretical great-circle predictions. This strategic choice mitigates epistemic uncertainties arising from teleseismic source mislocations and lateral refractive path deviations. Each network is a continuous function of its inputs rather than a set of values on a fixed grid. N_τ and N_A map a receiver coordinate, an event backazimuth, and a frequency to a scalar traveltime and amplitude, respectively, whereas N_c maps a coordinate and frequency to a phase velocity. Each of N_τ and N_A is a single network that represents the corresponding field continuously over the entire array and can be evaluated at the coordinates of all receivers. The backazimuth ϕ acts as a per-event conditioning input that specifies the incoming wavefront orientation. Evaluating N_τ at a fixed ϕ over all receivers reconstructs that event's continuous traveltime field. Because only relative traveltimes within the local array are modelled and teleseismic wavefronts cross the array as quasi-plane waves, the absolute source distance is not required as an input. Two events with the same backazimuth and frequency are represented by the same relative traveltime surface in N_τ, because the absolute arrival time is removed by per-event referencing. The backazimuth ϕ therefore provides the propagation direction needed to represent events arriving from different directions.

The two data terms measure the misfit between the network predictions and the observed traveltimes and amplitudes (Figure 3):

$$L_d(N_\tau, \tau^{\text{obs}}) = \sum_{k}^{N_{freq}} \sum_{j}^{N_{src}} \sum_{i}^{N_{rcv}} \left(N_\tau(\mathbf{x}_i, \phi_j, \omega_k) - \tau^{\text{obs}}_{i,j,k} \right)^2, \tag{8}$$

$$L_d(N_A, A^{\text{obs}}) = \sum_{k}^{N_{freq}} \sum_{j}^{N_{src}} \sum_{i}^{N_{rcv}} \left(N_A(\mathbf{x}_i, \phi_j, \omega_k) - A^{\text{obs}}_{i,j,k} \right)^2, \tag{9}$$

where N_{freq}, N_{src}, and N_{rcv} denote the numbers of frequencies, sources, and receivers, respectively. $\tau^{\text{obs}}_{i,j,k}$ and $A^{\text{obs}}_{i,j,k}$ are the observed phase traveltime and amplitude at location $\mathbf{x}_i = (x_i, y_i)$ and frequency ω_k for direction ϕ_j.

The overall loss function used to train these three networks simultaneously has the form:

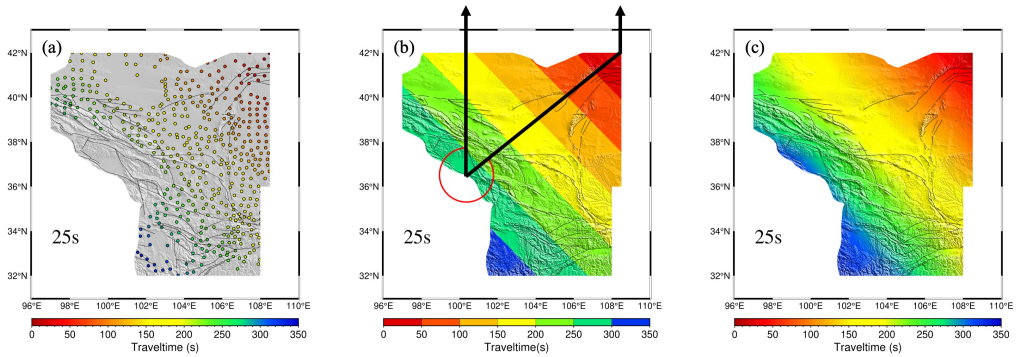


Figure 4. (a) Observed Rayleigh wave phase delay time for seismic event 201507270449 measured using the ASWMS. (b) Schematic illustration of the measured backazimuth calculation. The colored surface represents a 2-D linear interpolation of the observed traveltimes shown in (a), the arrow indicates geographic north, the red curve represents the measured backazimuth that is the clockwise angle between the normal vector to the observed traveltime surface and geographic north. (c) Predicted traveltime surface for seismic event 201507270449, evaluated from the trained N_τ network using the pinnTHT method.

$$\mathcal{L}(\theta_\tau, \theta_A, \theta_c) = L_d(N_\tau, \tau^{\text{obs}}) + \epsilon_A \, L_d(N_A, A^{\text{obs}}) + \epsilon_H \, L_H(N_\tau, N_A, N_c), \tag{10}$$

where ϵ_A and ϵ_H are the weighting factors for the amplitude data term and physical constraint term. The loss function design adopts a multi-term approach, including data constraint terms (ensuring prediction fits with observational data) and physical constraint terms (guaranteeing solutions satisfy the Helmholtz equation). The weight factors were determined based on synthetic tests to balance the contribution of each term, avoiding overfitting for any term. These loss weights play a role analogous to the smoothing parameter in the smoothing-spline formulation of Sergeant et al. (2025), as they control the balance between data misfit and enforcement of the physical constraint.

The training process aims to find the optimal parameters θ_τ, θ_A, and θ_c by minimizing the loss function in Equation 10:

$$\arg \min_{\theta_\tau, \theta_A, \theta_c} \{\mathcal{L}(\theta_\tau, \theta_A, \theta_c)\} \rightarrow \hat{\theta}_\tau, \hat{\theta}_A, \hat{\theta}_c. \tag{11}$$

After training, the phase velocities across all frequencies can be directly evaluated from the trained velocity NN N_c. We can also evaluate traveltime surfaces and amplitude maps from the trained traveltime NN N_τ and amplitude NN N_A.

4 Results

4.1 Training inputs and targets

We applied the pinnTHT method to teleseismic recordings collected from the dense seismic array deployed in the northeastern Tibetan Plateau to obtain Rayleigh wave phase velocities for periods of 20-80 s. As shown in Figure 3, the inputs to the NNs include the spatial coordinates (x, y) of stations, measured backazimuths ϕ, and frequencies ω. The training targets are observed traveltime (Figure 4 (a)) and

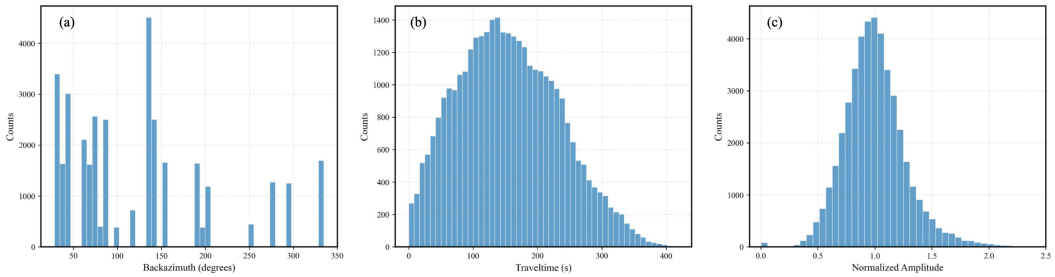


Figure 5. (a) Distribution of the measured backazimuth. (b) Distribution of the observed traveltime. (c) Distribution of the normalized amplitude.

amplitudes (Figure S2), and the residual of the Helmholtz equation. Figure 4 (a) shows the Rayleigh wave traveltime surface obtained from inter-station phase delays through the eikonal reconstruction step of ASWMS, displayed as traveltime values sampled at station locations. Although plotted as discrete points, the spatial distribution reveals a smooth and continuous traveltime pattern across the array, reflecting the underlying wavefront geometry. These station-based traveltime values are subsequently used as observational constraints in the pinnTHT. The NNs are trained to fit the observed data while simultaneously driving the residual of the physical equation towards zero. Figure 4 (c) shows the predicted traveltime field from the trained PINN, which closely reproduces the observed patterns in Figure 4 (a), demonstrating effective data fitting across the array. Similarly, the amplitude predictions (compared against observations in Figure S2) further validate the model's ability to capture both phase and amplitude variations. The spatial coordinates of stations were converted from latitude and longitude coordinates to the Cartesian coordinate system using the Universal Transverse Mercator (UTM) projection. As shown in Figure 4 (b), the backazimuths were determined from the phase delay times by fitting the 2-D linear traveltime surfaces to these delays and calculating the normal vectors to the fitted planes. The backazimuth is defined as the clockwise angle between geographic north and the traveltime surface normal direction, where the normal direction is measured from the station location to the point of minimum traveltime (source location), giving the effective source direction relative to the receivers without requiring precise source locations. This use of backazimuth, rather than direct source locations, offers several advantages. Since teleseismic earthquakes occur at great distances from the stations, projecting source and receiver locations from latitude/longitude to Cartesian coordinates can introduce errors due to Earth's curvature. The backazimuth, being derived from the direction of wave arrival at the local array, is less sensitive to these distance-related inaccuracies. Furthermore, the method of fitting 2-D linear traveltime surfaces and calculating their normal vectors provides a robust way to capture the wavefront's direction, thereby offering a local constraint for phase velocity determination. More sophisticated approaches, such as array beamforming, may further improve backazimuth estimates but require additional processing and computational effort. Evaluating the influence of beamforming-based backazimuth estimates on the inversion results is beyond the scope of the present study and remains a promising direction for future work.

Figure 5 displays the distributions of the measured backazimuth, observed traveltime, and normalized amplitude data, aggregated over the entire training dataset. In Figure 5 (a), the backazimuth spans nearly the full range of possible angles (0-360 degrees), though the distribution is uneven. It features several distinct peaks and gaps, with the most prominent peak occurring around 140-150 degrees, reaching nearly 4500 counts. The observed traveltime distribution corresponds to relative

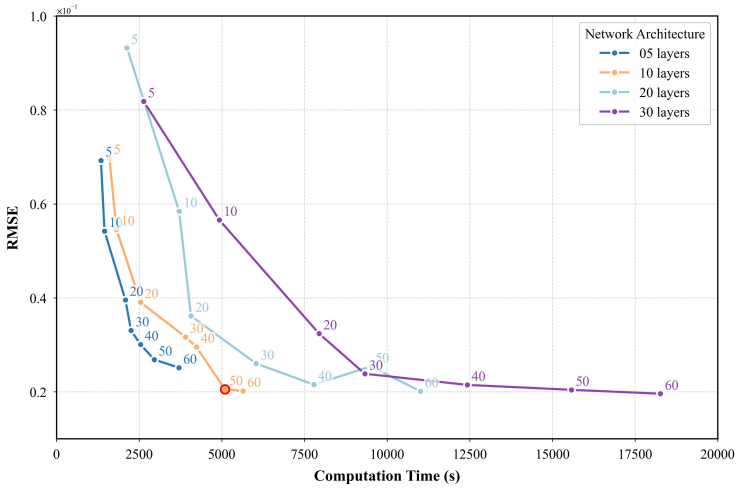


Figure 6. Trade-off between traveltime prediction errors and computational cost. The different colors and curves indicate the different number of layers, and the different points and labels indicate the different number of neurons in each layer. The red circle highlights an optimal configuration for traveltime NNs with 10 layers and 50 neurons per layer.

inter-station phase delays, obtained by sampling the traveltime surface reconstructed from eikonal tomography at the station locations. As shown in the Figure 5 (b), the traveltime distribution shows a roughly bell-shaped pattern with a clear concentration between approximately 50 to 250 s, and a gradual tail extending toward 400 s. Figure 5 (c) illustrates the distribution of normalized amplitude (original amplitudes divided by their mean). Amplitudes are shown normalized because only relative variations are physically meaningful, while the raw values exhibit a highly right-skewed distribution spanning orders of magnitude. The normalized amplitudes follow a unimodal, slightly right-skewed distribution centered near 1.0, with the majority of values concentrated between 0.5 and 1.5. All training inputs underwent normalization processing to ensure their values are on a comparable scale. Specifically, the spatial coordinates (x, y) and the angular frequency ω are linearly rescaled to a common reference interval based on their minimum and maximum values, the backazimuth ϕ is linearly rescaled by a constant factor to a comparable numerical range, the amplitudes are normalized by their mean $\bar{A}$, and the traveltimes are referenced relative to each event before being scaled:

$$\tilde{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \quad \tilde{y} = \frac{y - y_{\min}}{y_{\max} - y_{\min}}, \quad \tilde{\omega} = \frac{\omega - \omega_{\min}}{\omega_{\max} - \omega_{\min}}, \quad \tilde{A} = \frac{A}{\bar{A}}, \tag{12}$$

where a tilde denotes a normalized quantity. To ensure robust optimization, all individual terms thus undergo magnitude alignment via preprocessing normalization, preventing features with large numerical ranges (like traveltime in seconds) from dominating the NN's learning process over features with smaller ranges (like normalized amplitude). Without normalization, such scale imbalances can lead to biased gradients, slower convergence, or unstable training results.

4.2 Neural network design

pinnTHT employs three separate NNs (N_τ, N_c and N_A) to learn the traveltime, phase velocity, and amplitude fields, respectively. Each network is designed as a deep fully-connected NN with specific architecture to handle the complexity of seismic wave propagation. The input layers for each network are designed to fulfill

specific functions. N_τ and N_A take 4-D inputs (x, y, ϕ, ω), while N_c takes 3-D inputs (x, y, ω), as phase velocity is independent of propagation direction. The design of the PINN architecture follows a "step-by-step design, joint optimization" strategy. First, we designed the structures of the traveltime network (N_τ) and amplitude network (N_A) since these networks have direct observational data as evaluation criteria. To determine the optimal network architectures, we conducted several experiments evaluating the trade-off between prediction accuracy and computational efficiency. The analysis focused on both model performance and practical implementation considerations. We used the root mean square error (RMSE) to represent the discrepancy between the observations and the NN's predictions. For example, the RMSE between the predicted and observed traveltimes is defined as:

$$RMSE_\tau = \sqrt{\frac{1}{N_{rcv}} \sum_{i=1}^{N_{rcv}} (\tau_i^{pred} - \tau_i^{obs})^2}, \tag{13}$$

The computational time is the training time for the entire dataset. We tested various network configurations with combinations of different numbers of layers and neurons. Specifically, we had configurations with 5, 10, 20 and 30 layers, respectively. For each layer configuration, we also tested networks with 5, 10, 20, 30, 40, 50 and 60 neurons. As shown in Figure 6, there is a clear trade-off between the number of layers and neurons in the network and the computation time required. Generally, as the number of layers and neurons increases, traveltime $RMSE_\tau$ decreases, indicating better performance. However, this improvement in performance comes at the cost of increased computation time. For example, the $RMSE_\tau$ of 5-layer networks decreases rapidly as the number of neurons increases from 5 to 60, while 30-layer networks achieve lower $RMSE_\tau$ values over the same range but at cost of much longer computation times. Based on these tests, we determined that the best trade-off for N_τ contains 10 hidden layers with 50 neurons in each layer (red circle in Figure 6). This configuration achieves a good balance between prediction accuracy ($RMSE_\tau \approx 0.02$) and computational efficiency (computation time $\approx 5000\ s$), beyond which adding more neurons or layers yields diminishing returns in performance improvement. Here the computational cost is defined as the wall-clock training time required to reach the fixed number of epochs, measured on a single workstation with an Intel Core i7-8700 CPU at 3.20 GHz using the SciANN framework (Haghighat & Juanes, 2021), a Keras/TensorFlow wrapper for physics-informed deep learning. The reported $RMSE_\tau \approx 0.02$ is given in the scaled traveltime units used for training (traveltimes are divided by 100), corresponding to $\approx$ 2 s in the original units. It is the RMS residual of the traveltime network over all selected events and stations and is of the same order as, though not identical to, the station-level residual scatter of the measured traveltimes about a smooth wavefront (∼2 s) and the typical teleseismic phase-traveltime measurement uncertainty, indicating that the network fits a smooth field down to the data-scatter level rather than overfitting it. The number of layers and neurons also influences the effective smoothness of the recovered fields. Larger networks can represent finer-scale variations and thus impose weaker implicit regularization, whereas smaller networks act as a stronger smoothing prior. A similar trend holds for the amplitude $RMSE_A$, which decreases consistently with deeper and wider networks. Accordingly, we selected 10 layers with 50 neurons per layer as the optimal architecture for N_A. After determining the network structures of N_τ and N_A, we proceed to design the structure of the velocity network (N_c). Since N_c has no direct observational data, its evaluation mainly relies on the Helmholtz equation residual and the prediction accuracy of the entire system. By fixing the optimized structures of N_τ and N_A, testing different N_c configurations via a grid-search across combination depths of 5 to 30 layers and widths of 5 to 60 neurons, observing the convergence of the Helmholtz equation residual and the overall sys-

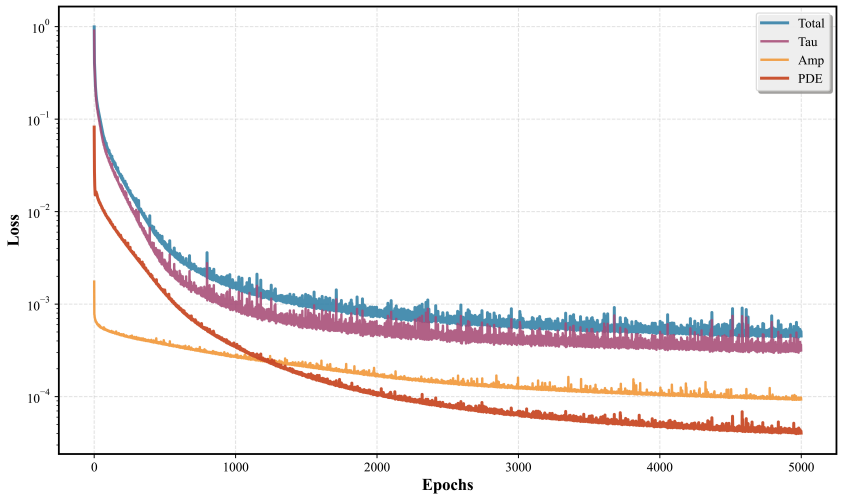


Figure 7. Convergence of the total loss function (Total), the data constraints including traveltime loss (Tau) and amplitude loss (Amp), and physical constraint (PDE) in Equation 10.

tem performance, we finally determined a 5-layer network with 20 neurons per layer as optimal NN design since it yields the lowest stable physical residual without triggering overfitting. This step-by-step design approach not only simplifies the optimization process but also ensures the physical consistency and numerical stability of the entire system.

4.3 Hyperparameter settings

All other network parameters were determined through systematic synthetic tests. We selected the Adam optimizer due to its ability to effectively handle non-convex optimization problems and excellent performance in processing large-scale data (Kingma & Ba, 2014). The loss function is minimized using Adam with mini-batch training, where each mini-batch approximately corresponds to the station measurements from a single event, followed by refinement using full-batch optimization. Network parameters are initialized using Glorot initialization with zero biases, and the default Adam constants $(\beta_1, \beta_2, \delta) = (0.9, 0.999, 10^{-7})$ are used. At iteration r, the parameters are updated as

$$\begin{aligned} \mathbf{m}_r &= \beta_1 \mathbf{m}_{r-1} + (1-\beta_1)\nabla_\theta \mathcal{L}_r, \qquad \mathbf{v}_r = \beta_2 \mathbf{v}_{r-1} + (1-\beta_2)(\nabla_\theta \mathcal{L}_r)^2, \\ \theta_{r+1} &= \theta_r - \eta \, \frac{\mathbf{m}_r/(1-\beta_1^r)}{\sqrt{\mathbf{v}_r/(1-\beta_2^r)} + \delta}, \end{aligned} \tag{14}$$

where $\mathbf{m}_r$ and $\mathbf{v}_r$ are the first- and second-moment estimates of the loss gradient $\nabla_\theta \mathcal{L}_r$ and η is the learning rate. No additional hidden regularization term is used. The learning rate was set to 0.0001, which demonstrated a good balance in synthetic tests. This value is neither too large to cause training instability nor too small to result in slow convergence. The total number of training epochs was set to 5000, determined based on the convergence characteristics of the loss functions. The relative weighting between traveltime and amplitude misfit terms is controlled by the parameter ϵ_A in Equation 10. Based on a series of sensitivity tests conducted over a theoretical noisy benchmark model where ϵ_A was stepped from 0.01 to 1.0, ϵ_A was set to 0.1, which allows amplitude information to act as a secondary, physically motivated constraint without dominating the inversion. This choice ensures that the amplitude term contributes to stabilizing the Helmholtz formulation by accounting for wave-

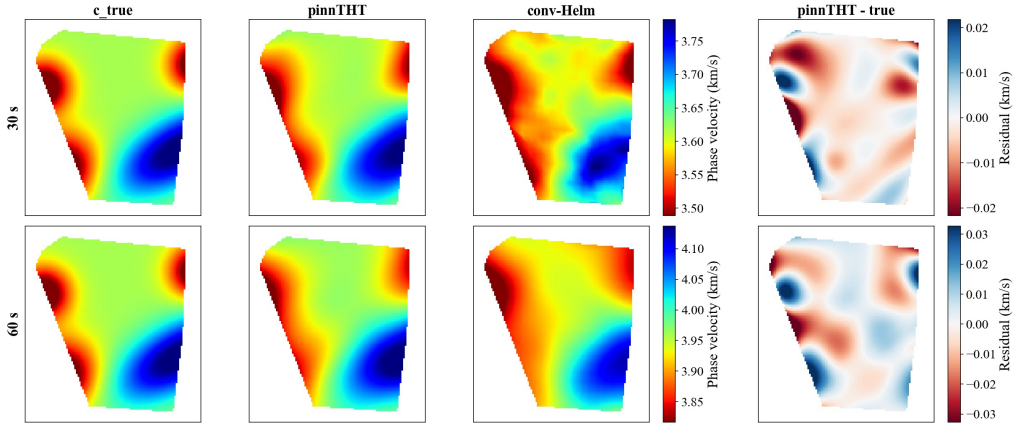


Figure 8. Synthetic ground-truth recovery test at periods of 30 and 60 s. The columns show, from left to right, the true velocity model used to generate the synthetic data (c_{true}), the model recovered by pinnTHT from ∼20% of the events (pinnTHT), the model recovered by conventional Helmholtz tomography from the full dataset (conv-Helm), and the difference between the pinnTHT recovery and the true model (pinnTHT − true). Results for the remaining periods (20, 25, 40, and 80 s) are shown in Figure S7.

front curvature and finite-frequency effects, while preserving the primary control of traveltime observations on phase velocity estimation. The weights for physical constraints (ϵ_H in Equation 10) were set to increase linearly with period. This empirical weighting scheme was selected based on the characteristics of the present dataset and the observed convergence behavior during training, rather than as a universal property of surface-wave observations. In our dataset, longer-period measurements provide comparatively weaker constraints on lateral heterogeneity owing to their broader spatial sensitivity. Increasing the weight of the physical constraint therefore helps stabilize the optimization and maintain physically consistent solutions. We used arctangent as the activation function, which determines the nonlinear approximation effect of the NN. These parameter selections not only ensure the model's ability to accurately and efficiently extract maximum details from field data but also guarantee the stability and reliability of the training process. Figure 7 shows the convergence processes of the data constraints, physical constraints, and total loss. The total and traveltime terms converge rapidly to below 10^{-3}, the amplitude misfit stabilizes at the order of 10^{-4}, and the physical term converges to below 10^{-4}, indicating fast learning speed and strong training stability. The adopted loss weights and network sizes are further verified against ground truth in Section 4.4.

4.4 Synthetic and field-data validation

To provide ground-truth validation of the method and to quantify its advantages relative to conventional Helmholtz tomography, we designed two complementary experiments using the real 676-station geometry and the measured event backazimuths.

In the synthetic recovery test, we constructed a geologically realistic heterogeneous velocity model designed to mimic the complex tectonic features and multi-scale velocity anomalies characteristic of the northeastern Tibetan Plateau. Synthetic traveltimes and amplitudes were generated within this known model by solving the 2-D frequency-domain Helmholtz equation with a high-fidelity finite-difference method using the true station-event geometry. To eliminate shared numerical dis-

Table 1. Ground-truth recovery accuracy inside the array coverage. For each period, the root-mean-square error (RMSE, in km/s) of the recovered phase velocity is listed with the corresponding correlation coefficient R in parentheses, together with the amplitude recovery ratio (unity indicates perfect recovery). pinnTHT uses only ∼20% of the events, whereas conventional Helmholtz tomography uses the full dataset. The mean row refers to the RMSE only.

	RMSE [km/s] (R)		Amplitude recovery ratio	
Period (s)	pinnTHT	conv-Helmholtz	pinnTHT	conv-Helmholtz
20	0.0126 (0.99)	0.0390 (0.93)	0.95	1.19
25	0.0097 (0.99)	0.0351 (0.95)	0.99	1.03
30	0.0080 (0.99)	0.0268 (0.98)	0.99	1.07
40	0.0075 (0.99)	0.0235 (0.98)	1.00	1.04
60	0.0111 (0.99)	0.0274 (0.97)	0.99	0.96
80	0.0146 (0.98)	0.0294 (0.95)	0.99	0.93
mean RMSE	0.0106	0.0302		

cretization between data generation and inversion, the forward modeling and the inversion use different numerical discretizations. Realistic noise was added and calibrated to match the observed field-data scatter ($\sigma_\tau \approx 2$ s, $\sigma_A \approx 13$–42%). To probe performance under constrained coverage, pinnTHT was trained with only a ∼20% subset of the events and compared against conventional Helmholtz tomography using the full dataset. Despite the severely reduced coverage, pinnTHT recovers the true structure with a mean RMS error of ∼0.011 km/s and high structural fidelity ($R = 0.98$–0.99), achieving a lower mean RMS error than conventional Helmholtz tomography even when the latter uses the full dataset (∼0.030 km/s). The period-by-period RMSE, correlation, and amplitude-recovery ratios for both methods are summarized in Table 1. The recovery at 30 and 60 s is shown in Figure 8, and the results for the remaining periods (20, 25, 40, and 80 s) are provided in Figure S7. At the adopted weights ($\epsilon_A = 0.1$ and a period-dependent ϵ_H), the known model is recovered accurately with no visually coherent large-scale bias, and short-period features are not artificially mapped into longer periods. This benchmark is intended to assess the numerical stability of the method and the robustness of the reconstruction to sparse, noisy data. Under these controlled conditions, pinnTHT reproduces the known velocity model accurately from far fewer events.

In the field-data robustness test, conventional Helmholtz tomography was applied to the same ∼20% event subset used by pinnTHT. Under identical data coverage, conventional Helmholtz tomography exhibits markedly stronger small-scale artifacts than pinnTHT, particularly near array edges and in regions of uneven coverage (Figure S8), whereas pinnTHT remains stable and in closer agreement with previously published phase-velocity models.

4.5 Phase velocity maps in northeastern Tibet

The Rayleigh wave phase velocity maps across periods of 20-80 s are shown in Figure 9, with corresponding velocity uncertainties presented in Figure S3 of the Supplementary Material. Figure 9 indicates the distribution of low velocity zones (LVZs) and high velocity zones (HVZs) is controlled by the strike of faults and the boundaries of blocks, with significant lateral heterogeneity. It can be seen that the phase velocity patterns in different geological units vary with different periods. In this work we do not aim to show the full tomographic inversion of the derived phase velocity maps but restrict ourselves to discussing the phase-sensitive velocity maps

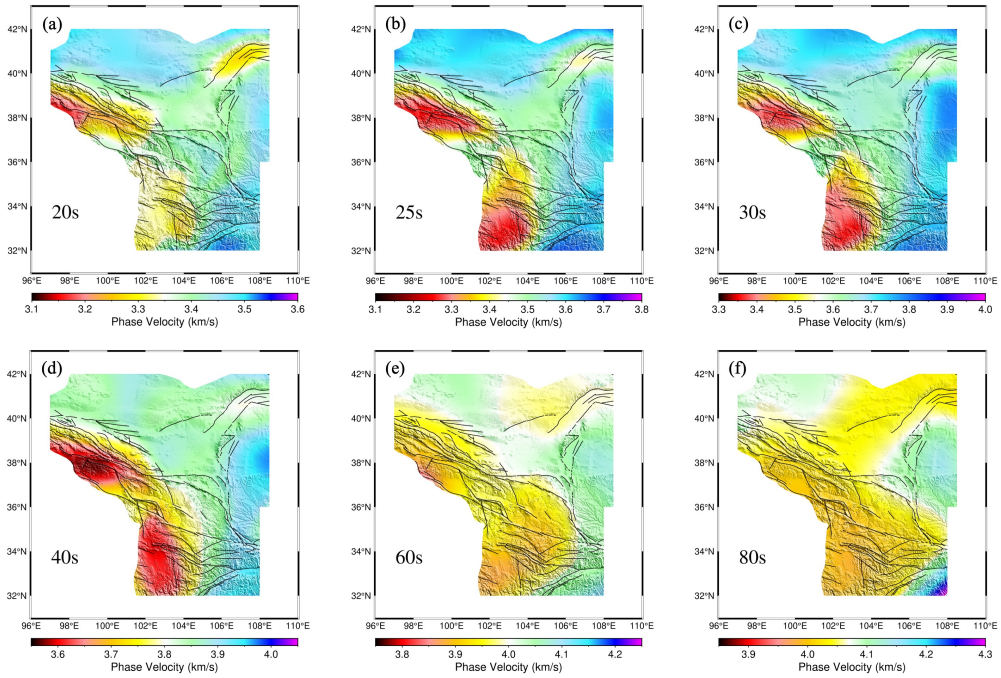


Figure 9. (a-f) Rayleigh wave phase velocities at periods of 20, 25, 30, 40, 60, and 80 *s* beneath northeastern Tibetan Plateau obtained from pinnTHT.

directly. The phase velocitity at each period is most sensitive to S-wave velocity at a depth of approximately one-third of the corresponding wavelength.

The phase velocities of Rayleigh waves with periods of 20-25 *s* are most sensitive to the S-wave velocities at a depth of 20-40 *km*, mainly reflecting the velocity variations in the mid-crust within the study area. While the actual depth sensitivity is described by continuous kernels rather than discrete depth ranges, this simplified interpretation helps identify the main structural features. Along the southwest boundary, prominent LVZs exist in the western Qilian Orogen, western Qinling Orogen, and Songpan-Ganzi Terrane (for regions see in Figure 1). From 20 *s* to 25 *s*, both the range and strength of these LVZs increase. The Hetao Graben also shows a relatively low-velocity anomaly. HVZs appear in the northern and eastern boundaries and southeastern corner of the area, corresponding to the southern central Asian Orogenic Belt, northern Alxa Block, Ordos Block, and central Qinling Orogen (the north of Sichuan Basin).

The phase velocities of Rayleigh waves with periods of 30-40 *s* are most sensitive to the S-wave velocities at a depth of 30-70 *km*, mainly reflecting the velocity variations in the lower crust and the uppermost mantle within the study area. Significant LVZs remain in the western Qilian Orogen, western Qinling Orogen, and Songpan-Ganzi Terrane, while the isolated LVZ in the Hetao Graben relatively weakens. The high-velocity anomalies in the southern Ordos Block and central Qinling Orogen become more pronounced compared to the surrounding areas. The high velocities in the central Asian Orogenic Belt and Alxa Block weaken as the period increases, showing generally uniform medium-high velocity characteristics.

The phase velocities of Rayleigh waves with periods of 60-80 *s* are most sensitive to the S-wave velocities at a depth of 70-120 *km*, mainly reflecting the velocity varia-

tions in the upper mantle within the study area. The lateral heterogeneities of the phase velocity in these blocks are further reduced. Overall, the Ordos Block and central Qinling Orogen show relatively high velocities, while the rest of the areas show relatively low velocities. The HVZs and LVZs are separated by the faults distributed around the blocks. A relative HVZ appears in the northwestern corner of the study area, which may be related to the sparse station distribution in this region. The LVZs along the southwestern boundary of this region almost merge together, with relatively low velocity remaining in the Songpan-Ganzi Terrane. Phase velocities in the central Alxa Block change from a relatively higher velocity compared to its surroundings to a similar velocity value from 60 s to 80 s.

5 Discussion

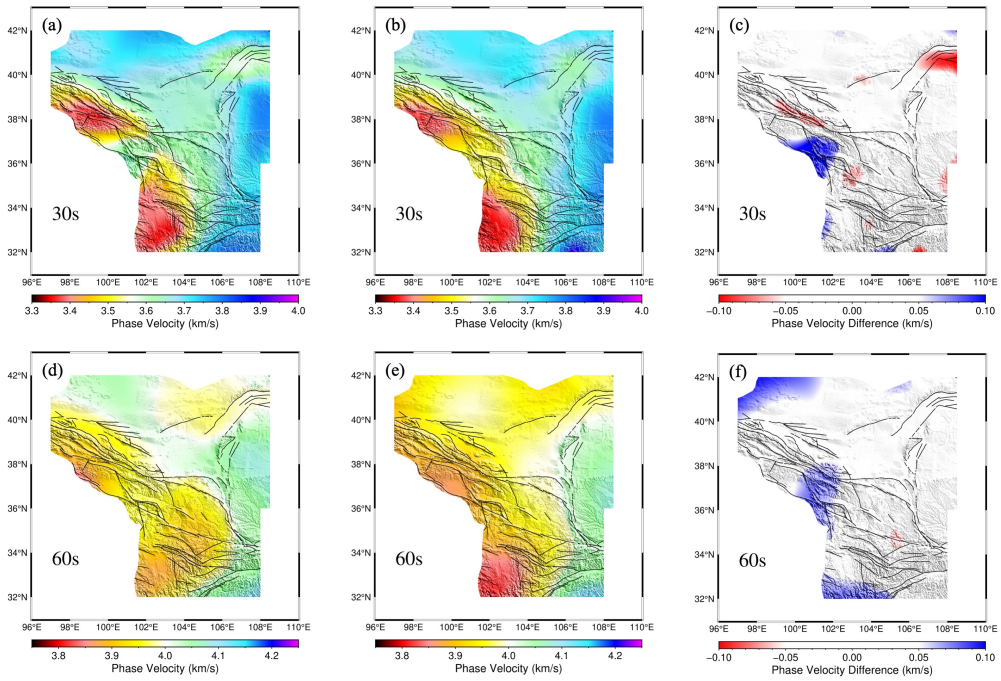


Figure 10. (a)(d) Rayleigh wave phase velocity at periods of 30 and 60 s beneath the northeastern Tibetan Plateau obtained from pinnTHT. (b)(e) Rayleigh wave phase velocity at periods of 30 and 60 s beneath the northeastern Tibetan Plateau obtained from conventional Helmholtz tomography. (c)(f) Difference of Rayleigh wave phase velocity between two methods at periods of 30 and 60 s.

In this paper, we successfully applied the PINNs to teleseismic Helmholtz tomography in the northeastern Tibetan Plateau. By incorporating the Helmholtz equation as a physical constraint within the loss function, the proposed framework jointly constrains traveltime and amplitude information, leading to physically consistent and stable Rayleigh wave phase velocity maps at periods of 20–80 s. In the following, we discuss the performance of the method, its comparison with other approaches, and the geological implications of our findings.

We first validate the reliability of pinnTHT by comparing its results with those from the traditional Helmholtz tomography method using ASWMS based on the same set of teleseismic events and station recordings. Here, the conventional Helmholtz

tomography uses a regular-grid representation and finite-difference operators in a two-stage procedure, in which phase velocities are first estimated from the traveltime and amplitude fields and then averaged across events (Lin & Ritzwoller, 2011; Jin & Gaherty, 2015). However, pinnTHT is trained using only a subset (approximately one-fifth) of the available observations. Unlike conventional tomography methods, which rely on dense spatial sampling to stabilize the inversion, pinnTHT embeds the Helmholtz equation directly into the loss function. Beyond this physical constraint, the shared, event-invariant network parameterization of $N_c(x, y, \omega)$ provides structural regularization across multiple sources, allowing the network to infer spatially continuous phase velocity fields even with sparse observational coverage. Figures 10 and S4 present the phase velocity results obtained from pinnTHT and the traditional method at representative periods, together with their absolute differences. Overall, the two approaches show excellent agreement in large-scale velocity patterns across all periods. At 30 s (Figures 10 (a) and (b)), both methods identify LVZs in the western Qilian Orogen, western Qinling Orogen, Songpan-Ganzi Terrane, and Hetao Graben, as well as HVZs in the Ordos Block and central Qinling Orogen. Velocity differences exceeding 0.05 km/s are mainly concentrated along major fault systems and regions characterized by strong lateral velocity gradients (Figure 10 (c)). The most prominent differences occur in the Hetao Graben and eastern Qilian Orogen. In the Hetao Graben, pinnTHT yields lower phase velocities than the traditional Helmholtz tomography, resulting in a more pronounced low-velocity anomaly relative to the surrounding regions. This feature is consistently observed across periods from 20 to 40 s (Figure S4) and shows better agreement with previously reported surface wave tomography results (Zhong et al., 2017; Wu et al., 2023). Because the velocity field is represented as a continuous function $N_c(\mathbf{x}, \omega)$, neighboring periods are coupled through the shared network parameterization, introducing implicit regularization without explicit cross-period smoothing terms. The synthetic baseline tests (Figure 8) confirm that this formulation does not artificially map short-period features into longer periods. Accordingly, the cross-period consistency of the Hetao Graben anomaly is partly an intrinsic property of the continuous parameterization and should not be viewed as independent confirmation, and our preference for the pinnTHT result there now rests primarily on its agreement with independent published regional models. In contrast, pinnTHT exhibits slightly higher phase velocities in a small region along the southwestern edge of the study area within the eastern Qilian Orogen (approximately 35.5-37° N, 99-102° E) compared to the traditional method. This region corresponds to a complex tectonic transition zone characterized by strong lateral velocity gradients and uneven ray coverage. Under such conditions, grid-based Laplacian calculations used in traditional Helmholtz tomography are prone to amplifying sharp velocity contrasts, potentially leading to overestimated low-velocity anomalies. However, pinnTHT evaluates the governing equation using a continuous neural representation, which naturally suppresses numerical artifacts from discrete differentiation and thus yields slightly higher velocity values in this area. At 60 s (Figures 10 (d) and (e)), the agreement between pinnTHT and traditional Helmholtz tomography further improves, reflecting the increased sensitivity of longer-period surface waves to deeper and more laterally homogeneous structures. Across most of the study area, phase velocity differences are well within 0.05 km/s, indicating robust and stable imaging by both approaches. Slightly higher phase velocities in the pinnTHT results are observed primarily in a small region in the eastern Qilian Orogen, similar to patterns seen at 30 s, as well as in the northwestern and southwestern corners of the study area. These regions coincide with the edges of the seismic array, where ray coverage is reduced and illumination is less uniform. The velocity differences at these locations are therefore most likely attributable to array-edge effects caused by the sparse and uneven ray coverage, rather than reflecting structural information. Considering the entire period range from 20 to 80 s (Figure 10 and S4), the differences between

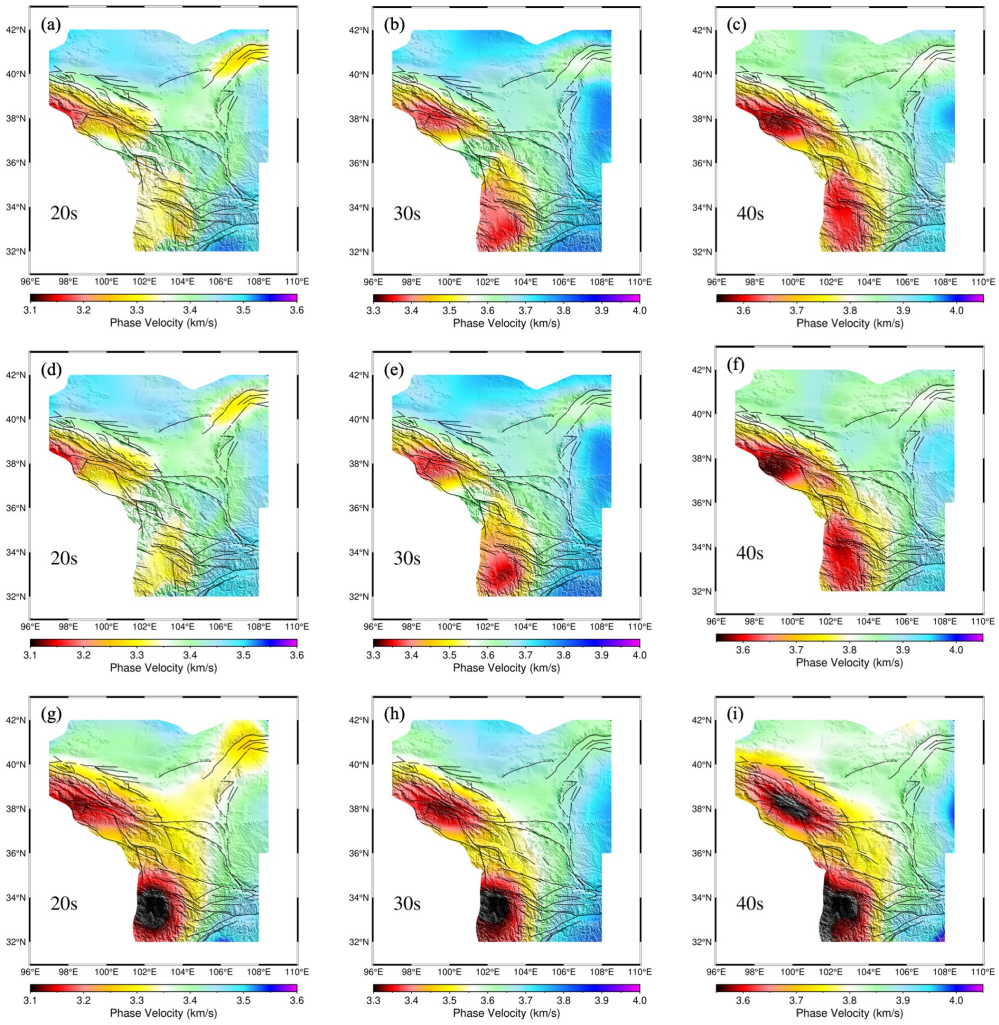


Figure 11. (a)-(c) Rayleigh wave phase velocity at periods of 20, 30 and 40 s beneath northeastern Tibetan Plateau obtained from pinnTHT. (d)-(f) Rayleigh wave phase velocity at periods of 20, 30 and 40 s beneath northeastern Tibetan Plateau obtained from pinnTET. (g)-(i) Rayleigh wave phase velocity at periods of 20, 30 and 40 s beneath northeastern Tibetan Plateau obtained from pinnANT (Chen et al., 2023).

the pinnTHT results and those obtained from conventional Helmholtz tomography are mainly confined to the boundaries of the study area, where data coverage is relatively sparse. Both the magnitude and spatial extent of these differences are overall reduced at longer periods, suggesting that pinnTHT becomes increasingly stable and consistent as the sensitivity shifts toward deeper and more laterally continuous structures. Quantitatively, phase velocity differences in most regions are within 0.05 km/s, indicating that pinnTHT reliably reproduces the primary structural features resolved by conventional Helmholtz tomography.

To further verify the reliability of the method and understand the contributions of different methodological components and data types, we present systematic comparisons at multiple levels. Figure 11 presents a comparison of Rayleigh wave phase

velocity maps obtained from pinnTHT, PINN-based teleseismic eikonal tomography (pinnTET), and PINN-based ambient noise tomography (pinnANT) (Chen et al., 2023) at periods of 20, 30 and 40 s. The pinnTET results for the remaining periods (25, 60 and 80s) and the velocity differences between pinnTHT and pinnTET are provided in Figures S5 and S6 of the Supplementary Material. At the scale of major tectonic units, the three approaches show strong overall consistency, imaging relatively low velocities beneath the western Qilian Orogen, western Qinling Orogen, Songpan-Ganzi Terrane, and relatively high velocities within the Ordos Block and central Qinling Orogen. This agreement indicates that the resolved large-scale structures are robust with respect to both the physical formulation and the data type.

To isolate the specific contribution of the amplitude corrections, we compare pinnTHT (Figures 11 (a)-(c)) with pinnTET (Figures 11 (d)-(f)) obtained using the same data and NN parameters. This controlled comparison eliminates data-related differences and NN architecture variations, illustrating purely the effect of including the amplitude correction term that distinguishes the Helmholtz equation from the eikonal equation. At 20 s, velocity differences are mainly confined to localized regions of the Hetao Graben (approximately 40-41° N, 107-109° E), where pinnTHT exhibits slightly lower phase velocities than pinnTET. This phenomenon suggests that the Helmholtz formulation may be more sensitive to strong lateral heterogeneity and low-velocity anomalies at shorter periods, owing to its explicit treatment of finite-frequency effects and wavefront curvature, which can enhance the expression of low-velocity features. At 30 s, the pattern of differences becomes more complex and spatially extensive. In addition to the differences observed in the Hetao Graben, alternating bands of relatively higher and lower velocities (with respect to pinnTET) appear along the southwestern boundary of the study area, extending across portions of the Qilian Orogen, western Qinling Orogen, and Songpan-Ganzi Terrane. These patterns may reflect a partial correspondence between the wavelength at this period and the dominant scale of subsurface heterogeneity, where complex wavefield interference effects are more effectively resolved by pinnTHT through amplitude corrections but comparatively smoothed by pinnTET. At 40 s, the velocity differences within the Hetao Graben, western Qinling Orogen, and Songpan-Ganzi Terrane largely disappear. In contrast, slightly higher phase velocities are observed in the Ordos Block (approximately 36-39° N, 108-110° E) in the pinnTHT model compared to the pinnTET. This discrepancy may arise from wavefront defocusing effects at longer periods, where surface waves encountering the high-velocity Ordos Block undergo ray path divergence. This reduces amplitudes, resulting in a positive Laplacian term ($\nabla^2 A > 0$) that leads the Helmholtz-based method (pinnTHT) to yield higher velocity estimates than the eikonal approximation (pinnTET). In general, the velocity differences between pinnTHT and pinnTET are less than 0.02 km/s in most of the study areas for all selected periods (20-80 s). The two methods show very similar results at shorter periods (20–25 s), while velocity differences become more spatially widespread and of increased magnitude from 30 s, reflecting the increasing influence of finite-frequency effects and wavefield complexity captured by the Helmholtz formulation but neglected in the eikonal approximation. The overall similarity between pinnTHT and pinnTET results, in contrast to the larger differences reported by Sergeant et al. (2025), can be understood from the structure of the amplitude term. Following Lehujeur and Chevrot (2020), the error incurred by neglecting the amplitude term varies sinusoidally with propagation azimuth and is strongly reduced by azimuthal averaging. Because our maps stack many teleseismic events with broad backazimuthal coverage (Figure S1), the net amplitude correction is small. The scattering term $\nabla^2 A/A$ is itself frequency-independent, but it enters the phase-slowness correction as $\nabla^2 A/(A\omega^2)$, so that its residual effect grows toward longer periods. The mapped magnitude of this correction is further limited by the large station spacing (40–70 km), the smooth teleseismic wavefields, and the noisier

amplitude measurements. We nonetheless retain the full Helmholtz formulation because it removes this bias event-by-event rather than relying on azimuthal averaging.

When comparing PINN-based teleseismic tomography (pinnTHT or pinnTET, Figures 11 (a)-(f)) with those from pinnANT (Figures 11 (g)-(i)), the differences become more pronounced, particularly in absolute phase velocity values. This comparison involves completely independent datasets with fundamentally different wave excitation and propagation, and therefore provides a more rigorous evaluation of methodological consistency and complementary sensitivity. In general, teleseismic surface wave tomography is capable of revealing the deeper structure in the lower crust and upper mantle based on surface waves at longer periods generated by earthquakes, whereas ambient noise tomography has the ability to leverage the surface wave at shorter periods, exhibiting higher resolution in the shallow crustal structures. Here we compared results of Rayleigh wave phase velocities at 20, 30 and 40 s periods, which represents an overlapping interval where these results exhibit reasonable resolution. At 20 s, teleseismic tomography generally produces higher phase velocities than ambient noise tomography, corresponding to the central Alxa Block, Qilian Orogen, western Qinling Orogen, Songpan-Ganzi Terrane, and north and south sides of the Hetao Graben. At 30 s, the magnitude and spatial extent of the velocity differences decrease significantly. The strongest discrepancies are confined to tectonically active regions, particularly along major orogenic belts, while stable regions show improved agreement between the two methods. At 40 s, the velocity differences further decrease and become spatially sparse. Teleseismic tomography reveals slightly lower phase velocities compared to ambient noise tomography in the narrow strip-like area on the southwestern edge of the western Qilian Orogen (approximately 37-39° N, 97-99° E). Overall, the observed period-dependent behavior highlights the complementary sensitivities of the two approaches. The amplitude term in pinnTHT is theoretically expected to enhance sensitivity to smaller-scale structures. However, teleseismic amplitude measurements are more affected by source complexity and scattering, which can reduce their reliability in practice. In contrast, ambient noise tomography benefits from more stable phase measurements, providing robust constraints on shallow to mid-crustal structures at shorter periods, whereas teleseismic tomography offers better sensitivity to deeper structures and upper-mantle velocities at longer periods. In particular, the sharper and stronger anomalies obtained by the ambient noise data reflect its higher signal-to-noise ratio and more stable measurements at crustal periods, whereas the smoother teleseismic images reflect the greater influence of source complexity and scattering on real teleseismic amplitude data, as well as the stronger effect of regularization in stabilizing the inversion. In addition, the underlying datasets are not comparable in size, because the teleseismic inversions are constrained by a limited number of wavefronts crossing the array, whereas the ambient noise result draws on interstation cross-correlations of the continuous recordings, providing far more paths with denser azimuthal sampling. Despite these differences, the spatial patterns of major velocity anomalies remain broadly consistent between the two approaches. The locations and first-order geometries of prominent LVZs and HVZs are well preserved across periods, reinforcing the robustness of the inferred lithospheric-scale structural features. This agreement suggests that the dominant tectonic framework of the region is reliably resolved regardless of the data type.

By combining ambient noise tomography results at short periods with teleseismic surface wave tomography results, we can develop a systematic understanding of the lithospheric structure along the northeastern margin of the Tibetan Plateau. In the sedimentary layer and upper crust (Chen et al., 2023), the western Qilian Orogen, western Qinling Orogen, and Songpan-Ganzi Terrane exhibit prominent low-velocity anomalies. In addition, localized LVZs are present in the Hetao Graben, central Alxa Block, and part of the Ordos Block, particularly in the Hetao Graben, which

are typically attributed to thick unconsolidated basin sediments (Fu & Xiao, 2020). In the middle and lower crust, extensive LVZs persist beneath the western Qilian Orogen, western Qinling Orogen, and Songpan-Ganzi Terrane, while the Hetao Graben maintains relatively low velocities. In contrast, the Ordos Block and central Qinling Orogen are characterized by prominent HVZs. These patterns reflect significant tectonic differences in the middle and lower crust between the northeastern margin of the Tibetan Plateau and its surrounding blocks. The LVZs are typically interpreted as regions of reduced mechanical strength of the medium, which are more susceptible to deformation under tectonic stress. In the upper mantle, the contrast between low and high-velocity anomalies becomes less pronounced, with relative low velocities extending beneath most of the study area except for the Ordos Block, central Qinling Orogen, and parts of the central Asian Orogenic Belt. The Songpan-Ganzi Terrane exhibits particularly distinct relative low velocities, which may be associated with asthenospheric upwelling following the delamination of a thickened lithospheric root (Li et al., 2017). The depth-dependent transition from pronounced crustal heterogeneity to more uniform upper mantle structure reflects the differential response of lithospheric layers to the ongoing Indo-Asian collision and plateau growth.

Finally, we summarize the main assumptions and limitations of the present implementation. First, the velocity network N_c is defined as a scalar phase-velocity field independent of propagation direction. The present study therefore assumes azimuthal isotropy and focuses on incorporating amplitude constraints into the Helmholtz framework, so the resulting maps should be interpreted as azimuthally averaged structures. This is a limitation of the current implementation rather than of PINNs in general, as anisotropic PINN formulations have been demonstrated previously (Chen et al., 2023) and can be incorporated in future work. Second, pinnTHT imposes no explicit damping or smoothing operator. Regularization is instead implicit, arising from the smooth functional form of the network and the physics-informed constraint, which together confine the solution to physically consistent fields. Because of this, a conventional resolution matrix cannot be formed, and structural recovery is assessed empirically through the synthetic ground-truth experiments (Figure 8). Consistent with this, Figure 6 shows that increasing model complexity does not meaningfully reduce reconstruction error beyond a plateau, indicating an appropriate balance between expressivity and data fit. Third, the uncertainty maps in Figure S3 are not a formal posterior model uncertainty. For each period, event-wise phase-velocity estimates derived from the traveltime-gradient formulation are used to compute a mean and standard deviation across events, and the mapped standard deviation reflects inter-event variability and data-sampling effects, so it should be regarded as an empirical dispersion proxy.

6 Conclusion

This study demonstrates the successful application of physics-informed neural networks to teleseismic Rayleigh wave Helmholtz tomography in the northeastern Tibetan Plateau. The proposed pinnTHT method achieves phase velocity models that are highly consistent with conventional Helmholtz tomography, while requiring substantially fewer observations to achieve comparable resolution and stability. By explicitly embedding the Helmholtz equation into the loss function, the method enforces physical constraints in a natural way through a continuous, grid-free representation that avoids the interpolation and finite-difference artifacts inherent in mesh-based schemes, while the intrinsic resolution remains controlled by the data wavelength content, resulting in stable solutions with phase velocity differences typically within 0.05 km/s relative to traditional approaches. The resulting phase velocity maps (20-80 s) reveal systematic structural variations that correlate well

with major tectonic units from the crust to the upper mantle. Low-velocity zones in the western Qilian Orogen, western Qinling Orogen, and Songpan-Ganzi Terrane contrast with high-velocity zones in the Ordos Block and central Qinling Orogen, reflecting significant tectonic differences in lithospheric strength and deformation behavior between actively deforming orogenic belts and adjacent stable blocks. With increasing periods, these lateral velocity contrasts become progressively weaker, indicating a transition toward more laterally homogeneous structures in the upper mantle. Methodologically, this work highlights the potential of combining deep learning with wave-equation-based physical constraints for seismic tomography, particularly in regions with sparse or uneven data coverage. The consistency between PINN-based teleseismic tomography, conventional methods, and ambient noise tomography suggests that this framework can reliably integrate complementary seismic observations within a physically consistent formulation. This integrated perspective enables robust imaging of lithospheric structure across a broad depth range and provides a coherent basis for interpreting tectonic processes along the northeastern margin of the Tibetan Plateau, while offering a flexible foundation for future joint inversions with other geophysical observations and applications in tectonically complex regions.

Open Research Section

The ChinArray-Himalaya II waveform data used in this study are archived by the International Earthquake Science Data Center of the Institute of Geophysics, China Earthquake Administration (`http://www.esdc.ac.cn/class/16`). In accordance with the data policy of the Institute of Geophysics, the raw waveforms are not publicly downloadable but are available to registered users on application through the online system of this data center. The traveltime and amplitude datasets, the Rayleigh wave phase velocity models, and the pinnTHT code used to generate them are openly available at Y. Chen (2026): `https://doi.org/10.5281/zenodo.22241036`.

Conflict of Interest Statement

The authors have no conflicts of interest to disclose.

Acknowledgments
We thank all those who participated in the field deployment of ChinArray-Himalaya II, and the China Seismic Array Data Management Center at the Institute of Geophysics, China Earthquake Administration, for providing the seismic records used in this study. We thank Rui Wang for his help in revising the manuscript. This work was supported by the National Natural Science Foundation of China (Grants 42430308, 42371067, and 92058212) and the Deep Earth Probe and Mineral Resources Exploration–National Science and Technology Major Project (Grant 2024ZD1000300). Sjoerd A.L. de Ridder wishes to acknowledge support through IEC\NSFC\223488.

References

Agata, R., Shiraishi, K., & Fujie, G. (2023). Bayesian seismic tomography based on velocity-space stein variational gradient descent for physics-informed neural network. *IEEE Transactions on Geoscience and Remote Sensing*.

Anderson, D. L. (1987). Global mapping of the uppermantle by surface wave tomography. *Composition, Structure and Dynamics of the Lithosphere-Asthenosphere System*, *16*, 89–97.

Araya-Polo, M., Jennings, J., Adler, A., & Dahlke, T. (2018). Deep-learning tomography. *The Leading Edge*, *37*(1), 58–66.

Baydin, A. G., Pearlmutter, B. A., Radul, A. A., & Siskind, J. M. (2018). Automatic differentiation in machine learning: a survey. *Journal of machine learning research*, *18*(153), 1–43.

Bianco, M. J., Gerstoft, P., Olsen, K. B., & Lin, F.-C. (2019). High-resolution seismic tomography of long beach, ca using machine learning. *Scientific reports*, *9*(1), 14987.

bin Waheed, U., Haghighat, E., Alkhalifah, T., Song, C., & Hao, Q. (2021). Pinneik: Eikonal solution using physics-informed neural networks. *Computers & Geosciences*, *155*, 104833.

Chai, X., Gu, Z., Long, H., Liu, S., Yang, T., Wang, L., ... Cao, W. (2024). Modeling multisource multifrequency acoustic wavefields by a multiscale fourier feature physics-informed neural network with adaptive activation functions. *Geophysics*, *89*(3), T79–T94.

Chen, P., Zhao, L., & Jordan, T. H. (2007). Full 3d tomography for the crustal structure of the los angeles region. *Bulletin of the Seismological Society of America*, *97*(4), 1094–1120.

Chen, Y. (2026). *Teleseismic helmholtz tomography of northeastern tibet using physics-informed neural networks. [dataset and software].* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.22241036` doi: 10.5281/zenodo.22241036

Chen, Y., de Ridder, S. A., Rost, S., Guo, Z., Wu, X., & Chen, Y. (2022). Eikonal tomography with physics-informed neural networks: Rayleigh wave phase velocity in the northeastern margin of the tibetan plateau. *Geophysical Research Letters*, *49*(21), e2022GL099053.

Chen, Y., de Ridder, S. A., Rost, S., Guo, Z., Wu, X., Li, S., & Chen, Y. (2023). Physics-informed neural networks for elliptical-anisotropy eikonal tomography: Application to data from the northeastern tibetan plateau. *Journal of Geophysical Research: Solid Earth*, *128*(12), e2023JB027378.

Elsayed, H. S., Saad, O. M., Soliman, M. S., Chen, Y., & Youness, H. A. (2023). Eqconvmixer: A deep learning approach for earthquake location from single-station waveforms. *IEEE Geoscience and Remote Sensing Letters*, *20*, 1–5.

Forsyth, D. W., Li, A., Levander, A., & Nolet, G. (2005). Array analysis of two-dimensional variations in surface wave phase velocity and azimuthal anisotropy in the presence of multipathing interference. *Geophysical Monograph-American Geophysical Union*, *157*, 81.

Forsyth, D. W., Webb, S. C., Dorman, L. M., & Shen, Y. (1998). Phase velocities of rayleigh waves in the melt experiment on the east pacific rise. *Science*, *280*(5367), 1235–1238.

Fu, L., Pan, L., Ma, Q., Dong, S., & Chen, X. (2021). Retrieving s-wave velocity from surface wave multimode dispersion curves with dispinet. *Journal of Applied Geophysics*, *193*, 104430.

Fu, Y., & Xiao, Z. (2020). Ambient noise tomography of rayleigh and love wave in northeast tibetan plateau and adjacent regions. *Chinese Journal of Geophysics*, *63*(3), 860–870.

Gaherty, J. B. (2001). Seismic evidence for hotspot-induced buoyant flow beneath the reykjanes ridge. *Science*, *293*(5535), 1645–1647.

Gaherty, J. B. (2004). A surface wave analysis of seismic anisotropy beneath eastern north america. *Geophysical Journal International*, *158*(3), 1053–1066.

Gaherty, J. B., & Jordan, T. H. (1995). Lehmann discontinuity as the base of an anisotropic layer beneath continents. *Science*, *268*(5216), 1468–1471.

Gaherty, J. B., Jordan, T. H., & Gee, L. S. (1996). Seismic structure of the upper mantle in a central pacific corridor. *Journal of Geophysical Research: Solid Earth*, *101*(B10), 22291–22309.

Gee, L. S., & Jordan, T. H. (1992). Generalized seismological data functionals. *Geophysical Journal International*, *111*(2), 363–390.

Golos, E., Fang, H., Yao, H., Zhang, H., Burdick, S., Vernon, F., ... Van der Hilst, R. (2018). Shear wave tomography beneath the united states using a joint inversion of surface and body waves. *Journal of Geophysical Research: Solid Earth*, *123*(6), 5169–5189.

Gou, R., Zhang, Y., Zhu, X., & Gao, J. (2023). Bayesian physics-informed neural networks for the subsurface tomography based on the eikonal equation. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1–12.

Grossmann, T. G., Komorowska, U. J., Latz, J., & Schönlieb, C.-B. (2024). Can physics-informed neural networks beat the finite element method? *IMA Journal of Applied Mathematics*, hxae011.

Haghighat, E., & Juanes, R. (2021). Sciann: A keras/tensorflow wrapper for scientific computations and physics-informed deep learning using artificial neural networks. *Computer Methods in Applied Mechanics and Engineering*, *373*, 113552.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, *2*(5), 359–366.

Huang, X., & Alkhalifah, T. (2022a). Pinnup: Robust neural network wavefield solutions using frequency upscaling and neuron splitting. *Journal of Geophysical Research: Solid Earth*, *127*(6), e2021JB023703.

Huang, X., & Alkhalifah, T. (2022b). Single reference frequency loss for multi-frequency wavefield representation using physics-informed neural networks. *IEEE Geoscience and Remote Sensing Letters*.

Jin, G., & Gaherty, J. B. (2015). Surface wave phase-velocity tomography based on multichannel cross-correlation. *Geophysical Journal International*, *201*(3), 1383–1398.

Jo, J. H., & Ha, W. (2023). Seismic traveltime tomography using deep learning. *IEEE Transactions on Geoscience and Remote Sensing*.

Karimpouli, S., & Tahmasebi, P. (2020). Physics informed machine learning: Seismic wave equation. *Geoscience Frontiers*, *11*(6), 1993–2001.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Lehujeur, M., & Chevrot, S. (2020). On the validity of the eikonal equation for surface-wave phase-velocity tomography. *Geophysical Journal International*, *223*(2), 908–914.

Li, A., & Li, L. (2015). Love wave tomography in southern africa from a two-plane-wave inversion method. *Geophysical Journal International*, *202*(2), 1005–1020.

Li, H., Shen, Y., Huang, Z., Li, X., Gong, M., Shi, D., ... Li, A. (2014). The distribution of the mid-to-lower crustal low-velocity zone beneath the northeastern tibetan plateau revealed from ambient noise tomography. *Journal of Geophysical Research: Solid Earth*, *119*(3), 1954–1970.

Li, L., Cai, C., Fu, Y., & Fang, H. (2023). Multiple surface wave tomography methods and their applications to the tibetan plateau. *Reviews of Geophysics and Planetary Physics*, *54*(2), 174–196.

Li, S., Guo, Z., Yu, Y., Wu, X., & Chen, Y. J. (2022). Imaging the northeastern crustal boundary of the tibetan plateau with radial anisotropy. *Geophysical Research Letters*, *49*(23), e2022GL100672.

Li, Y., Pan, J., Wu, Q., & Ding, Z. (2017). Lithospheric structure beneath the northeastern Tibetan Plateau and the western Sino-Korea Craton revealed by Rayleigh wave tomography. *Geophysical Journal International*, *210*(2), 570–584.

Li, Y., Pensoneault, A., Zhang, Y., Zhu, X., Gou, R., & Gao, J. (2025). Efficient bayesian physics-informed neural networks for surface tomography via ensemble kalman inversion. *Geophysics*, *90*(3), U1–U14.

Lin, F.-C., & Ritzwoller, M. H. (2011). Helmholtz surface wave tomography for isotropic and azimuthally anisotropic structure. *Geophysical Journal International*, *186*(3), 1104–1120.

Lin, F.-C., Ritzwoller, M. H., & Snieder, R. (2009). Eikonal tomography: surface wave tomography by phase front tracking across a regional broad-band seismic array. *Geophysical Journal International*, *177*(3), 1091–1110.

Liu, B., Yang, S., Ren, Y., Xu, X., Jiang, P., & Chen, Y. (2021). Deep-learning seismic full-waveform inversion for realistic structural models. *Geophysics*, *86*(1), R31–R44.

Lu, C., Wang, Y., Zou, X., Zong, J., & Su, Q. (2024). Elastic full-waveform inversion via physics-informed recurrent neural network. *IEEE Transactions on Geoscience and Remote Sensing*.

Maguire, R., Schmandt, B., Wang, R., Kong, Q., & Sanchez, P. (2024). Generalization of deep-learning models for classification of local distance earthquakes and explosions across various geologic settings. *Seismological Research Letters*, *95*(4), 2229–2238.

Moseley, B., Markham, A., & Nissen-Meyer, T. (2020). Solving the wave equation with physics-informed deep learning. *arXiv preprint arXiv:2006.11894*.

Mousavi, S. M., Beroza, G. C., Mukerji, T., & Rasht-Behesht, M. (2024). Applications of deep neural networks in exploration seismology: A technical survey. *Geophysics*, *89*(1), WA95–WA115.

Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C. (2020). Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature communications*, *11*(1), 3952.

Muller, A. P., Bom, C. R., Costa, J. C., Klatt, M., Faria, E. L., Silva, B. d. S., ... de Albuquerque, M. P. (2023). Deep-tomography: iterative velocity model building with deep learning. *Geophysical Journal International*, *232*(2), 975–989.

Nishida, K., Montagner, J.-P., & Kawakatsu, H. (2009). Global surface wave tomography using seismic hum. *Science*, *326*(5949), 112–112.

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2017). Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations. *arXiv preprint arXiv:1711.10561*.

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, *378*, 686–707.

Rasht-Behesht, M., Huber, C., Shukla, K., & Karniadakis, G. E. (2022). Physics-Informed Neural Networks (PINNs) for Wave Propagation and Full Waveform Inversions. *Journal of Geophysical Research: Solid Earth*, *127*(5), e2021JB023120.

Ren, P., Rao, C., Chen, S., Wang, J.-X., Sun, H., & Liu, Y. (2024). Seismicnet: Physics-informed neural networks for seismic wave modeling in semi-infinite domain. *Computer Physics Communications*, *295*, 109010.

Sabra, K. G., Gerstoft, P., Roux, P., Kuperman, W., & Fehler, M. C. (2005). Surface wave tomography from microseisms in southern california. *Geophysical Research Letters*, *32*(14).

Sato, Y. (1955). Analysis of dispersed surface waves by means of fourier transform i. *Bull. Earthq. Res. Inst*, *33*, 33–48.

Schaeffer, A., & Lebedev, S. (2013). Global shear speed structure of the upper mantle and transition zone. *Geophysical Journal International*, *194*(1), 417–449.

Sergeant, A., Delage, K., & Chevrot, S. (2025). Iterative helmholtz surface wave tomography using generalized wave equation smoothing splines—application to western europe. *Journal of Geophysical Research: Solid Earth*, *130*(2), e2024JB030087.

Seydoux, L., Balestriero, R., Poli, P., Hoop, M. d., Campillo, M., & Baraniuk, R. (2020). Clustering earthquake signals and background noises in continuous seismic data with unsupervised deep learning. *Nature communications*, *11*(1), 3972.

Shapiro, N. M., Campillo, M., Stehly, L., & Ritzwoller, M. H. (2005). High-resolution surface-wave tomography from ambient seismic noise. *Science*, *307*(5715), 1615–1618.

Song, C., Alkhalifah, T., & Waheed, U. B. (2021). Solving the frequency-domain acoustic vti wave equation using physics-informed neural networks. *Geophysical Journal International*, *225*(2), 846–859.

Song, C., & Alkhalifah, T. A. (2021). Wavefield reconstruction inversion via physics-informed neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–12.

Song, C., Geng, H., Wang, Y., Waheed, U. B., & Liu, C. (2025). Simultaneous p-and s-wave seismic traveltime tomography using physics-informed neural networks. *Geophysical Prospecting*.

Song, C., & Wang, Y. (2023). Simulating seismic multifrequency wavefields with the fourier feature physics-informed neural network. *Geophysical Journal International*, *232*(3), 1503–1514.

Tsai, V. C., & Atiganyanun, S. (2014). Green's functions for surface waves in a generic velocity structure. *Bulletin of the Seismological Society of America*, *104*(5), 2573–2578.

Wielandt, E. (1993). Propagation and structural interpretation of non-plane waves. *Geophysical Journal International*, *113*(1), 45–53.

Woodhouse, J. H., & Dziewonski, A. M. (1984). Mapping the upper mantle: Three-dimensional modeling of earth structure by inversion of seismic waveforms. *Journal of Geophysical Research: Solid Earth*, *89*(B7), 5953–5986.

Wu, X., Guo, Z., Li, S., Yu, Y., Bai, Q., & Chen, Y. J. (2023). Seismic azimuthal anisotropy of northeastern tibetan plateau from ambient noise double beamforming tomography: Implications for crustal deformation. *Journal of Geophysical Research: Solid Earth*, *128*(6), e2022JB026109.

Wu, Y., Aghamiry, H. S., Operto, S., & Ma, J. (2023). Helmholtz-equation solution in nonsmooth media by a physics-informed neural network incorporating quadratic terms and a perfectly matching layer condition. *Geophysics*, *88*(4), T185–T202.

Xue, Z., Wang, Y., Wu, X., & Ma, J. (2025). Multi-geophysical information neural network for seismic tomography. *Geophysics*, *90*(3), R89–R100.

Yang, F., & Ma, J. (2019). Deep-learning inversion: A next-generation seismic velocity model building method. *Geophysics*, *84*(4), R583–R599.

Yang, F., & Ma, J. (2023). Fwigan: Full-waveform inversion via a physics-informed generative adversarial network. *Journal of Geophysical Research: Solid Earth*, *128*(4), e2022JB025493.

Yao, H., van Der Hilst, R. D., & De Hoop, M. V. (2006). Surface-wave array tomography in se tibet from ambient seismic noise and two-station analysis—i. phase velocity maps. *Geophysical Journal International*, *166*(2), 732–744.

Yao, H., Xu, G., Zhu, L., & Xiao, X. (2005). Mantle structure from inter-station rayleigh wave dispersion and its tectonic implication in western china and neighboring regions. *Physics of the Earth and Planetary Interiors*, *148*(1), 39–54.

Zhang, J., Zhu, H., Li, Z., Wu, X., & Zhang, J. (2025). Multistation seismic location via machine learning: application to oklahoma and southern california. *Geophysical Journal International*, *241*(3), 1852–1866.

Zhang, Q., Sandvol, E., Ni, J., Yang, Y., & Chen, Y. J. (2011). Rayleigh wave tomography of the northeastern margin of the tibetan plateau. *Earth and Planetary Science Letters*, *304*(1-2), 103–112.

Zhang, X., Zhang, J., Yuan, C., Liu, S., Chen, Z., & Li, W. (2020). Locating induced earthquakes with a network of seismic stations in oklahoma via a deep learning method. *Scientific reports*, *10*(1), 1941.

Zhong, S.-J., Wu, J.-P., Fang, L.-H., Wang, W.-L., Fan, L.-P., & Wang, H.-F. (2017). Surface wave eikonal tomography in and around the northeastern margin of the tibetan plateau. *Chinese Journal of Geophysics*, *60*(6), 2304–2314.

Zhu, J., Fang, L., Miao, F., Fan, L., Zhang, J., & Li, Z. (2024). Deep learning and transfer learning of earthquake and quarry-blast discrimination: applications to southern california and eastern kentucky. *Geophysical Journal International*, *236*(2), 979–993.

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

Zhu, W., Biondi, E., Li, J., Yin, J., Ross, Z. E., & Zhan, Z. (2023). Seismic arrival-time picking on distributed acoustic sensing data using semi-supervised learning. *Nature Communications*, *14*(1), 8192.