
GlobalTomo: A global dataset for physics-ML seismic wavefield modeling and FWI

Shiqian Li ^{*,1,3,4} shiqianli@stu.pku.edu.cn	Zhi Li ^{*,2} zl382@pku.edu.cn	Zhancun Mu ¹ muzhancun@stu.pku.edu.cn
Shiji Xin ¹ shijixin@g.harvard.edu	Zhixiang Dai ⁵ zhixiangd@nvidia.com	Kuangdai Leng ⁶ kuangdai.leng@stfc.ac.uk
Ruihua Zhang ⁵ ritaz@nvidia.com	Xiaodong Song ^{2,1,✉} xdsong@pku.edu.cn	Yixin Zhu ^{3,1,4,✉} yixin.zhu@pku.edu.cn

✱ equal contributors ✉ corresponding authors

¹ Institute for Artificial Intelligence, Peking University

² State Key Laboratory of Deep Earth and Mineral Exploration,
School of Earth and Space Sciences, Peking University

³ School of Psychological and Cognitive Sciences, Peking University

⁴ Beijing Key Laboratory of Behavior and Mental Health, Peking University

⁵ NVIDIA ⁶ Rutherford Appleton Laboratory

The code and data for reproducing the results are available at <https://global-tomo.github.io>

Abstract

Global seismic tomography, taking advantage of seismic waves from natural earthquakes, provides essential insights into the earth’s internal dynamics. Advanced Full-Waveform Inversion (FWI) techniques, whose aim is to meticulously interpret every detail in seismograms, confront formidable computational demands in forward modeling and adjoint simulations on a global scale. Recent advancements in Machine Learning (ML) offer a transformative potential for accelerating the computational efficiency of FWI and extending its applicability to larger scales. This work presents the first 3D global synthetic dataset tailored for seismic wavefield modeling and full-waveform tomography, referred to as the Global Tomography (GlobalTomo) dataset. This dataset is comprehensive, incorporating explicit wave physics and robust geophysical parameterization at realistic global scales, generated through state-of-the-art forward simulations optimized for 3D global wavefield calculations. Through extensive analysis and the establishment of ML baselines, we illustrate that ML approaches are particularly suitable for global FWI, overcoming its limitations with rapid forward modeling and flexible inversion strategies. This work represents a cross-disciplinary effort to enhance our understanding of the earth’s interior through physics-ML modeling.

1 Introduction

Global seismic tomography is a crucial yet intricate field within the earth sciences that facilitates understanding the earth’s internal structure and dynamics. This discipline has wide-ranging applications, from the discovery of natural resources [69] and the assessment of seismic hazards [75] to the exploration of our planet’s evolutionary history [11]. It utilizes seismic signals—earthquake-induced ground vibrations—recorded at surface stations to invert and map the earth’s interior. The fundamental challenge in this field lies in the integration of forward modeling of seismic wave propagation with FWI, both of which are essential for accurately correlating seismic data with velocity structures [55, 16]. However, accurately simulating seismic waves through the earth’s complex structures and inverting this data is challenging due to the heterogeneous nature of terrestrial media and the inherently ill-posed characteristics of FWI [65, 41].

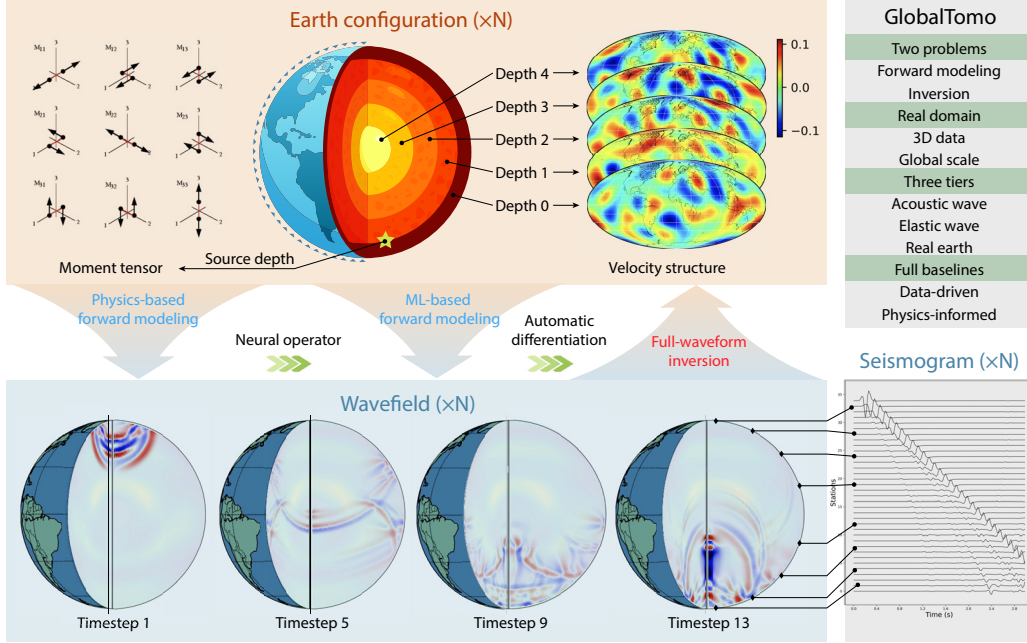


Figure 1: **Overview of GlobalTomo.** GlobalTomo is meticulously designed to tackle the pressing challenges associated with global seismic wavefield modeling and full-waveform inversion via cutting-edge physics-ML methodologies. In the **forward modeling** process, given specific source and velocity structures, the goal is to predict the wavefield at various time steps and the resulting seismograms at surface stations. The **inversion** process utilizes these seismograms as observational data to deduce the underlying velocity structures. Advanced ML techniques enhance these processes through neural operator learning and rapid automatic differentiation, substantially improving the efficiency of both forward modeling and inversion tasks.

Modern numerical simulation methods, such as finite difference methods [23], finite element methods [78, 43], and spectral element methods [24, 25, 27], are fundamental in the forward modeling of seismic waves. Although these methods have proven effective, they impose significant computational demands, particularly when deploying higher-resolution models necessary for detailed geophysical interpretation. High-resolution imaging requires the simulation of high-frequency waves, which in turn necessitates smaller time steps and finer spatial discretization. This is essential to prevent numerical dispersion and maintain the stability of the simulations, as mandated by the Courant-Friedrichs-Lewy condition [9]. Consequently, as the desired resolution increases, so does the computational burden, exponentially escalating the required computational resources and time, thereby introducing a substantial bottleneck [37].

The computational intensity of forward modeling significantly constrains the efficiency of FWI. FWI aims to iteratively refine a model of the earth’s structure by minimizing the discrepancies between observed and simulated seismic data [48]. The prolonged computation times required for forward modeling restrict the exploration of the model parameter space and limit the number of feasible iterations [3, 29]. This limitation is particularly problematic when attempting high-resolution inversions over extensive areas. Consequently, this situation highlights the critical need for more efficient methodologies in both seismic forward modeling [45] and inversion strategies [1].

Recent advances in ML have shown significant promise in mitigating the aforementioned computational challenges in seismic tomography. Techniques such as Physics-Informed Neural Network (PINN) [7] and Neural Operator Learning [39, 35, 70] efficiently model complex physical systems by approximating solutions to the underlying Partial Differential Equations (PDEs). These methods have been successfully applied in fields including fluid dynamics prediction [4], new material discovery [68], climate change modeling [46, 47], and enhancing biomedical treatments [38]. By significantly reducing computational time compared to traditional numerical simulations, these ML techniques enable more efficient high-resolution seismic modeling and inversion. Furthermore, modern ML approaches can generalize beyond trained scenarios, facilitating the exploration of previously uncharted earth’s interior.

However, the efficacy of ML methods in geophysical applications is heavily contingent upon the availability of extensive, high-quality training data. Currently, there is a significant lack of benchmark synthetic datasets specifically designed for global seismic tomography and the ML community. Although datasets to support FWI [10, 15] have been introduced, they are mostly limited to subsurface exploration scenarios and do not fully represent the earth’s entire scale, which is crucial for high-resolution wave propagation modeling. This deficiency highlights a critical gap in the resources needed to advance ML applications in this area.

In response to these challenges, we introduce GlobalTomo, the first comprehensive 3D global full-waveform dataset specifically crafted for the ML community. Designed to support both forward modeling and inverse problem studies, GlobalTomo includes high-resolution seismic simulations that extend from the earth’s surface to its core. It features three data tiers and integrates acoustic and elastic wave equations to cater to diverse seismic phenomena, enabling both data-driven and physics-informed training. The dataset’s scope varies from a local scale of a 1-km radius simulated at 20Hz to a global scale encompassing the entire earth’s radius of 6,371 km at a 30-second period.

GlobalTomo’s design is grounded in well-established geological studies [13, 18, 65] that provide a credible distribution of earth’s internal properties, ensuring our model is adhering to recognized physical and structural characteristics. For the specific application of earth tomography, we avoid out-of-distribution scenarios, as they involve hypothetical or artificially generated distributions that may lack geological relevance. We employ spherical harmonics to effectively parameterize velocity structures and account for variable seismic sources. Extensive benchmark analyses demonstrate GlobalTomo’s capabilities in fostering generalization, accelerating forward modeling, and addressing the inherent ill-posedness of seismic inversion tasks. By providing GlobalTomo, we aim to bridge the existing gap between geophysical sciences and the ML community, thereby enabling more precise and efficient seismic modeling and inversion processes through ML-driven innovations [44].

1.1 Related Work

Full-Waveform Inversion FWI has been extensively applied across a variety of domains, including shallow crustal imaging [62, 20], continental tomography [5, 77], medical imaging [56, 71], and nondestructive testing [50]. In seismology, classical FWI employs numerical simulations to derive seismograms from an initial earth model and iteratively optimizes this model using the adjoint-state method [63, 66]. While it has achieved significant success in both local and global inversions [17, 19], FWI still confronts substantial challenges such as susceptibility to local minima [65] and high computational demands especially when Bayesian methods are applied [73]. Emerging ML approaches, bolstered by comprehensive global seismic datasets, present promising avenues to address these challenges, potentially leading to more efficient and precise FWI solutions.

Neural Operator Learning Neural operator learning has garnered significant attention for solving parametric PDEs, as evidenced by a range of pioneering studies [26, 36, 35, 70, 39]. Among these, DeepONet stands out for its flexibility in modeling mesh-free spaces and producing continuous solutions [39, 70]. Conversely, the Fourier Operator Learning (FNO) operates with mesh-based inputs and outputs yet efficiently models the frequency domain and demonstrates rapid convergence during training [35, 40]. Once trained, neural operators can swiftly predict solutions for new parameter sets with just a single forward pass. This capability is particularly advantageous in seismic applications, enabling the instant modeling of wave propagation under various velocity structures during inference. Such efficiency significantly streamlines the inversion process when compared to PINN methods [57, 51].

2 Dataset Construction

2.1 Problem Definition

The primary goal of forward modeling here is to solve the 3D acoustic and elastic wave equations governed by interior PDEs and boundary conditions detailed in [Appendix B.1](#). The objective is to predict the wavefield $\phi_{c,s}(p, t)$ at any spatial location $p = (x, y, z)$ and time t across various velocity structures c and source configurations s . This requires constructing and learning a nonlinear operator G , which maps velocity and source information into a wavefield output function, $G(c, s)$.

The velocity structure c is captured by sensor measurements at predetermined points, expressed as $c = [c(p_1), c(p_2), \dots, c(p_m)]$. The source s includes spatial locations and moment tensors. Given inputs p and t , the function $G(c, s)$ computes the wavefield $\phi_{c,s}(p, t)$. This approach models wave propagation dynamics, aiding in accurate simulations and deeper geophysical understanding.

In particular, we monitor the displacements received by surface stations, known as seismograms, to infer the earth’s underlying structures. The inversion problem can be formulated as:

$$c^* = \arg \min_c J(\mathbf{d}, \Phi(c, s)) + \lambda F(c, s). \quad (1)$$

Here, \mathbf{d} represents the observed seismogram data. The function $\Phi(c, s)$ computes the seismograms based on the velocity structure c and the source s . The term $F(c, s)$ is a regularization function that integrates prior knowledge and assumptions about the structure, while λ is a coefficient that modulates the influence of the regularization term. The notation J denotes the L1 or L2 norm, used here to measure the misfit between the observed data and the predictions.

Traditional FWI relies on numerical simulations to compute $\Phi(c, s)$ and uses adjoint methods for optimizing c . Due to the intensive computational demand of simulations, the optimization process is typically limited to a small number of iterations. We view ML methods as a potent tool to expedite the forward simulation process and enable more flexible optimization strategies. The solutions of modern ML approaches in addressing inversion problems are further discussed in [Appendix G.1](#).

2.2 Model Configuration

For generating global wavefield and seismogram data for our dataset, we employ AxiSEM3D [32], a forward simulator renowned for its efficiency in simulating global seismic wave propagation. AxiSEM3D is exceptionally adaptable, allowing customization to various model complexities [31]. This simulator effectively utilizes the axisymmetry of global wavefields in a source-centered frame and extends calculations into the azimuthal Fourier domain. Further details are discussed in [Appendix B.2](#).

Our dataset comprises three tiers of increasing complexity, each supporting different training scales and applications. The first tier simulates acoustic wave propagation through a 1-km radius fluid sphere using a 20Hz mesh, providing a foundation in basic wave physics. The second tier extends simulations to elastic wave propagation through an isotropic solid medium of the same scale, with source variations. This scale is comparable to 600 km continental simulations at a 30-second period, allowing exploration of complex P/S converted interactions. The third tier targets real-world earth applications, integrating acoustic and elastic simulations from the earth’s surface to its core over a 30-second period, addressing planetary-scale wave propagation for detailed geophysical analyses.

Each model’s 3D structure within these tiers is developed by introducing perturbations ranging from -10% to 10% to a 1D background model. These perturbations are meticulously designed to mirror the subtle yet significant 3D heterogeneity of the earth’s interior, as uncovered by global tomographic studies [13, 18, 64]; even earth’s most extreme features, such as subduction zones and mantle plumes, show velocity variations of less than 10% [54, 64]. Such adjustments are crucial for demonstrating the impact of tomographic features on earth’s geochemical and geodynamical processes, including mantle convection [67], subduction of tectonic plates [60], and deep mantle heterogeneity such as the large low-shear-velocity provinces [6]. Further details on the configurations of each model are provided in [Appendix C](#).

Parameterization Perturbations within each defined layer are parameterized using real spherical harmonics [8] up to degree 8, with radial values interpolated linearly between layers. Spherical harmonics, employed as the basis for parameterization, are mathematical functions that define a set of orthogonal functions on the sphere. These functions are particularly advantageous in geophysical applications because they can naturally represent the spherical geometry of the earth. Although some current mantle tomographic models offer higher resolution [53], spectrum analysis indicates that the significant power predominantly resides at lower degrees [42, 54]. Consequently, our choice to limit structural parameterization to degree 8 in the published dataset effectively captures the predominant long-wavelength heterogeneity of the earth’s interior [59]. This approach reduces the number of parameters and enhances the solvability of this inverse problem.

Table 1: **Simulation input and output size across three dataset tiers.** The upper section of the table provides basic information about the three data tiers, including specific parameters and configurations. The lower section of the table details the dimensions of time, spatial elements, and predicted seismic data for each tier.

Category	Variables	Acoustic			Elastic			Real Earth		
Input	Sample number	10,000			30,000			10,000		
	Background	Uniform			Uniform			PREM		
	Time range	0-3 s			0-3 s			0-6000 s		
	Source	0			6			9		
	Structure	405			1215			5427		
Dimension		Time	Element	Seis	Time	Element	Seis	Time	Element	Seis
Output	Wavefield	15	58,368	1	15	58,368	3	20	263,520	3
	Fourier	15	3,648	16	15	3,648	48	20	16,470	48
	Seismogram	150	1,369	1	150	1,369	3	6,000	703	3
	Storage	88.54 GB			657.28 GB			1795.24 GB		

2.3 Data Generation

Table 1 details simulation inputs and outputs. Using Latin Hypercube Sampling (LHS), we generated a set of random velocity structures and source configurations for extensive parameter coverage in the area of interest. Simulations were performed on a Dell PowerEdge R740 cluster with Intel Xeon Scalable processors (2,336 cores, up to 2.9GHz, 13,056GB RAM). Forward modeling for Acoustic and Elastic tiers required about 2,800 and 8,400 CPU hours respectively, while seismic simulations for the Real Earth tier needed about 100,000 CPU hours. Despite the high computational demands, this method is more cost-effective than traditional ones [64].

Our dataset comprises two types of output data. The first type is surface seismogram data, which includes one-dimensional time series recordings of ground displacement at each surface station. These seismograms are observable globally and are crucial for the inversion of the earth’s internal structure. The second type is seismic wavefield data, capturing the propagating seismic wavefield within the earth’s interior. Although this spatially dense wavefield cannot be directly measured and used for traditional inversion, it reveals complex interactions between seismic waves and the earth’s three-dimensional structure. The intricacies of the wavefield, which contain rich details, have recently gained attention in global wavefield studies [30] but are often too challenging to identify manually. We provide this intermediate wavefield data to facilitate the training of neural networks. For detailed configurations and distribution of each type of output data, please refer to [Appendix C](#). We compare the related dataset with GlobalTomo from various aspects in [Appendix E](#).

3 Experiments

We experiment with several ML baselines as the general solution operator for the 3D acoustic and elastic wave equations to accompany GlobalTomo. These baselines, trained on the Acoustic and Elastic tiers, serve as practical workflows for forward modeling tasks, predicting wavefields and seismograms based on given velocity structures and source parameters. Wavefield predictions include 7 timesteps from 0 to 3 seconds, with each timestep consisting of 16 slices and 3,648 points per slice. Seismogram predictions cover a series of time series of length 150 for 1,369 stations. These forward modeling baselines are not only utilized to simulate data but also employed in inversion tasks to refine the velocity structures based on their predictions. See [Appendix D](#) for pipelines.

3.1 Baselines

We implement our baselines using NVIDIA Modulus Sym [22], a framework designed for high-performance scientific computing.

Mean Model (MM) The MM baseline calculates the average results from the training data and uses these averages as predictions for the test data. Although this method does not capture complex data patterns, it serves as a useful benchmark for assessing the performance of more sophisticated models. Additionally, it provides insights into the overall distribution of the data.

Table 2: **Quantitative results of baselines.** Performance of baseline models in forward modeling is evaluated on the test set using RL2 and R. Results include both mean and standard deviation values. The average inference time for each model is also presented.

Wave	Model	Representation	Wavefield		Seismogram		Time (ms)
			RL2 ↓	R ↑	RL2 ↓	R ↑	
Acoustic	MM	Vector	0.495±0.048	0.871±0.026	0.597±0.069	0.802±0.053	-
	MLP	Vector	0.356±0.035	0.937±0.012	0.446±0.051	0.896±0.026	1.771
	H-Fourier	Vector	0.397±0.054	0.917±0.066	0.594±0.073	0.820±0.068	2.604
	DeepONet	Point	0.368±0.050	0.927±0.066	0.503±0.061	0.862±0.067	2.898
	GNOT	Point	0.300±0.014	0.954±0.004	0.564±0.041	0.829±0.026	2.985
Elastic	MM	Vector	1.000±0.001	-0.006±0.113	1.000±0.001	0.001±0.116	-
	MLP	Vector	0.617±0.038	0.790±0.027	0.534±0.068	0.848±0.039	1.773
	H-Fourier	Vector	0.735±0.061	0.689±0.048	0.613±0.067	0.790±0.049	2.397
	DeepONet	Point	0.682±0.035	0.734±0.031	0.565±0.065	0.826±0.043	2.775
	GNOT	Point	0.665±0.032	0.752±0.035	0.621±0.062	0.769±0.079	2.875

Multilayer Perceptron (MLP) The MLP baseline consists of a 6-layer fully connected neural network, each layer having a hidden size of 500 and utilizing SiLU activation functions. This model takes velocity structures as inputs and predicts both wavefields and seismograms.

Highway Fourier Net (H-Fourier) Considering the tendency of neural networks to favor low-frequency solutions [49], we explore the use of a H-Fourier. This model combines the gating mechanisms of Highway Networks [58] with the frequency domain learning capabilities of Fourier Neural Networks [61]. The H-Fourier is designed to handle deep learning tasks requiring both complex layer-to-layer transformations and frequency-specific learning, potentially more adept at managing intricate patterns in seismic data.

DeepONet Instead of directly mapping velocity structures to fixed output grids, this approach involves building a framework that models both wave propagation and seismogram generation on a continuous spatial-temporal domain. We utilize the DeepONet [39], where the trunk net processes inputs of positions and times, and the branch net handles velocity structures. The outputs from both nets are combined through a dot product before passing through a linear output layer to produce the final prediction. To improve training efficiency, we developed an optimized method for parallel multiplication of the trunk and branch net outputs; see [Appendix F.2](#).

Physics-Informed DeepONet (PIDO) While DeepONet is trained using standard supervised learning, which might limit its generalization in high-resolution domains, we explore additional capabilities with PIDO [70]. This model enforces physically consistent representations by incorporating interior and free-surface constraints during training. These constraints are integrated by calculating the derivatives within the PDEs using automatic differentiation.

General Neural Operator Transformer (GNOT) The GNOT [21] leverages the strengths of both neural operators and transformer architectures to effectively capture intricate spatial and temporal dependencies in data. By utilizing a self-attention mechanism, GNOT can efficiently model long-range interactions and handle high-dimensional inputs, enabling accurate predictions of dynamic systems like seismic wave propagation.

3.2 Results

3.2.1 Forward Modeling

Generalization to unseen structures To investigate the ability of the baseline models to generalize across unseen seismic structures, we trained them using the last 90% of the velocity structures available in our dataset. The first 10% of the structures were used as a test set to evaluate the models' performance. Quantitative assessments were carried out using Relative L2 Loss (RL2) and Person Correlation Coefficient (R) metrics.

Differences in baseline performance highlight the strengths of various modeling approaches. As detailed in [Table 2](#), MLP achieves high correlation coefficients and low RL2 values due to its effi-

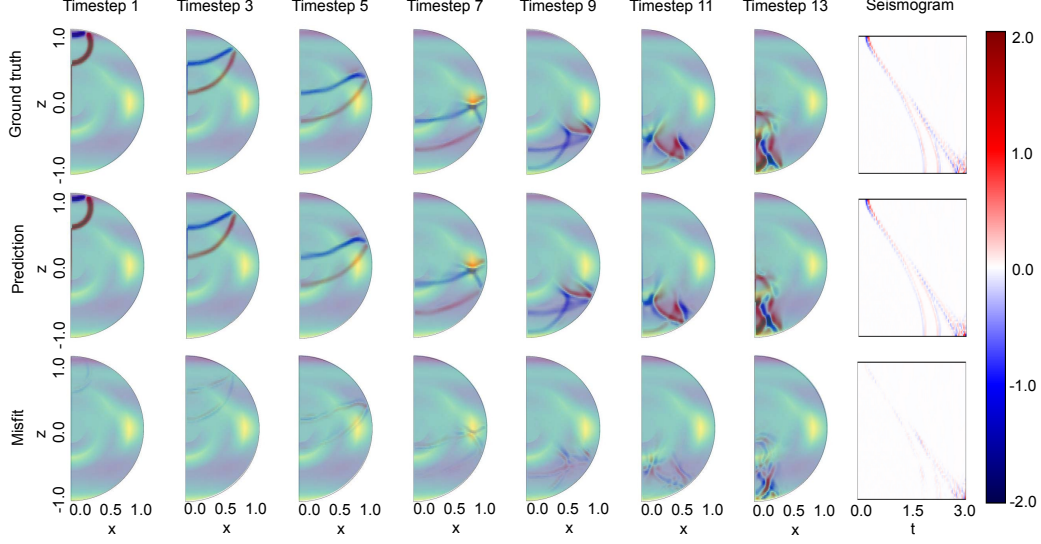


Figure 2: **Qualitative forward modeling prediction by MLP in the Acoustic tier.** A single slice is displayed. The source is located at $x = 0.00$ and $z = 0.80$. The background illustrates the velocity structure. The seismogram depicts the time series received by stations around the surface of this slice.

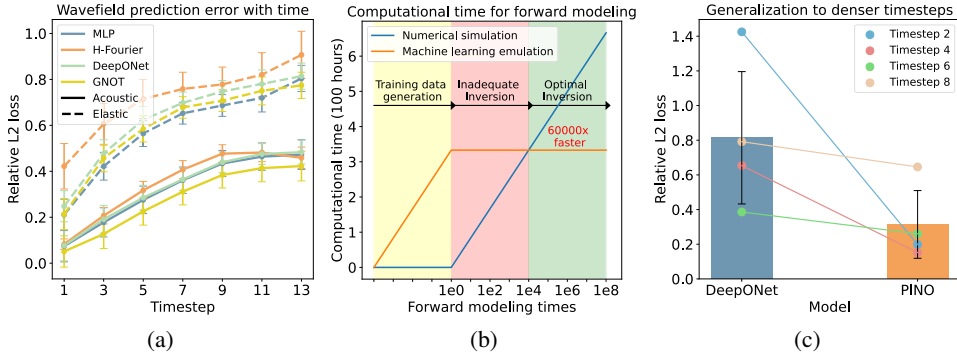


Figure 3: **Meta-analysis of forward modeling.** (a) RL2, used to quantify error in wavefield prediction, shows increasing trends over time across baseline models. (b) Once trained, ML emulation significantly outpaces numerical simulations in speed, facilitating more iterations in inversion processes. (c) DeepONet, when trained on timesteps 1, 3, 5, and 7, struggles to generalize to intermediate timesteps such as 2, 4, 6, and 8. Incorporating physical constraints during training improves model performance on these denser timesteps.

cacy in capturing global patterns from vector-based representations even when the seismic signals are sparse. Observations from training indicate that the H-Fourier model reaches low training errors quickly, suggesting its proficiency in capturing frequency-based signals; however, it tends to overfit the periodic details, resulting in limited generalization capabilities compared to the MLP. While vector-based methods excel with structured data, they face challenges in predicting arbitrary coordinates and time steps. In contrast, the point-based DeepONet and GNOT offer greater flexibility in generalizing across different coordinates and time steps, supporting mesh-free predictions. Nonetheless, it requires more computational resources during training to achieve comparable performance to the MLP. With similar width and depth, the model parameters of MLPs are 20 times (104.31M v.s. 5.06M) larger than those in the point-based models. Further training details are available in [Appendix F](#). Qualitative evaluations, illustrated in [Figure 2](#), corroborate these findings. The ML baselines closely match the actual data across several time steps, capturing the complex dynamics of wave propagation on unseen structures. As time advances, emulation errors tend to increase, reflecting the growing complexity of wave dynamics over periods, as depicted in [Figure 3a](#).

Complexity analysis The complexity of the acoustic and elastic waves comes from the intrinsic mechanical properties and mathematical formulation. For instance, the acoustic wavefield only has compressional (P) waves, while the elastic wavefield contains complex P/S converted energy. The

performance of MM can serve as an indicator to reflect the dataset’s complexity and distribution characteristics. Smaller error margins suggest more consistent patterns across velocity structures. The Elastic tier, evidenced by the poorer performance of MM, exhibits greater complexity than the Acoustic tier due to the intrinsic properties of the elastic wave equation and variations in source dynamics. This trend is also corroborated by the performance observed across ML baselines.

Forward modeling speed A single round of forward simulation using a numerical solver requires 120 seconds with 24 CPU cores. In contrast, ML approaches achieve forward modeling speeds typically ranging from 1 to 3 milliseconds on a single GPU, making them about 60,000 times faster than traditional methods. Note that the acceleration rate may vary with the degree of parallelism. While most computational costs for ML seismic modeling are incurred during the generation of training data, once trained, the neural operator can predict all wavefields in a single inference step. This underscores the potential of ML approaches for real-time seismic modeling and extensive forward modeling applications. For a comparative visualization, see [Figure 3b](#).

Generalization to higher temporal resolution To evaluate the flexibility of FWI in real-world scenarios through continuous wavefield prediction across the entire temporal domain, we tested the generalizability of a trained neural operator to higher resolution snapshots. Specifically, we utilized DeepONet, which accommodates mesh-free inputs. DeepONet was trained on a uniform velocity structure using acoustic waves at timesteps 1, 3, 5, and 7, and subsequently evaluated on a denser sequence from timesteps 1 to 8, within the same temporal range. The findings, illustrated in [Figure 3c](#), show that DeepONet, being purely data-driven, initially struggled with adapting to this higher-resolution domain. To enhance its generalization capabilities, we incorporated physical constraints (detailed in [Appendix B.1](#)) into its training process. This integration of physical principles significantly improved DeepONet’s predictions on previously unseen timesteps, reducing its dependency on extensive labeled training datasets for developing a super-resolution model.

3.2.2 Inversion

In this section, we explore how the generalization and acceleration capabilities of ML methods in forward modeling can significantly improve the inversion process. We present three strategies to address the challenges inherent in traditional FWI and enhance the efficiency of inversion.

Gradient-based optimization Following traditional FWI methods, we initially set fixed parameters for established baselines and iteratively refine the velocity model to minimize the L1 misfit between the predicted and actual seismograms. Our optimization strategy employs 200 iterations of gradient descent using the LBFGS algorithm, with a learning rate of 0.08 and a history size of 30. A regularization term is integrated to promote stable convergence. We explore the effectiveness of gradient information from ML forward models in aiding the inversion optimization process. For each test structure, five initial points are randomly selected, and optimization is performed using the forward models pre-trained on the acoustic tier. The inversion performance is assessed using R across various structural scales with 80, 245, and 405 free parameters corresponding to degrees 4, 6, and 8, respectively. We show the optimization process of the best-performing model in [Figure 4](#). The correlation between the inverted structure and the ground truth strengthens with each iteration step, confirming that ML-derived gradients can effectively guide the inversion within 200 steps.

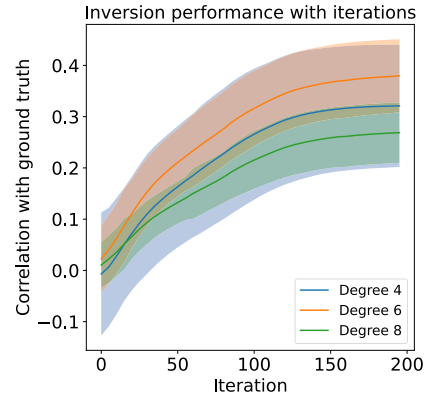


Figure 4: **The inversion optimization process.** Performance improves through optimization across 200 iterations. Higher degrees capture increasingly shorter-wavelength structures, enhancing model fidelity.

Sampling with multiple starting points The initial conditions significantly influence the performance of inversion. To enhance inversion, we experimented with starting the optimization from multiple initial points. Specifically, we used ML forward modeling to generate 1,000 random start-

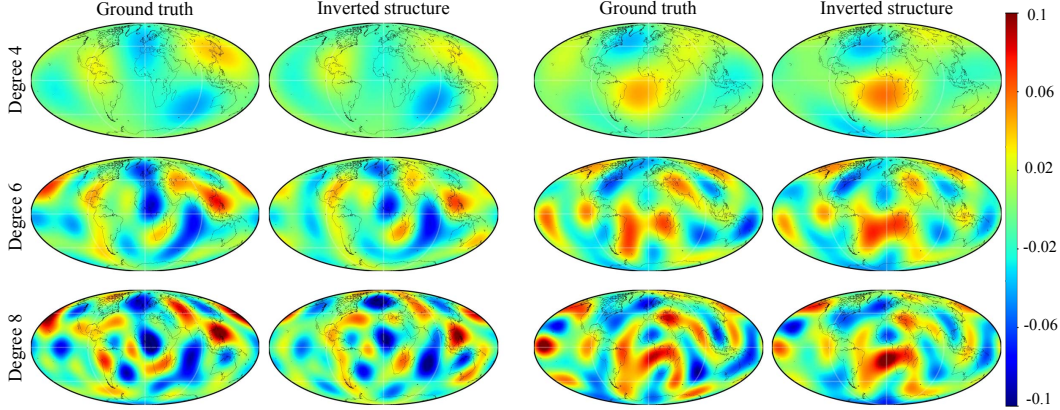


Figure 6: **Qualitative inversion results.** The results compare the inverted structure of the earth’s interior, derived from acoustic data, with the ground truth across different spherical harmonic degrees. The color scale, transitioning from blue to red, indicates negative to positive perturbations respectively. For additional visualizations, please refer to [Appendix I](#).

ing points for each structure in the test set using LHS. Each starting point underwent optimization with the chosen algorithm, and the best result was selected for analysis. As illustrated in [Figure 5](#), performance consistently improves with an increase in the number of starting points across various structural scales. This suggests that using more starting models could further boost inversion efficacy.

In summary, both increasing the number of optimization steps and using more starting models significantly enhance inversion performance. This underscores the ability of ML emulation to effectively tackle the challenges of ill-posedness and local minima—issues that are often problematic in traditional numerical solvers due to the high computational demands of extensive forward modeling. The ML methods allow for a more comprehensive exploration of solution spaces, thus achieving accurate and robust inversion.

Direct inversion mapping An alternative inversion strategy is to directly map observed seismograms to velocity structures. This method bypasses the conventional requirement for a forward ML model, potentially addressing the optimization challenges inherent in traditional FWI. We evaluated this approach by training the InversionNet-3D [10] with a 3D CNN backbone and an InversionMLP with 6 fully connected layers and 1000 hidden units, designed to input acoustic wave seismogram data and output velocity structures. Post-training, these models predicted reasonable velocity structures across various degrees in an unseen test set.

The InversionMLP achieved the best average R of 0.826, MAE of 0.253, and RMSE of 0.335. Visualization of these results is provided in [Figure 6](#). This direct inversion mapping approach offers promising insights into solving ML-based inversion problems more efficiently. The quantitative inversion results for both optimization-based inversion and direct inversion can be found in [Table 3](#). A check-board test in [Appendix I.3](#) further validates the robustness of the direct inversion method.

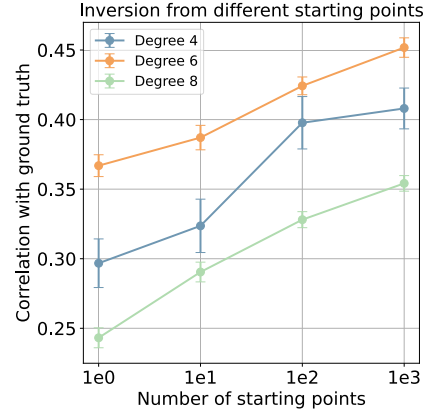


Figure 5: **Inversion with increasing starting points.** Correlation between the inverted models and the ground truth strengthens as the number of starting points increases. Performance on different degrees shows a consistent trend.

Table 3: **Quantitative inversion results.** The OI strategy uses pre-trained forward models. The DI strategy predicts velocity structures directly.

Model	R \uparrow	MAE \downarrow	RMSE \downarrow
MLP+L-BFGS	0.415	0.437	0.528
DeepONet+L-BFGS	0.225	0.482	0.569
H-Fourier+L-BFGS	0.161	0.618	0.750
InversionNet-3D	0.794	0.272	0.356
InversionMLP	0.826	0.253	0.335

3.2.3 Discussion and Future Work

- **Uniqueness of GlobalTomo.** The universality and representativeness of the dataset are critical for ensuring robust performance when applying the trained network to real-world data. In the realm of earth science, we benefit from concentrating on a single, unique object—the earth—from its surface to its core. Unlike datasets from localized explorational sites, which can vary significantly, our dataset is crafted to encapsulate the global attributes of the earth’s interior. This focus enhances the dataset’s applicability as all simulations and models are framed within a consistent setup. Future work could expand by directly incorporating more degrees of spherical harmonics to explore the earth’s interior with greater detail.
- **Broader interest of GlobalTomo.** Beyond application in earth, our dataset is also crucial for advancing planetary sciences, especially for celestial bodies like Mars and the Moon. Seismic data for these planets are rare and costly to obtain. A pre-trained model that simulates seismic activity on these bodies can significantly improve our understanding of their interior.
- **Potential of ML.** ML presents an efficient approach to addressing long-standing scientific challenges. Utilizing a finite set of simulated data that incorporates physics-inspired efficient representations, ML can perform global forward modeling and inversion across various structures and scales with minimal time investment. The accuracy of these methods could be further enhanced by increasing dataset size and adopting more advanced models. Furthermore, ML can be developed to reconcile gaps between synthetic simulation outcomes and real-world data.

4 Conclusion

In this work, we introduce GlobalTomo, a comprehensive 3D global full waveform dataset for physics-ML research in forward modeling and FWI. By synthesizing high-fidelity seismic data across a range of scales and complexities with efficient spherical parameterization, GlobalTomo facilitates the advanced training and evaluation of ML models in the realm of seismic tomography. Our extensive evaluations confirm that ML-based approaches, supported by GlobalTomo, substantially enhance the efficiency and scalability of seismic wavefield modeling and FWI processes. Moreover, this dataset offers new possibilities for integrating machine learning with physics-based modeling. Through this cross-disciplinary endeavor, we anticipate that the convergence of advanced machine learning techniques and traditional geophysical methods will accelerate discoveries in earth science, promoting a deeper understanding of the earth’s interior dynamics and supporting critical applications such as resource discovery and hazard assessment.

Impact statements

A deeper understanding of the earth’s internal structure brings numerous societal benefits, including improved disaster preparedness and mitigation, safer urban planning and infrastructure development, more effective and sustainable management of natural resources, and environmental preservation. This research also fosters global scientific and technological collaboration, further amplifying its positive societal impacts.

Acknowledgments and Disclosure of Funding

The authors would like to thank Miss Chen Zhen (BIGAI) for making the nice figures, Prof. Yizhou Wang (Peking University) for supporting the project, and NVIDIA for their valuable technical support for accelerated computing and artificial intelligence. Z. Li and X. Song are supported in part by the National Natural Science Foundation of China (42394111) and the China National Science and Technology Major Project (2024ZD1001101). S. Li, Z. Mu, S. Xin, and Y. Zhu are supported in part by the National Natural Science Foundation of China (62376009), the State Key Lab of General AI at Peking University, the PKU-BingJi Joint Laboratory for Artificial Intelligence, and the National Comprehensive Experimental Base for Governance of Intelligent Society, Wuhan East Lake High-Tech Development Zone. Z. Li is supported in part by the China Postdoctoral Science Foundation (2025T180102) and the Emerging Engineering Interdisciplinary Young Scholars Project. K. Leng is supported in part by the EPSRC grant, Blueprinting for AI for Science at Exascale (BASE-II, EP/X019918/1).

References

- [1] Adourian, S., Lyu, C., Masson, Y., Munch, F., and Romanowicz, B. (2023). Combining different 3-d global and regional seismic wave propagation solvers towards box tomography in the deep earth. *Geophysical Journal International*, 232(2):1340–1356. [2](#)
- [2] Battaglia, P. W., Hamrick, J. B., and Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences (PNAS)*. [A12](#)
- [3] Bozdağ, E., Peter, D., Lefebvre, M., Komatitsch, D., Tromp, J., Hill, J., Podhorszki, N., and Pugmire, D. (2016). Global adjoint tomography: first-generation model. *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society*, 207(3):1739–1766. [2](#)
- [4] Cai, S., Mao, Z., Wang, Z., Yin, M., and Karniadakis, G. E. (2021). Physics-informed neural networks (pinns) for fluid mechanics: A review. *Acta Mechanica Sinica*, 37(12):1727–1738. [2](#)
- [5] Chen, P., Zhao, L., and 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. [3](#)
- [6] Cottaar, S. and Lekic, V. (2016). Morphology of seismically slow lower-mantle structures. *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society*, 207(2):1122–1136. [4](#)
- [7] Cuomo, S., Di Cola, V. S., Giampaolo, F., Rozza, G., Raissi, M., and Piccialli, F. (2022). Scientific machine learning through physics-informed neural networks: Where we are and what’s next. *Journal of Scientific Computing*, 92(3):88. [2](#)
- [8] Dahlen, F. and Tromp, J. (2020). Theoretical global seismology. In *Theoretical Global Seismology*. Princeton university press. [4](#)
- [9] De Moura, C. A. and Kubrusly, C. S. (2013). The courant–friedrichs–lewy (cfl) condition. *AMC*, 10(12):45–90. [2](#)
- [10] Deng, C., Feng, S., Wang, H., Zhang, X., Jin, P., Feng, Y., Zeng, Q., Chen, Y., and Lin, Y. (2022). Openfwi: Large-scale multi-structural benchmark datasets for full waveform inversion. In *Advances in Neural Information Processing Systems (NeurIPS)*. [3](#), [9](#), [A8](#)
- [11] Dziewonski, A. and Romanowicz, B. (2015). Deep earth seismology: An introduction and overview. *Treatise on Geophysics*, 2:1–28. [1](#)
- [12] Dziewonski, A. M. and Anderson, D. L. (1981). Preliminary reference earth model. *Physics of the earth and planetary interiors*, 25(4):297–356. [A3](#)
- [13] Dziewonski, A. M. and Woodhouse, J. H. (1987). Global images of the earth’s interior. *Science*, 236(4797):37–48. [3](#), [4](#)
- [14] Feng, S., Wang, H., Deng, C., Feng, Y., Liu, Y., Zhu, M., Jin, P., Chen, Y., and Lin, Y. (2024). E-fwi: Multiparameter benchmark datasets for elastic full waveform inversion of geophysical properties. *Advances in Neural Information Processing Systems (NeurIPS)*. [A8](#)
- [15] Feng, S., Wang, H., Deng, C., Feng, Y., Zhu, M., Jin, P., Chen, Y., and Lin, Y. (2023). Openfwi 2.0: Benchmark datasets for elastic full-waveform inversion. In *SEG International Exposition and Annual Meeting*. [3](#)
- [16] Fichtner, A. (2010). *Full seismic waveform modelling and inversion*. Springer Science & Business Media. [1](#)
- [17] Fichtner, A., Igel, H., Bunge, H.-P., and Kennett, B. L. (2009). Simulation and inversion of seismic wave propagation on continental scales based on a spectral-element method. *Journal of Numerical Analysis, Industrial and Applied Mathematics*, 4(1-2):11–22. [3](#)
- [18] French, S. and Romanowicz, B. A. (2014). Whole-mantle radially anisotropic shear velocity structure from spectral-element waveform tomography. *Geophysical Journal International*, 199(3):1303–1327. [3](#), [4](#)
- [19] French, S. W. and Romanowicz, B. (2015). Broad plumes rooted at the base of the earth’s mantle beneath major hotspots. *Nature*, 525(7567):95–99. [3](#)
- [20] Górszczyk, A., Operto, S., and Malinowski, M. (2017). Toward a robust workflow for deep crustal imaging by fwi of obs data: The eastern nankai trough revisited. *Journal of Geophysical Research: Solid Earth*, 122(6):4601–4630. [3](#)

- [21] Hao, Z., Wang, Z., Su, H., Ying, C., Dong, Y., Liu, S., Cheng, Z., Song, J., and Zhu, J. (2023). Gnot: A general neural operator transformer for operator learning. In *International Conference on Machine Learning (ICML)*. 6
- [22] Hennigh, O., Narasimhan, S., Nabian, M. A., Subramaniam, A., Tangsali, K., Fang, Z., Rietmann, M., Byeon, W., and Choudhry, S. (2021). Nvidia simnet™: An ai-accelerated multi-physics simulation framework. In *International conference on computational science*. 5
- [23] Kelly, K. R., Ward, R. W., Treitel, S., and Alford, R. M. (1976). Synthetic seismograms: A finite-difference approach. *Geophysics*, 41(1):2–27. 2
- [24] Komatitsch, D. and Tromp, J. (2002a). Spectral-element simulations of global seismic wave propagation—i. validation. *Geophysical Journal International*, 149(2):390–412. 2
- [25] Komatitsch, D. and Tromp, J. (2002b). Spectral-element simulations of global seismic wave propagation—ii. three-dimensional models, oceans, rotation and self-gravitation. *Geophysical Journal International*, 150(1):303–318. 2
- [26] Kovachki, N., Li, Z., Liu, B., Azizzadenesheli, K., Bhattacharya, K., Stuart, A., and Anandkumar, A. (2023). Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning Research (JMLR)*, 24(89):1–97. 3
- [27] Krischer, L., Afanasiev, M., Boehm, C., Driel, M. v., and Fichtner, A. (2021). Salvus: A flexible open-source package for waveform modelling and inversion from laboratory to global scales. 2
- [28] Kubricht, J. R., Holyoak, K. J., and Lu, H. (2017). Intuitive physics: Current research and controversies. *Trends in cognitive sciences*. A11
- [29] Lei, W., Ruan, Y., Bozdağ, E., Peter, D., Lefebvre, M., Komatitsch, D., Tromp, J., Hill, J., Podhorszki, N., and Pugmire, D. (2020). Global adjoint tomography—model glad-m25. *Geophysical Journal International*, 223(1):1–21. 2
- [30] Leng, K., Korenaga, J., and Nissen-Meyer, T. (2020). 3-d scattering of elastic waves by small-scale heterogeneities in the earth’s mantle. *Geophysical Journal International*, 223(1):502–525. 5
- [31] Leng, K., Nissen-Meyer, T., and van Driel, M. (2016). Efficient global wave propagation adapted to 3-d structural complexity: a pseudospectral/spectral-element approach. *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society*, 207(3):1700–1721. 4
- [32] Leng, K., Nissen-Meyer, T., van Driel, M., Hosseini, K., and Al-Attar, D. (2019). Axisem3d: broadband seismic wavefields in 3-d global earth models with undulating discontinuities. *Geophysical Journal International*, 217(3):2125–2146. 4
- [33] Li, S., Wu, K., Zhang, C., and Zhu, Y. (2022a). On the learning mechanisms in physical reasoning. *Advances in Neural Information Processing Systems (NeurIPS)*. A12
- [34] Li, S., Wu, K., Zhang, C., and Zhu, Y. (2023). I-phyre: Interactive physical reasoning. In *International Conference on Learning Representations (ICLR)*. A12
- [35] Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., and Anandkumar, A. (2020a). Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*. 2, 3
- [36] Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Stuart, A., Bhattacharya, K., and Anandkumar, A. (2020b). Multipole graph neural operator for parametric partial differential equations. *Advances in Neural Information Processing Systems*, 33:6755–6766. 3
- [37] Li, Z., Leng, K., Jenkins, J., and Cottaar, S. (2022b). Kilometer-scale structure on the core–mantle boundary near hawaii. *Nature Communications*, 13(1):2787. 2
- [38] Liu, M., Liang, L., and Sun, W. (2020). A generic physics-informed neural network-based constitutive model for soft biological tissues. *Computer methods in applied mechanics and engineering*, 372:113402. 2
- [39] Lu, L., Jin, P., Pang, G., Zhang, Z., and Karniadakis, G. E. (2021). Learning nonlinear operators via deep-onet based on the universal approximation theorem of operators. *Nature machine intelligence*, 3(3):218–229. 2, 3, 6
- [40] Lu, L., Meng, X., Cai, S., Mao, Z., Goswami, S., Zhang, Z., and Karniadakis, G. E. (2022). A comprehensive and fair comparison of two neural operators (with practical extensions) based on fair data. *Computer Methods in Applied Mechanics and Engineering*, 393:114778. 3, A9

- [41] Lyu, C., Capdeville, Y., Al-Attar, D., and Zhao, L. (2021). Intrinsic non-uniqueness of the acoustic full waveform inverse problem. *Geophysical Journal International*, 226(2):795–802. 1
- [42] Meschede, M. and Romanowicz, B. (2015). Lateral heterogeneity scales in regional and global upper mantle shear velocity models. *Geophysical Journal International*, 200(2):1078–1095. 4
- [43] Moczo, P., Kristek, J., Galis, M., Pazak, P., and Balazovjeh, M. (2007). The finite-difference and finite-element modeling of seismic wave propagation and earthquake motion. *Acta physica slovacica*, 57(2). 2
- [44] Mousavi, S. M. and Beroza, G. C. (2022). Deep-learning seismology. *Science*, 377(6607):eabm4470. 3
- [45] Nissen-Meyer, T., van Driel, M., Stähler, S. C., Hosseini, K., Hempel, S., Auer, L., Colombi, A., and Fournier, A. (2014). Axisem: broadband 3-d seismic wavefields in axisymmetric media. *Solid Earth*, 5(1):425–445. 2
- [46] O’Gorman, P. A. and Dwyer, J. G. (2018). Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change, and extreme events. *Journal of Advances in Modeling Earth Systems*, 10(10):2548–2563. 2
- [47] Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., Kurth, T., Hall, D., Li, Z., Azizzadenesheli, K., et al. (2022). Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214*. 2
- [48] Plessix, R.-E. (2006). A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167(2):495–503. 2
- [49] Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., Bengio, Y., and Courville, A. (2019). On the spectral bias of neural networks. In *International Conference on Learning Representations (ICLR)*. 6
- [50] Rao, J., Ratsasapp, M., and Fan, Z. (2016). Guided wave tomography based on full waveform inversion. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 63(5):737–745. 3
- [51] Rasht-Behesht, M., Huber, C., Shukla, K., and 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. 3, A11
- [52] Rawlinson, N., Fichtner, A., Sambridge, M., and Young, M. K. (2014). Seismic tomography and the assessment of uncertainty. *Advances in geophysics*, 55:1–76. A28
- [53] Ritsema, J., Deuss, A., Van Heijst, H., and Woodhouse, J. (2011). S40rts: a degree-40 shear-velocity model for the mantle from new rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements. *Geophysical Journal International*, 184(3):1223–1236. 4
- [54] Ritsema, J. and Lekić, V. (2020). Heterogeneity of seismic wave velocity in earth’s mantle. *Annual Review of Earth and Planetary Sciences*, 48:377–401. 4
- [55] Romanowicz, B. (2003). Global mantle tomography: progress status in the past 10 years. *Annual Review of Earth and Planetary Sciences*, 31(1):303–328. 1
- [56] Schreiman, J., Gisvold, J., Greenleaf, J. F., and Bahn, R. (1984). Ultrasound transmission computed tomography of the breast. *Radiology*, 150(2):523–530. 3
- [57] Song, C. and Alkhalifah, T. A. (2021). Wavefield reconstruction inversion via physics-informed neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–12. 3
- [58] Srivastava, R. K., Greff, K., and Schmidhuber, J. (2015). Training very deep networks. In *Advances in Neural Information Processing Systems (NeurIPS)*. 6
- [59] Su, W.-j. and Dziewonski, A. M. (1991). Predominance of long-wavelength heterogeneity in the mantle. *Nature*, 352(6331):121–126. 4
- [60] Sun, D., Miller, M. S., Agostinetti, N. P., Asimow, P. D., and Li, D. (2014). High frequency seismic waves and slab structures beneath italy. *Earth and Planetary Science Letters*, 391:212–223. 4
- [61] Tancik, M., Srinivasan, P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J., and Ng, R. (2020). Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in Neural Information Processing Systems (NeurIPS)*. 6

- [62] Tape, C., Liu, Q., Maggi, A., and Tromp, J. (2009). Adjoint tomography of the southern california crust. *Science*, 325(5943):988–992. [3](#)
- [63] Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(8):1259–1266. [3](#), [A12](#)
- [64] Thrastarson, S., van Herwaarden, D.-P., Noe, S., Schiller, C., and Fichtner, A. (2024). Reveal: A global full-waveform inversion model. *Bull. Seismol. Soc. Am.* [4](#), [5](#)
- [65] Tromp, J. (2020). Seismic wavefield imaging of earth’s interior across scales. *Nature Reviews Earth & Environment*, 1(1):40–53. [1](#), [3](#), [A11](#), [A12](#), [A28](#)
- [66] Tromp, J., Tape, C., and Liu, Q. (2005). Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophysical Journal International*, 160(1):195–216. [3](#), [A12](#)
- [67] van der Hilst, R. D., Widiyantoro, S., and Engdahl, E. (1997). Evidence for deep mantle circulation from global tomography. *Nature*, 386(6625):578–584. [4](#)
- [68] Vasudevan, R., Pilia, G., and Balachandran, P. V. (2021). Machine learning for materials design and discovery. *Journal of Applied Physics*, 129(7). [2](#)
- [69] Virieux, J. and Operto, S. (2009). An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26. [1](#)
- [70] Wang, S., Wang, H., and Perdikaris, P. (2021). Learning the solution operator of parametric partial differential equations with physics-informed deepnets. *Science advances*, 7(40):eabi8605. [2](#), [3](#), [6](#)
- [71] Wiskin, J., Malik, B., Natesan, R., Pirshafiey, N., Lenox, M., and Klock, J. (2019). Full wave 3d inverse scattering: 21st century technology for whole body imaging. *The Journal of the Acoustical Society of America*, 145(3_Supplement):1857–1857. [3](#)
- [72] Woollam, J., Münchmeyer, J., Tilmann, F., Rietbrock, A., Lange, D., Bornstein, T., Diehl, T., Giunchi, C., Haslinger, F., Jozinović, D., et al. (2022). Seisbench—a toolbox for machine learning in seismology. *Seismological Society of America*, 93(3):1695–1709. [A8](#)
- [73] Yang, L., Meng, X., and Karniadakis, G. E. (2021). B-pinns: Bayesian physics-informed neural networks for forward and inverse pde problems with noisy data. *Journal of Computational Physics*, 425:109913. [3](#)
- [74] Yang, Y., Gao, A. F., Azizzadenesheli, K., Clayton, R. W., and Ross, Z. E. (2023). Rapid seismic waveform modeling and inversion with neural operators. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–12. [A11](#), [A12](#)
- [75] Zang, N., Zhang, W., and Chen, X. (2021). An overset-grid finite-difference algorithm for simulating elastic wave propagation in media with complex free-surface topography. *Geophysics*, 86(4):T277–T292. [1](#)
- [76] Zhao, M., Xiao, Z., Chen, S., and Fang, L. (2023). Diting: A large-scale chinese seismic benchmark dataset for artificial intelligence in seismology. *Earthquake Science*, 36(2):84–94. [A8](#)
- [77] Zhu, H., Bozdağ, E., Peter, D., and Tromp, J. (2012). Structure of the european upper mantle revealed by adjoint tomography. *Nature Geoscience*, 5(7):493–498. [3](#)
- [78] Zienkiewicz, O. C., Taylor, R. L., and Zhu, J. Z. (2005). *The finite element method: its basis and fundamentals*. Elsevier. [2](#)

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [\[Yes\]](#)

Justification: The abstract and introduction have stated the two problems in geophysics and the advantages and contributions of using ML approaches to tackle the challenges.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [\[Yes\]](#)

Justification: See [Section 3.2.3](#).

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [\[NA\]](#)

Justification: [NA]

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: See [Appendix F](#).

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
 - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: See [Appendix A](#).

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: See [Appendix F](#).

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: See [Section 3.2](#).

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer “Yes” if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: See [Appendix F](#).

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines?>

Answer: [Yes]

Justification: We adhere to the NeurIPS Code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: See [Section 3.2.3](#).

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to

generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.

- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: [NA]

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: We explicitly mentioned the license of AxiSEM3D in [Appendix C](#).

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. New assets

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: We provided a dataset on Hugging Face with data structure, license, and code.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. Crowdsourcing and research with human subjects

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: [NA]

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. Institutional review board (IRB) approvals or equivalent for research with human subjects

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: [NA]

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. Declaration of LLM usage

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [NA]

Justification: [NA]

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.

Contents of Supplementary Material

A	Data and Code Access	A1
B	Physical Representations	A1
B.1	Physical Constraints	A1
B.2	Expansion to Azimuthal Fourier Domain	A1
B.3	Weak Formulation and Dimension-Reduced Form	A2
C	Earth Configuration Details	A2
C.1	Concept Illustration	A2
C.2	Three Dataset Tiers	A3
C.3	Data Distribution	A4
D	Forward and Inversion Pipelines	A8
E	Comparison with Other Related Datasets	A8
F	Baseline Model Details	A8
F.1	Evaluation Metrics	A8
F.2	Efficient DeepONet Design	A9
F.3	Hyperparameters	A10
G	Discussion	A10
G.1	Comparison of Modern ML-based Inversion	A10
G.2	Is Predicting Wavefield Necessary for Inversion?	A11
G.3	Connection with Other Fields	A11
H	Datasheet	A12
H.1	Motivation	A12
H.2	Composition	A12
H.3	Collection Process	A13
H.4	Preprocessing/cleaning/labeling	A13
H.5	Uses	A13
H.6	Distribution	A13
H.7	Maintenance	A14
I	More Visualization	A14
I.1	Forward Modeling	A14
I.2	Inversion	A17
I.3	A Checkboard Test	A28

A Data and Code Access

We have released our dataset on our website <https://global-tomo.github.io/>. The data can also be directly accessed in Hugging Face <https://huggingface.co/datasets/lishiqianhugh/globaltomo/>. The code for data loading, training, and evaluation has been documented and published on <https://github.com/lishiqianhugh/GlobalTomo/>. We will continuously update and maintain both the code and data to ensure they reflect the latest advancements.

B Physical Representations

B.1 Physical Constraints

We describe the physical constraints of the seismic wave propagation that can be used in the training process of neural networks. For the acoustic dataset tier, the wave propagation follows the acoustic wave equation in the fluid domain Ω_f :

$$\frac{\partial_t^2 \chi}{\kappa} = \nabla \cdot \left(\frac{\nabla \chi}{\rho} \right), \quad \text{in } \Omega_f. \quad (\text{A1})$$

The output displacement $\mathbf{u} = \nabla \chi / \rho$, χ denotes the wave potential, ρ represents the fluid density, and κ represents the bulk modulus. The acoustic dataset has a pressure-free boundary condition that can be stated by:

$$\chi = 0, \quad \text{on } \partial\Omega_f. \quad (\text{A2})$$

The elastic wave propagation in the solid domain Ω_s follows:

$$\rho \partial_t^2 \mathbf{u} - \nabla \cdot (\mathbf{C} : \nabla \mathbf{u}) = 0, \quad \text{in } \Omega_s. \quad (\text{A3})$$

where \mathbf{C} represents the elasticity tensor. The elastic dataset has a traction-free surface boundary condition:

$$\hat{\mathbf{n}} \cdot (\mathbf{C} : \nabla \mathbf{u}) = 0, \quad \text{on } \partial\Omega_s. \quad (\text{A4})$$

The Real Earth tier is a combination of these two scenarios. The seismic wave propagation in the fluid outer core follows Equation (A1) while in the resting solid places follows Equation (A3). This tier still has a traction-free boundary condition same as Equation (A4). The solid-fluid coupling is controlled by the displacement- and stress-continuity condition on the fluid-solid interface Σ :

$$\hat{\mathbf{n}} \cdot \mathbf{u} = \chi_r / \rho, \quad \hat{\mathbf{n}} \cdot (\mathbf{C} : \nabla \mathbf{u}) = \partial_t^2 \chi \hat{\mathbf{n}} \quad \text{on } \Sigma. \quad (\text{A5})$$

B.2 Expansion to Azimuthal Fourier Domain

Describing the 3D earth is memory-consuming. Inspired by the smoothness and symmetry found in the physical properties of the earth, representing the wavefield in the Fourier domain can achieve a more efficient description. If the time-dependent equations of motion can be generalized in computational domain Ω as:

$$\mathcal{L} \mathbf{u} = \mathbf{f} \quad (\text{A6})$$

The displacement \mathbf{u} , operator \mathcal{L} , and external forces \mathbf{f} will be characterized to the required azimuth resolution defined by n_u :

$$\mathbf{u}(r, \theta, \phi; t) = \sum_{|\alpha| \leq n_u} \mathbf{u}^\alpha(r, \theta; t) \exp(i\alpha\phi) \quad (\text{A7})$$

$$\mathcal{L}(r, \theta, \phi; t) = \sum_{|\beta| \leq n_u} \mathcal{L}^\beta(r, \theta; t) \exp(i\beta\phi) \quad (\text{A8})$$

$$\mathbf{f}(r, \theta, \phi; t) = \sum_{|\gamma| \leq n_u} \mathbf{f}^\gamma(r, \theta; t) \exp(i\gamma\phi) \quad (\text{A9})$$

Thus this method reduces the dimension of azimuth that is needed to solve a true 3D problem and saves huge computation costs while also ensuring the accuracy of the solution. If we apply the

structural parameterization of spherical harmonics and determine the highest degree of the model, AxiSEM3D currently has the best theoretical efficiency for calculating their global wavefields, which only need $2n_u$ times the computation of a 2D problem. In GlobalTomo, we provide the Fourier series of the wavefield for building more robust forward models with efficient representation in the future.

B.3 Weak Formulation and Dimension-Reduced Form

The weak formulation of the elastic equation can be obtained by taking dot product with an arbitrary test function \mathbf{w} and integrating by parts over the volume Ω_s :

$$\int_{\Omega_s} \left(\rho \partial_t^2 \mathbf{u} \cdot \mathbf{w} + \nabla \mathbf{u} : \mathbf{C} : \nabla \mathbf{w} \right) d\mathbf{x}^3 = \int_{\Omega_s} \mathbf{f} \cdot \mathbf{w} d\mathbf{x}^3 \quad (\text{A10})$$

The stress-free boundary condition is thus naturally merged into the weak formulation of the constraint.

Its corresponding dimension-reduced form in a 2D domain D can be written down as:

$$\sum_{|\alpha| \leq n_u} [\langle \rho^{-(\alpha+\beta)} \partial_t^2 \mathbf{u}^\alpha, \mathbf{w}^\beta \rangle_D + a_D^{-(\alpha+\beta)}(\mathbf{u}^\alpha, \mathbf{w}^\beta)] = \langle \mathbf{f}^{-\beta}, \mathbf{w}^\beta \rangle_D \quad (\text{A11})$$

$\forall \mathbf{w}^\beta \text{ in } D, \quad \forall \beta \in \{-n_u, \dots, n_u\}$

$$\langle \mathbf{u}^\alpha, \mathbf{w}^\beta \rangle_D := \int_D u_i^\alpha w_i^\beta s ds dz \quad (\text{A12})$$

$$a_D^\gamma(\mathbf{u}^\alpha, \mathbf{w}^\beta) := \int_D u_{i,j}^\alpha C_{ijkl}^\gamma w_{k,l}^\beta s ds dz \quad (\text{A13})$$

where the azimuthal expansion of \mathbf{u} , \mathbf{f} , ρ , and C follows:

$$\mathbf{u} = \sum_{|\alpha| \leq n_u} \mathbf{u}^\alpha(s, z; t) \Psi^\alpha(\phi) \quad (\text{A14})$$

$$\mathbf{f} = \sum_{|\alpha| \leq n_u} \mathbf{f}^\alpha(s, z; t) \Psi^\alpha(\phi) \quad (\text{A15})$$

$$\rho = \sum_{|\alpha| \leq n_u} \rho^\alpha(s, z; t) \Psi^\alpha(\phi) \quad (\text{A16})$$

$$C_{ijkl} = \sum_{|\alpha| \leq n_u} C_{ijkl}^\alpha(s, z; t) \Psi^\alpha(\phi) \quad (\text{A17})$$

Here, $\Psi^\alpha(\phi)$, $\alpha \in \{-n_u, \dots, n_u\}$ denote the azimuthal Fourier modes,

$$\Psi^\alpha(\phi) := \exp(i\alpha\phi), i = \sqrt{-1}. \quad (\text{A18})$$

C Earth Configuration Details

C.1 Concept Illustration

- **Source:** Described by location and moment tensor, a source of seismic energy can be either natural or artificial.
- **Velocity Structure:** Refers to the distribution of seismic wave velocities within the earth's interior, determined by the physical and chemical properties of the materials.
- **Wavefield:** The spatial pattern of wave propagation within the earth.
- **Seismogram:** The recordings of ground motions received by stations on the surface.
- **Forward Modeling:** The process of computing numerical solutions of the wave propagation from a defined model of the earth's interior.
- **Full Waveform Inversion (FWI):** An advanced seismic inversion technique that leverages the entire seismic waveform signals to develop high-resolution models of the earth's interior.

C.2 Three Dataset Tiers

We introduce the details of the three tiers generated by AxiSEM3D. The simulator is licensed under the MIT license at <https://github.com/AxiSEMunity/AxiSEM3D>.

Our dataset includes three tiers of synthetic seismic data. The first tier models acoustic wave propagation through a fluid sphere with a 1-km radius using a 20 Hz mesh. The second tier models elastic wave propagation through an isotropic solid medium of the same scale, with variations in the source. The third tier is designed for real-world earth applications, integrating both acoustic and elastic simulations over a 30-second period.

We export the 3D velocity perturbations in Cartesian coordinates to facilitate rapid interpolation and retain the original spherical harmonics coefficients for use in network training. To ensure the adherence to physical constraints, we disable attenuation.

Acoustic Ball The acoustic ball model is designed on a 1000 m radius ball with pure fluid medium and a mesh accurate to 20 Hz. This set of models has a pressure-free boundary condition. The wave propagation in the model obeys the acoustic wave equation. The medium only has the P-wave velocity (vp) attribute. The 1D background model has a uniform vp value of 1.0 km/s across the ball. The 3D superimposed velocity perturbation is parameterized by spherical harmonics up to degree 8, leading to 81 spherical parameters per layer, multiplied by 5 layers, resulting in 405 free parameters in total.

The source is located at a fixed 200 m depth at the north pole with a fixed fluid pressure source.

Elastic Ball The elastic ball model is also designed on a 1000 m radius ball with pure elastic medium and a mesh accurate to 20 Hz. This set of models has a stress-free boundary condition. The wave propagation in the model obeys the elastic wave equation. The isotropic medium only has three independent attributes: P-wave velocity (vp), S-wave velocity (vs), and density (rho). The 1D background model has a uniform vp value of 1.5 km/s, vs value of 1.0 km/s, and rho of 1.0g/cm³ across the ball. The 3D superimposed velocity perturbation is parameterized by spherical harmonics up to degree 8, leading to 81 spherical parameters per layer, multiplied by 5 layers and 3 medium attributes, resulting in 1215 free parameters in total.

The source is located at a fixed 200 m depth at the north pole with a variable moment tensor source mechanism. This introduces 6 extra free parameters in source terms.

Real Earth This set of models is designed for the ultimate real-earth applications. We utilize the isotropic PREM [12] background model and a global mesh with a period accurate to 30 seconds that encompasses various earth's layers including the crust, mantle, outer core, and inner core, each characterized by their unique seismic attributes. The Real Earth tier has a stress-free boundary condition. The 3D velocity perturbations are imposed on the earth's internal 23 discontinuities, resulting in a total of 5427 free parameters up to degree 8 across the whole earth.

The source is located at stochastic depths between 0 - 800 km at arbitrary geographical coordinates with a variable moment tensor source mechanism. This introduces 9 extra free parameters in source terms.

Surface Seismogram Output The seismogram data contains a one-dimensional time series recording ground displacement at each surface station. Each trace is low-pass filtered to 30 seconds to avoid numerical noise. We place such stations evenly along latitude and longitude to represent a spatially abundant coverage. The actual station deployments limit the real seismic data coverage, however, this issue can be considered by dumping part of the data during the application. The seismogram data can be observed globally and be used in the inversion of the earth's internal structure.

Seismic Wavefield Output The seismic wavefields are saved in discrete 15 or 20 time snapshots, (total time = 3 s or 6,000 s), represented by 16 azimuthal slices with 3,648 or 16,470 (from surface to the core) element points in each. Besides, we also saved the azimuthal Fourier coefficients of the global wavefields for a continuous and compact representation in three dimensions. The wavefields and the seismograms are saved as HDF5 files for fast loading.

C.3 Data Distribution

We visualize the data distribution of three tiers in [Figure A1](#), [Figure A2](#), and [Figure A3](#), respectively. The spherical harmonics coefficients are sampled by LHS, which is finally used to construct velocity perturbations. The exact representation of the spherical harmonics basis function is constructed by:

$$Y_{lm}(\theta, \phi) = \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos \theta) e^{im\phi} \quad (\text{A19})$$

where P_l^m are associated Legendre polynomials defined by

$$P_l^m(x) = (-1)^m (1-x^2)^{m/2} \frac{d^m}{dx^m} P_l(x). \quad (\text{A20})$$

We realized this function in SciPy module. The velocity structure is given by:

$$F(\theta, \phi) = \sum_{l=0}^8 \sum_{m=-l}^l LHS_{coeff} \cdot Y_{lm}(\theta, \phi) \quad (\text{A21})$$

Thus, the variability of the model is guaranteed by the LHS coefficient of spherical harmonics. The final constructed velocity perturbation can also be viewed as a linear transformation based upon the spherical basis function.

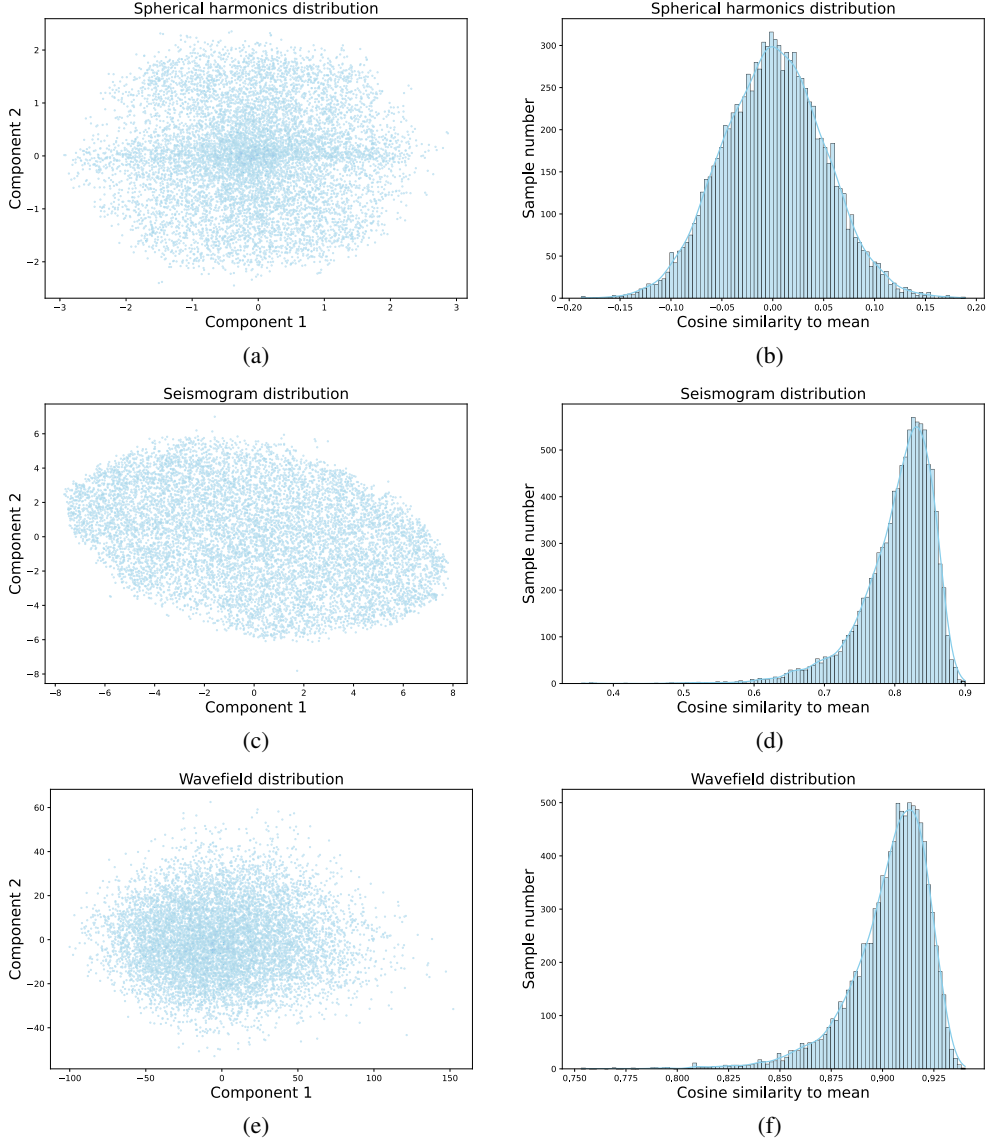


Figure A1: Data distribution of the Acoustic tier. We visualize the data distribution using t-SNE (t-distributed Stochastic Neighbor Embedding) in subfigures (a), (c), and (e), where the noise-like shape indicates the variability in our dataset. Additionally, we present the distribution of Cosine similarity to the mean in subfigures (b), (d), and (f) to illustrate how the distribution of input spherical harmonics shifts towards those of the output seismogram and wavefield. The low similarity of spherical harmonics, sampled from -0.025 to 0.025 in the range of interest, further demonstrates the variability of velocity structures. Note that the high average similarity in seismogram and wavefield data stems from the intrinsic simplicity of the acoustics waveform and doesn't indicate the input structures are not diverse.

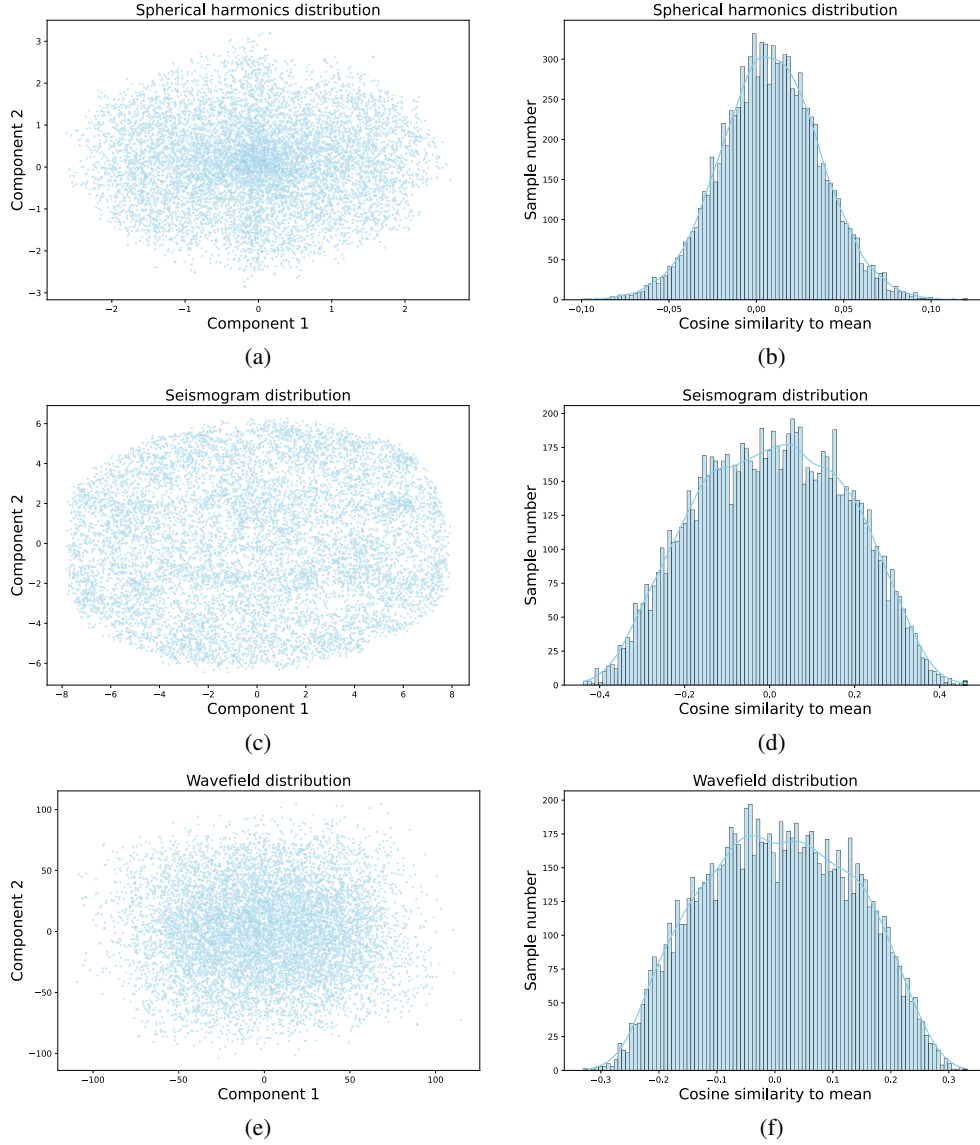


Figure A2: Data distribution of the Elastic tier. We visualize the data distribution using t-SNE in subfigures (a), (c), and (e), where the noise-like shape indicates the variability in our dataset. Additionally, we present the distribution of Cosine similarity to the mean in subfigures (b), (d), and (f) to illustrate how the distribution of input spherical harmonics shifts towards those of the output seismogram and wavefield. The low average similarity in seismogram and wavefield data stems from the intrinsic complexity of the elastic waveform and its sensitivity to different velocity structures.

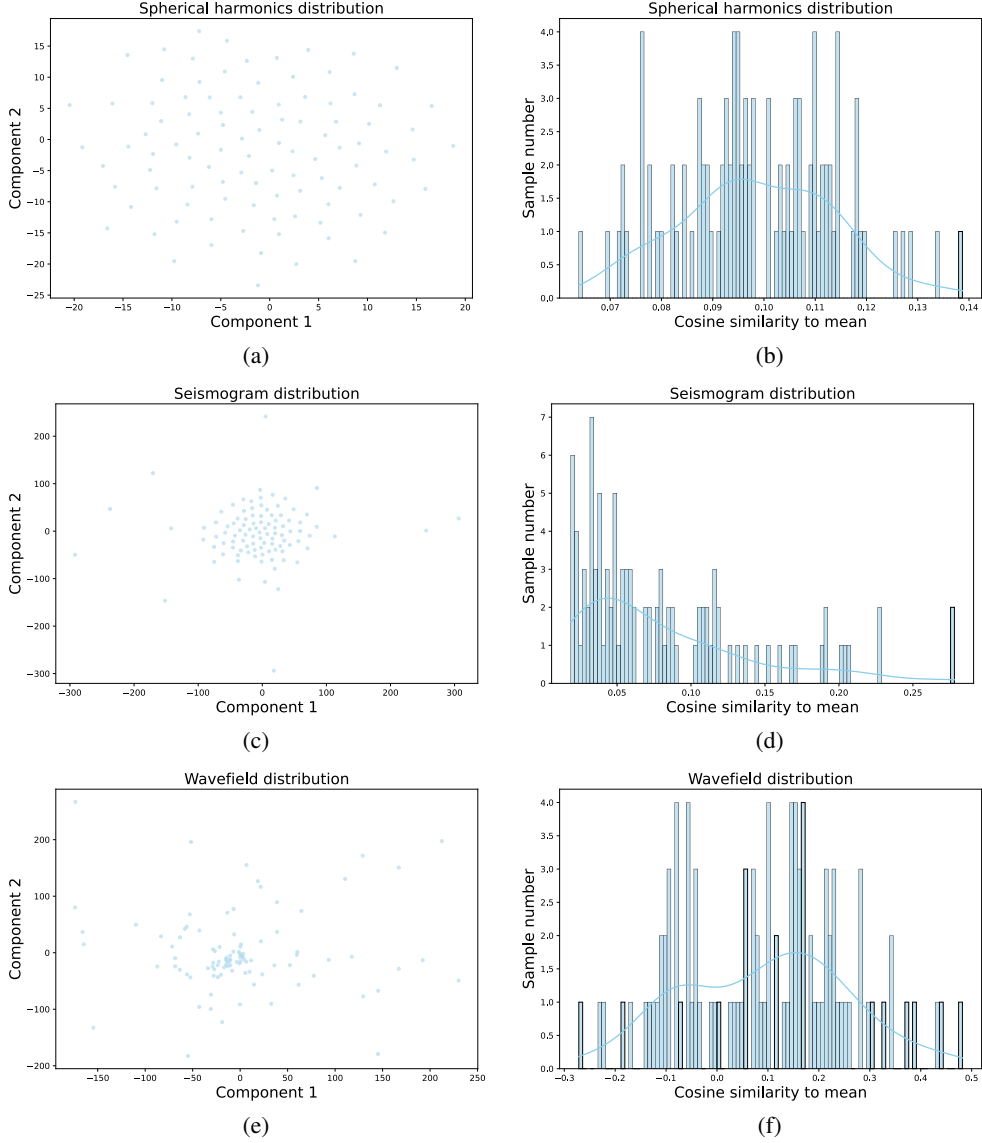


Figure A3: **Data distribution of the Earth tier (only 100 samples are shown due to memory cost).** We visualize the data distribution using t-SNE in subfigures (a), (c), and (e), where the noise-like shape indicates the variability in our dataset. Additionally, we present the distribution of Cosine similarity to the mean in subfigures (b), (d), and (f) to illustrate how the distribution of input spherical harmonics shifts towards those of the output seismogram and wavefield.

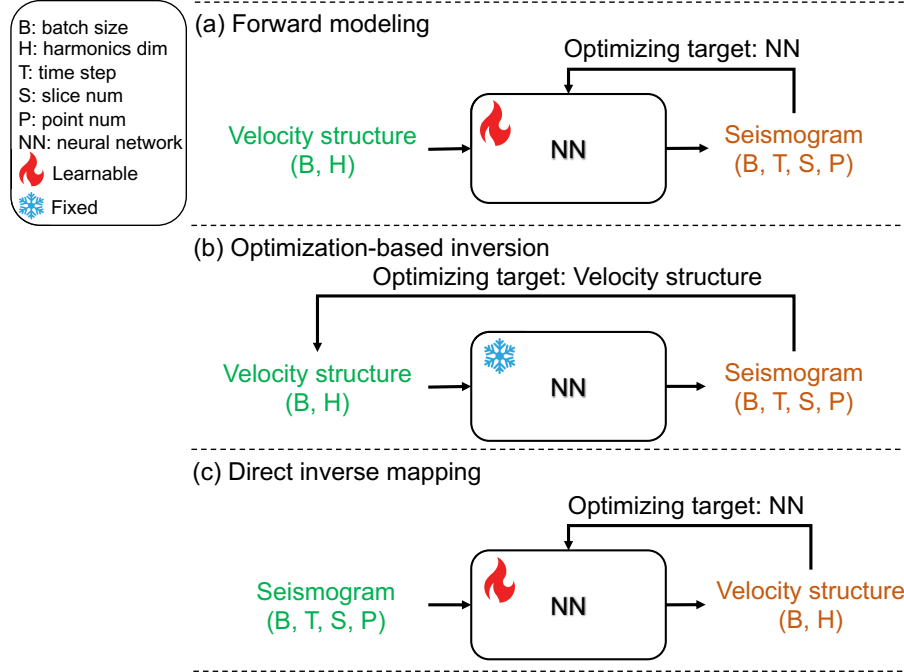


Figure A4: **The pipeline for forward modeling, optimization-based inversion, and direct inverse mapping.** (a) In forward modeling, the neural network (NN) is trained by using velocity structures as input and generating the corresponding seismogram as output. Once trained, the model parameters are fixed. (b) In the optimization-based inversion, the velocity structure is adjusted to match the observed seismogram. (c) In direct inverse mapping, a neural network is trained to directly predict the velocity structure from the observed seismogram input.

D Forward and Inversion Pipelines

We visualize the pipelines of forward modeling, optimization-based inversion, and direct inverse mapping in our work in [Figure A4](#). Specifically, the input, output, and optimizing target are highlighted in different processes.

E Comparison with Other Related Datasets

We contrast the GlobalTomo dataset with prior works across several key aspects in [Table A1](#). The comparison highlights the differences in scale, data utilization, parameterization, and simulation methods. GlobalTomo is first designed for global-scale inversion, incorporating the entire wave-field and receiver data, utilizing spherical harmonics for structural parameterization, and employing spectral element methods (SEM) for efficient simulation. In contrast, prior works (e.g., OpenFWI [10], E-FWI [14]) typically focus on local-scale applications, use only receiver data, rely on mesh representation, and often employ finite difference methods (FDM) for simulation. DiTing [76] and SeisBench [72] are particularly designed for phase picking, denoising, and earthquake detection instead of earth tomography.

F Baseline Model Details

F.1 Evaluation Metrics

The RL2 metric is particularly valuable in scenarios where it's crucial to evaluate the performance of predictive models across datasets of varying scales or inherent variability. By normalizing the Mean Squared Error (MSE) with the variance of the actual values, this metric effectively renders the error measurement scale-independent. The formulation is as follows:

Table A1: Comparison with prior related datasets.

Category	Perspective	OpenFWI	E-FWI	DiTing	SeisBench	GlobalTomo
Scale	Global				✓	✓
	Local	✓	✓	✓	✓	
Wave type	Acoustic	✓				✓
	Elastic		✓	✓	✓	✓
Data utilization	Wavefield					✓
	Receiver	✓	✓	✓	✓	✓
Structural parameterization	Sph. harmonics					✓
	Mesh	✓	✓			
Simulation method	FDM	✓	✓			
	SEM					✓

$$\text{Relative } L_2 \text{ Loss} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (\text{A22})$$

where y_i and \hat{y}_i denote the actual and predicted values of the i -th observation, respectively. \bar{y} is the mean of all actual values in the dataset, calculated as $\frac{1}{n} \sum_{i=1}^n y_i$.

We also use the Person Correlation Coefficient (R) to measure the similarity between the prediction and ground truth. The R is given by:

$$R = \frac{\sum (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum (y_i - \bar{y})^2 \sum (\hat{y}_i - \bar{\hat{y}})^2}} \quad (\text{A23})$$

F.2 Efficient DeepONet Design

The original DeepONet takes paired velocity structure and spatial-temporal points for the trunk network and branch network. The output of DeepONet is calculated by combining the outputs from the branch and trunk networks, typically through a multiplication operation. Specifically, if the output from the trunk network is $a = [a_1, a_2, \dots, a_d]$ and the output from the branch network is $b = [b_1, b_2, \dots, b_d]$, then the predicted value $\hat{\phi}(p, t)$ at output position p and time t is given by:

$$\hat{\phi}(p, t) = \sum_{i=1}^d a_i \cdot b_i(p, t) \quad (\text{A24})$$

Here, a_i and $b_i(p, t)$ are the features obtained from the branch and trunk networks, respectively, and their product followed by summation forms the estimate of the output function.

During training, this network is optimized to minimize the error between the true output $\phi(p, t)$ and the predicted output $\hat{\phi}(p, t)$, typically using the mean squared error as the loss function:

$$\text{Loss} = \frac{1}{n} \sum_{j=1}^n \left(\phi(p_j, t_j) - \hat{\phi}(p_j, t_j) \right)^2 \quad (\text{A25})$$

where n is the number of output points.

However, this is memory-consuming since the velocity structures are duplicated for each point. Indeed, one velocity structure can be paired with a group of spatial and temporal points P, T simultaneously. Inspired by [40], we employ a more efficient DeepONet that can predict the whole 3D outputs just using one forward of branch net. The calculation of the output function now is given by:

$$\hat{\Phi}(P, T) = \sum_{i=1}^d a_i \cdot B_i(P, T) \quad (\text{A26})$$

where $B_i(P, T)$ indicates a vector of the i -th latent feature from a batch of points in the same velocity structure. $\hat{\Phi}(P, T)$ indicates a group of predicted output in the same structure. Suppose the input dimensions of the velocity structure and the spatial-temporal point are m and 4, respectively, the memory usage in the modified DeepONet can be $\frac{4+m}{4+m/n}$ times smaller than the original one. The training speed is about 3 times faster.

F.3 Hyperparameters

For reproducibility, we present the training configuration of baselines in Table A2 and Table A3. We employ the SiLu activation function and use the Adam optimizer for parameter tuning. The initial learning rate is set at 0.0003, and it decreases exponentially by a factor of 0.95 every 1000 steps. Both the MLP and H-Fourier models consist of six layers with 500 neurons each. The DeepONet model features six layers of 600 neurons each, and includes trunk and branch projection layers, each with 1000 neurons. The final output is computed via a dot product of the latent trunk and branch vectors. The GNOT model has 4 attention layers with a hidden size of 256.

The MLP and H-Fourier models are trained on a single A800 SXM4 80GB GPU. In contrast, DeepONet and GNOT require 8 A800 SXM4 80GB GPUs for training due to the substantial increase in data size associated with point-based prediction.

Table A2: **Wavefield model details.**

Wave type	Model	Layer num	Hidden size	Training steps	Input size	Output size
Acoustic	MLP	6	500	10000	405	408576
	H-Fourier	6	500	3000	405	408576
	DeepONet	6	600	100000	405+3	1
	GNOT	4	256	200000	405+3	1
Elastic	MLP	6	500	30000	1221	408576
	H-Fourier	6	500	7000	1221	408576
	DeepONet	6	600	40000	1221+3	1
	GNOT	4	256	300000	1221+3	1

Table A3: **Seismogram model details.**

Wave type	Model	Layer num	Hidden size	Training steps	Input size	Output size
Acoustic	MLP	6	500	10000	405	205350
	H-Fourier	6	500	3000	405	205350
	DeepONet	6	600	100000	405+3	1
	GNOT	4	256	200000	405+3	1
Elastic	MLP	6	500	10000	1221	205350
	H-Fourier	6	500	4000	1221	205350
	DeepONet	6	600	100000	1221+3	1
	GNOT	4	256	300000	1221+3	1

G Discussion

G.1 Comparison of Modern ML-based Inversion

Here we discuss the difference between modern ML-based inversion approaches. The two most dominant schools are PINN and neural operator learning.

We first recall the formulation of the inversion problem.

$$c^* = \arg \min_c J(\mathbf{d}, \Phi(c, s)) + \lambda F(c, s). \quad (\text{A27})$$

In comparing the inversion strategies of PINN and operator learning, we observe distinct paths of optimization. The PINN-based approach seeks to optimize the model parameters c in conjunction with the forward modeling function Φ [51]. Thus, the c and Φ are closely coupled and the inversion of different structures requires training from scratch. This method limits the model’s ability to generalize across different velocity structures.

In contrast, the operator learning approach addresses this limitation by initially learning a universal operator to model forward problems across all c [74]. Subsequently, it fine-tunes the model parameters c for each specific case using the pre-trained Φ . Once trained, the neural operator can achieve inversion on any structures in inference time.

Our study adopts the latter approach to achieve fast and flexible inversion. We study both the sampling-based and optimization-based methods when using the pre-trained neural operator to solve the inversion problem.

G.2 Is Predicting Wavefield Necessary for Inversion?

As shown in Section 3.2.2, direct inverse mapping demonstrates better inversion performance than the optimization-based method. We want to discuss whether predicting wavefield is still necessary from the aspects of complexity, interpretability, and flexibility.

Complexity Direct ML inversion methods offer an efficient alternative to traditional optimization-based approaches by using neural networks to approximate the inverse mapping from data to subsurface properties directly. This method significantly cuts down on computational time and resources, simplifying the traditional iterative inversion process into a single step. In contrast, splitting the FWI process into two stages might increase complexity regarding the implementation and integration of different computational frameworks or software tools. Additionally, more stages could lead to greater cumulative errors.

Interpretability Predicting the wavefield explicitly as part of the inversion process can enhance the interpretability of the results. In FWI, the aim is to reconstruct the subsurface properties using wave propagation data. Although direct inverse mapping can produce a result of the inferred structure, it doesn’t explicitly model the intermediate process and tends to lack transparency in its operations. By decomposing the inversion into two distinct phases—predicting the forward wavefield and then optimizing the subsurface model—researchers can trace the effects of specific inputs or changes in model parameters on the outcomes. This division aligns closely with traditional FWI methods where each step of the process can be examined and understood separately. Furthermore, explicitly modeling the wavefield allows for a more detailed examination of the physical accuracy and stability of the inversion process, enabling researchers to validate and refine the underlying physical models being used.

Flexibility Predicting the wavefield explicitly provides opportunities to apply a broader range of optimization strategies tailored to specific challenges or characteristics of the data. For example, if the predicted wavefield is treated as a separable component, researchers can implement advanced gradient methods such as adaptive step-sizing or preconditioning. These techniques can address specific dynamics of wave propagation in different geological settings, potentially enhancing the accuracy and efficiency of the inversion. Preconditioning, in particular, can help tackle issues like the ill-posed nature of the inversion problem or the presence of sharp local minima by altering the optimization landscape.

G.3 Connection with Other Fields

Similar to the challenges of interpreting the human mind in the field of intuitive physics, the task of inverting earth’s structure involves comparable approaches and discussions.

Both seismic inversion and intuitive physics deal with inverse problems—inferring causes from effects. Seismic inversion back-calculates to infer the earth’s interior from surface measurements [65]. Similarly, intuitive physics tries to deduce the cognitive rules or mental models that people use to predict physical outcomes from their observed behavior in experimental settings [28].

Regarding methodologies, both fields heavily utilize computational models. In seismic studies, numerical methods and statistical models predict how geological features influence seismic waveforms [63, 66]. Similarly, in intuitive physics, physics engines simulate how humans might predict physical interactions in their surroundings [2, 34]. Machine learning, particularly neural networks, plays a significant role in both areas. In seismic inversion, neural networks can be trained to recognize patterns in seismic data that correlate with specific subsurface structures [74]. In intuitive physics, neural networks model human prediction and reasoning processes about physical laws, learning to mimic human judgment [33].

Uncertainty management is crucial in both domains. In seismic inversion, uncertainties arise from limited sensor coverage and noise in data. In intuitive physics, uncertainties stem from the variability in human perception and cognitive biases. Both fields develop methodologies to handle these uncertainties, often through probabilistic models [65, 2].

H Datasheet

H.1 Motivation

- **For what purpose was the dataset created?** GlobalTomo was created to promote the physics-ML research on solving seismic full waveform modeling and earth structure inversion.
- **Who created the dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?** The dataset was created by a group of geophysical scientists and ML researchers.

H.2 Composition

- **What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)?** GlobalTomo contains many simulated wavefield sequences and seismogram time series documented as arrays in HDF5 files.
- **How many instances are there in total (of each type, if appropriate)?** The acoustic tier has 10,000 cases of simulation results. The elastic tier has 30,000 cases and the real elastic tier has 10,000 cases.
- **Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set?** Yes, the dataset contains all possible instances for training and evaluation.
- **What data does each instance consist of?** Each instance consists of inputs and outputs. The inputs have velocity structures and source parameters. The outputs contains three types of features including the wavefield, the Fourier wavefield, and the seismogram. The wavefield includes a sequence of snapshots and each snapshot consists of 16 slices, together with the 3D coordinates and timesteps. The Fourier wavefield resembles the wavefield but presents each 3D snapshot in terms of 16 Fourier series. The seismogram consists of time series data from multiple stations, which also provide the 3D coordinates of each station and corresponding timesteps.
- **Is there a label or target associated with each instance?** Yes, each instance has clearly defined input and output variables.
- **Is any information missing from individual instances?** No.
- **Are there recommended data splits (e.g., training, development/validation, testing)?** We recommend using 90% of the data for training/validation and the remaining 10% for testing.
- **Are there any errors, sources of noise, or redundancies in the dataset?** The velocity structures are replicated across various features to facilitate data loading and training. Given the small size of these structures, the redundancy in storage can be considered negligible.
- **Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)?** The dataset is self-contained.

- **Does the dataset contain data that might be considered confidential?** No.
- **Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety?** No.

H.3 Collection Process

- **How was the data associated with each instance acquired?** Was the data directly observable (e.g., raw text, movie ratings), reported by subjects (e.g., survey responses), or indirectly inferred/derived from other data (e.g., part-of-speech tags, model-based guesses for age or language)? The data was collected from a set of simulations using a 3D global seismic model named AxiSEM3D.
- **What mechanisms or procedures were used to collect the data (e.g., hardware apparatuses or sensors, manual human curation, software programs, software APIs)?** We utilized a cluster of CPUs to run the AxiSEM3D.
- **Who was involved in the data collection process (e.g., students, crowdworkers, contractors) and how were they compensated (e.g., how much were crowdworkers paid)?** Geophysics researchers in the author list were involved in the data collection process and no crowdworkers were involved.

H.4 Preprocessing/cleaning/labeling

- **Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)?** We examine the failure cases in simulation and remove them from training and evaluation. The data is processed into HDF5 files with labeled variables for easy loading.
- **Was the “raw” data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)?** Yes, the raw data is saved in our cluster.
- **Is the software that was used to preprocess/clean/label the data available?** Yes, we uploaded the code on GitHub.

H.5 Uses

- **Has the dataset been used for any tasks already?** No.
- **What (other) tasks could the dataset be used for?** The dataset could be used for seismic forward modeling, earth structure inversion, and source inversion.
- **Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses?** No.
- **Are there tasks for which the dataset should not be used?** No.

H.6 Distribution

- **Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created?** The dataset is open to the public.
- **How will the dataset be distributed (e.g., tarball on website, API, GitHub)?** The dataset will be distributed on Google Drive and the code will be published on GitHub.
- **Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)?** The dataset will be distributed under the CC BY-NC-SA license.
- **Have any third parties imposed IP-based or other restrictions on the data associated with the instances?** No.
- **Do any export controls or other regulatory restrictions apply to the dataset or to individual instances?** No.

H.7 Maintenance

- **Who will be supporting/hosting/maintaining the dataset?** The GlobalTomo group will support, host, and maintain the dataset.
- **Is there an erratum?** No.
- **Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)?** Yes, the dataset will be continuously updated if there is a necessity to improve accuracy and other related information. The updates will be released on the website.
- **If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (e.g., were the individuals in question told that their data would be retained for a fixed period of time and then deleted)?** The dataset is not related to people.
- **Will older versions of the dataset continue to be supported/hosted/maintained?** Yes, older versions of the dataset will continue to be supported, maintained, and hosted.
- **If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so?** Yes, the contributor can contact us through email.

I More Visualization

I.1 Forward Modeling

We further present the visualization of forward modeling on four novel velocity structures to demonstrate that the ML baselines are capable of discerning differences in wave propagation. We display snapshots at timesteps 5, 7, 9, 11, and 13, illustrating that the variance in wavefields grows over time. Refer to [Figures A5 to A9](#) for detailed visualizations. We also show how our model predicts elastic wave propagation in [Figure A10](#).

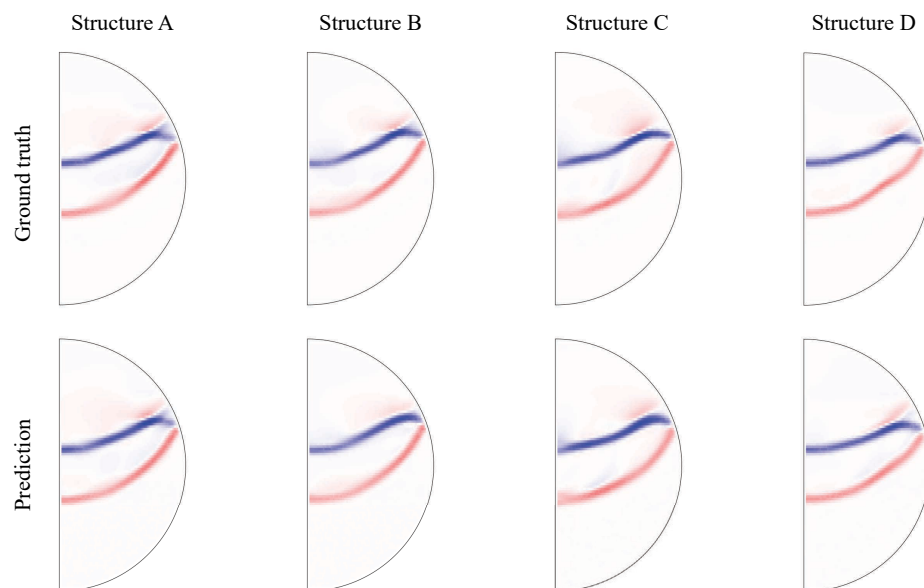


Figure A5: **Snapshot 5 of forward modeling.** We show the ground truth and predicted wavefield snapshot within four velocity structures: A, B, C, and D.

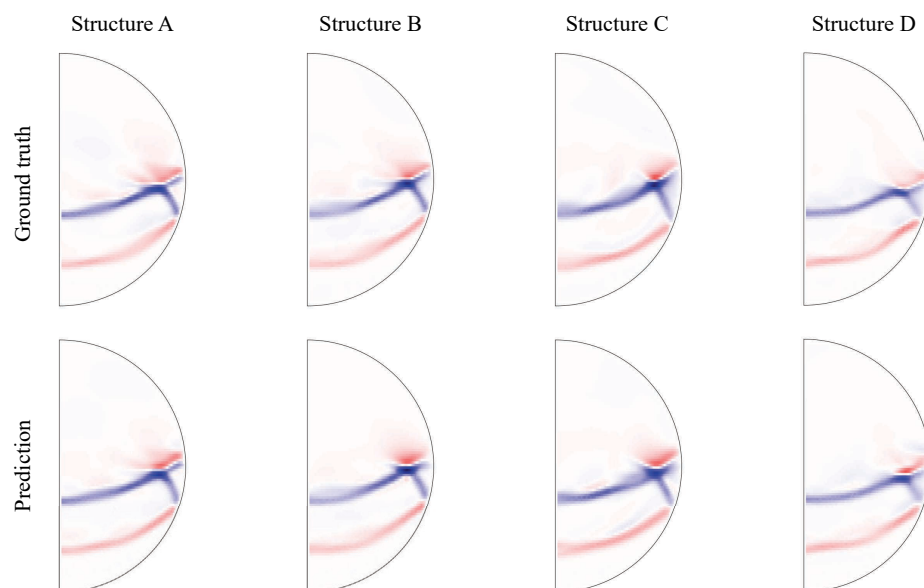


Figure A6: **Snapshot 7 of forward modeling.** We show the ground truth and predicted wavefield snapshot within four velocity structures: A, B, C, and D.

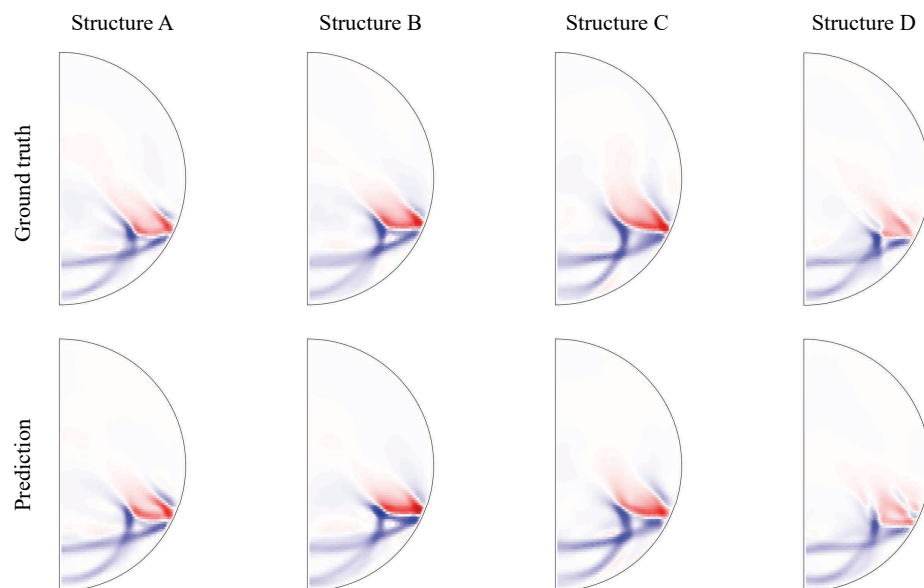


Figure A7: **Snapshot 9 of forward modeling.** We show the ground truth and predicted wavefield snapshot within four velocity structures: A, B, C, and D.

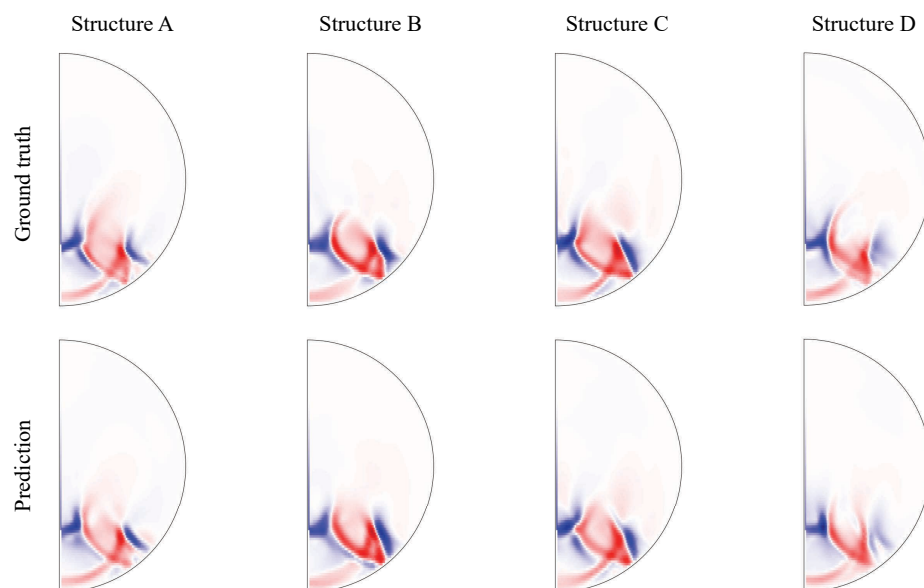


Figure A8: **Snapshot 11 of forward modeling.** We show the ground truth and predicted wavefield snapshot within four velocity structures: A, B, C, and D.

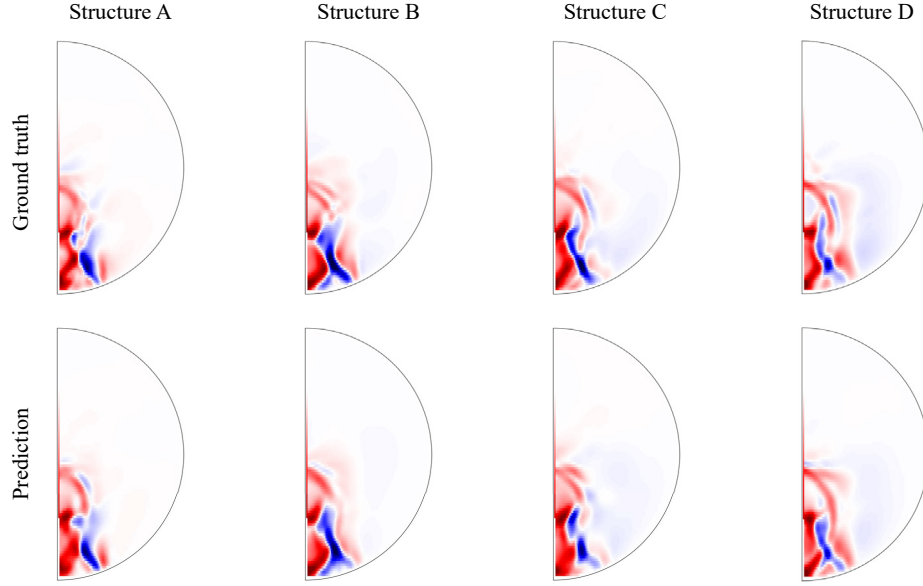


Figure A9: **Snapshot 13 of forward modeling.** We show the ground truth and predicted wavefield snapshot within four velocity structures: A, B, C, and D.

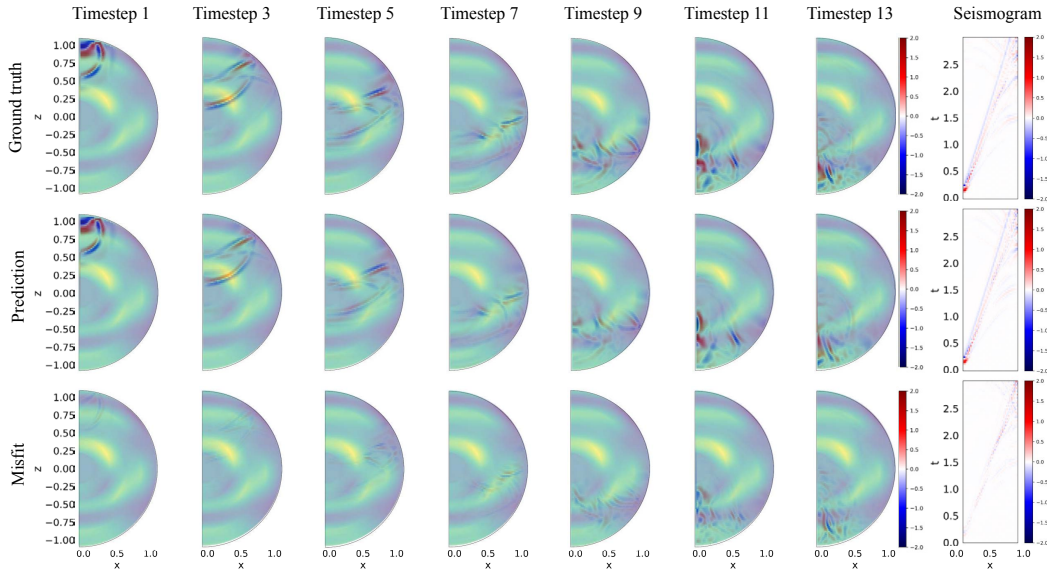


Figure A10: **Qualitative forward modeling prediction by MLP in the Elastic tier.** A single slice is displayed. The source is located at $x = 0.00$ and $z = 0.80$. The background illustrates the velocity structure. The seismogram depicts the time series received by stations around the surface of this slice.

I.2 Inversion

We present additional visualization of inversion results using various unseen seismogram data. The figures display both the actual and the inverted velocity perturbations across five depths, along with their correlations. Refer to [Figures A11 to A20](#).

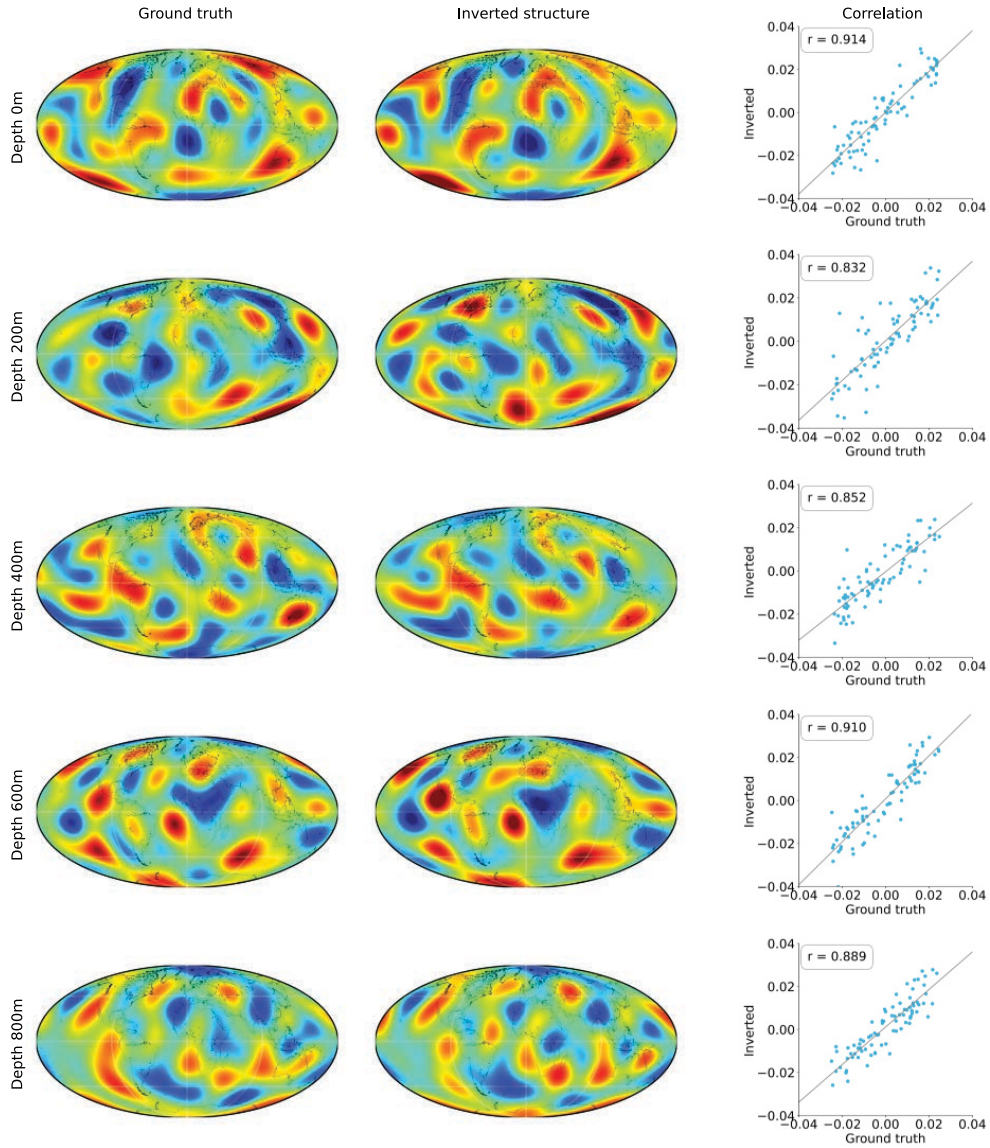


Figure A11: **Example 1 of acoustic inversion.** The colors range from blue to red, representing velocity perturbations from negative to positive. There are five rows, each depicting the velocity structure at different depths: 0m, 200m, 400m, 600m, and 800m. The first column displays the actual velocity structure, the second column shows the structure as derived from direct inverse mapping, and the third column illustrates the correlation between the spherical harmonics of the actual structure and those of the inverted structure.

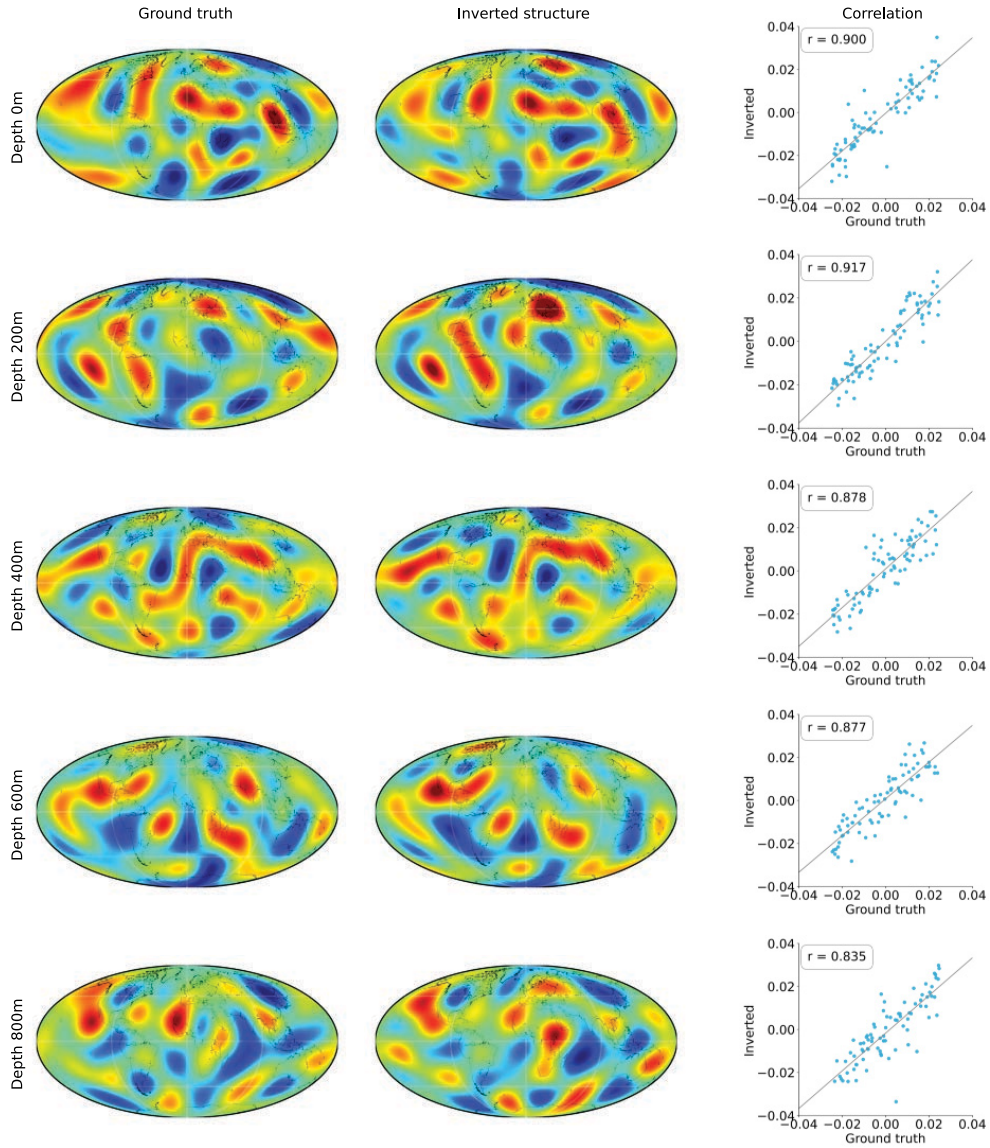


Figure A12: **Example 2 of acoustic inversion.** The colors range from blue to red, representing velocity perturbations from negative to positive. There are five rows, each depicting the velocity structure at different depths: 0m, 200m, 400m, 600m, and 800m. The first column displays the actual velocity structure, the second column shows the structure as derived from direct inverse mapping, and the third column illustrates the correlation between the spherical harmonics of the actual structure and those of the inverted structure.

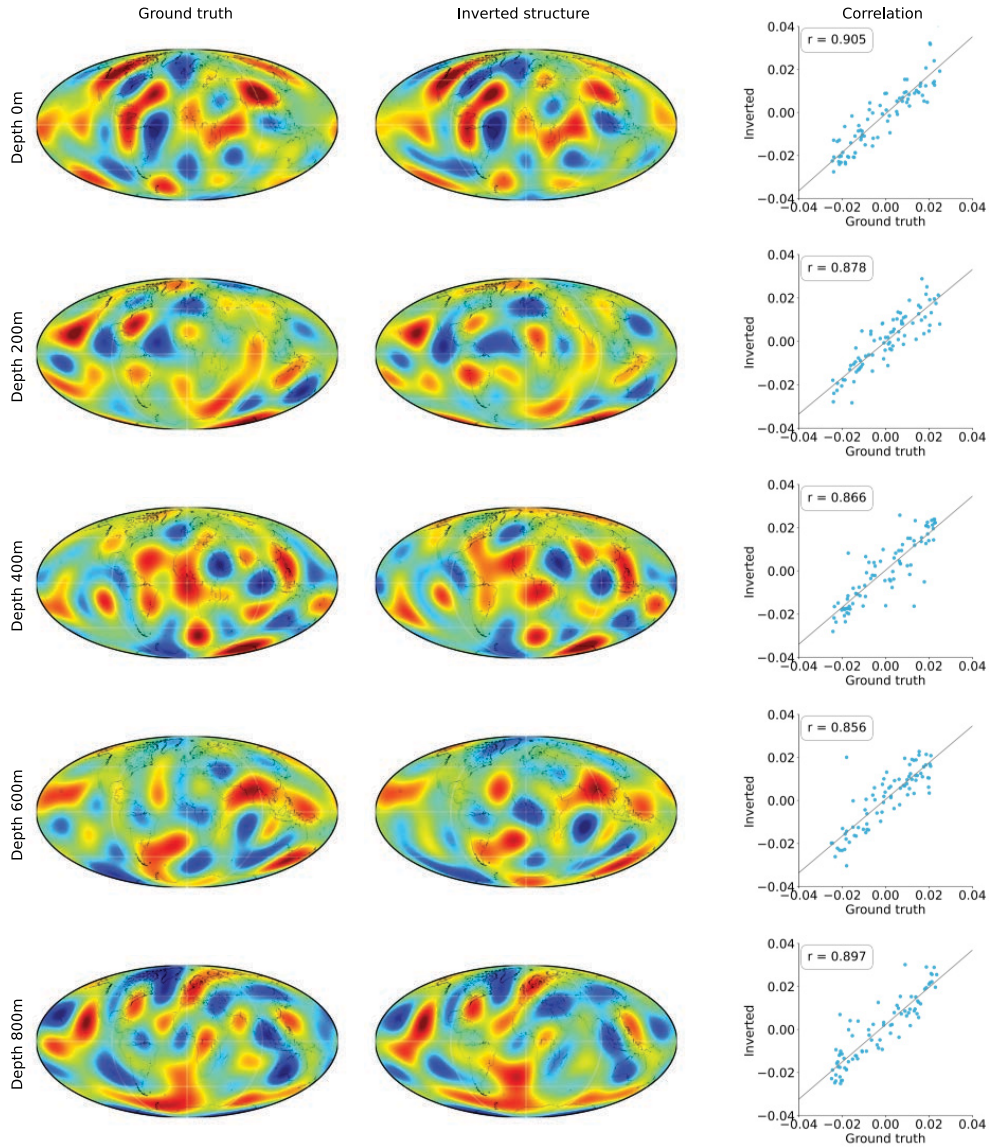


Figure A13: **Example 3 of acoustic inversion.** The colors range from blue to red, representing velocity perturbations from negative to positive. There are five rows, each depicting the velocity structure at different depths: 0m, 200m, 400m, 600m, and 800m. The first column displays the actual velocity structure, the second column shows the structure as derived from direct inverse mapping, and the third column illustrates the correlation between the spherical harmonics of the actual structure and those of the inverted structure.

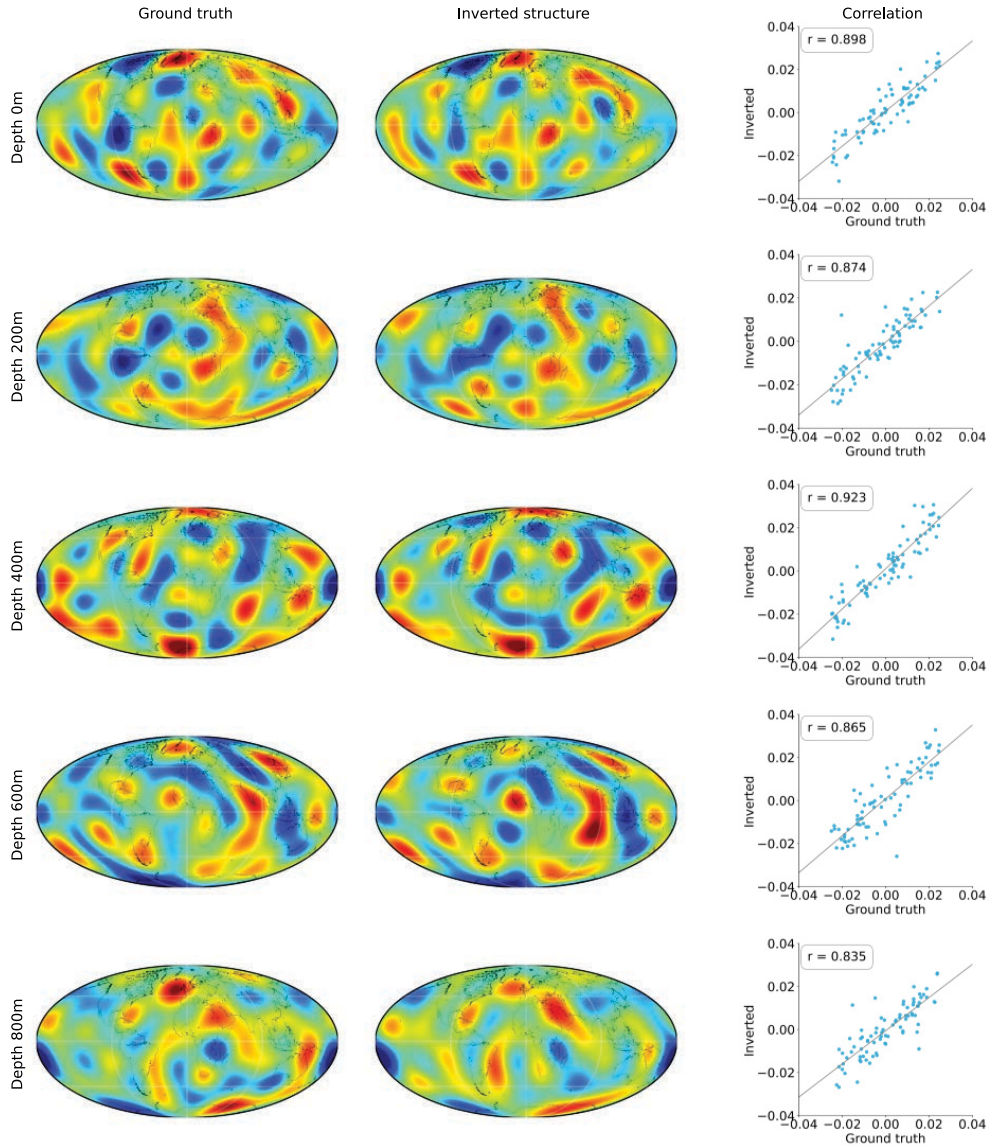


Figure A14: **Example 4 of acoustic inversion.** The colors range from blue to red, representing velocity perturbations from negative to positive. There are five rows, each depicting the velocity structure at different depths: 0m, 200m, 400m, 600m, and 800m. The first column displays the actual velocity structure, the second column shows the structure as derived from direct inverse mapping, and the third column illustrates the correlation between the spherical harmonics of the actual structure and those of the inverted structure.

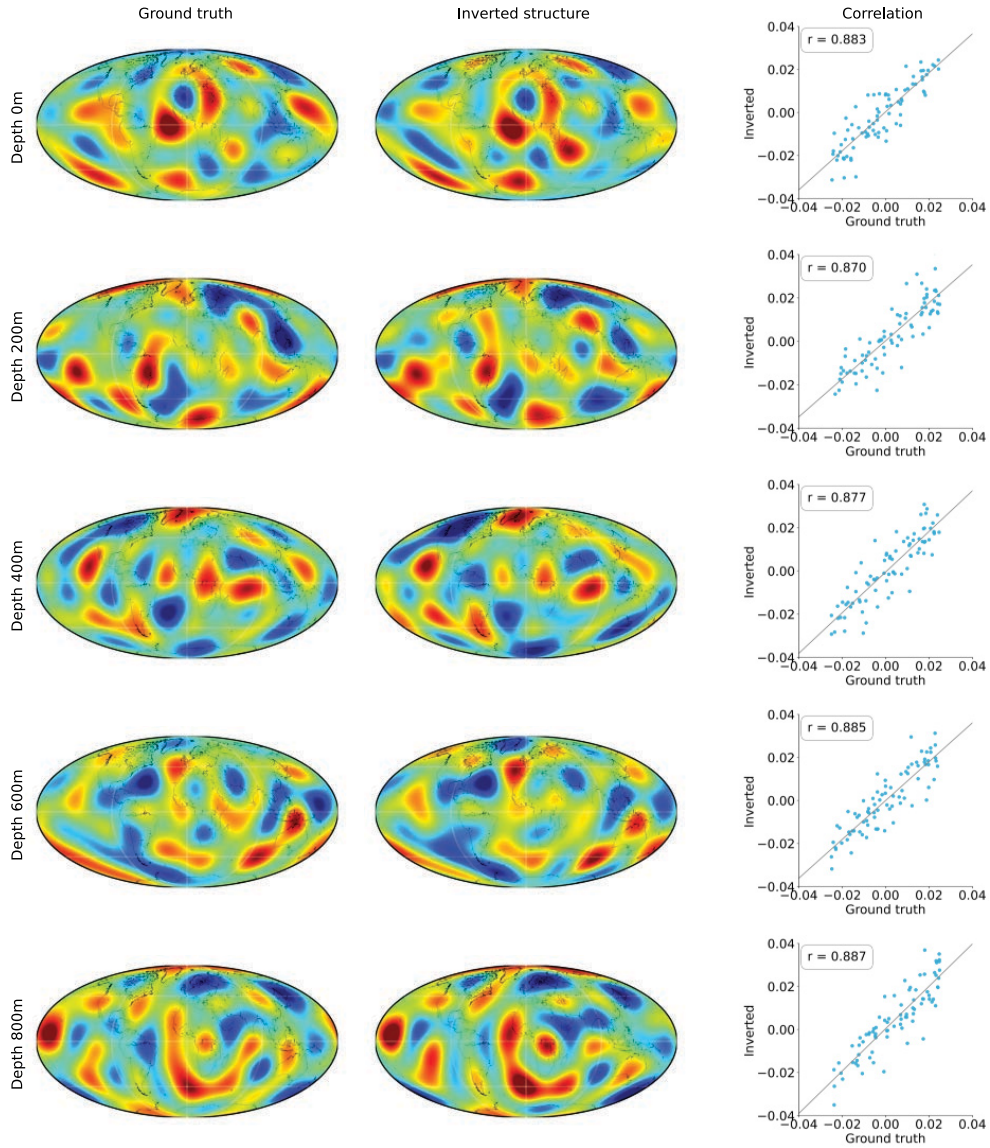


Figure A15: **Example 5 of acoustic inversion.** The colors range from blue to red, representing velocity perturbations from negative to positive. There are five rows, each depicting the velocity structure at different depths: 0m, 200m, 400m, 600m, and 800m. The first column displays the actual velocity structure, the second column shows the structure as derived from direct inverse mapping, and the third column illustrates the correlation between the spherical harmonics of the actual structure and those of the inverted structure.

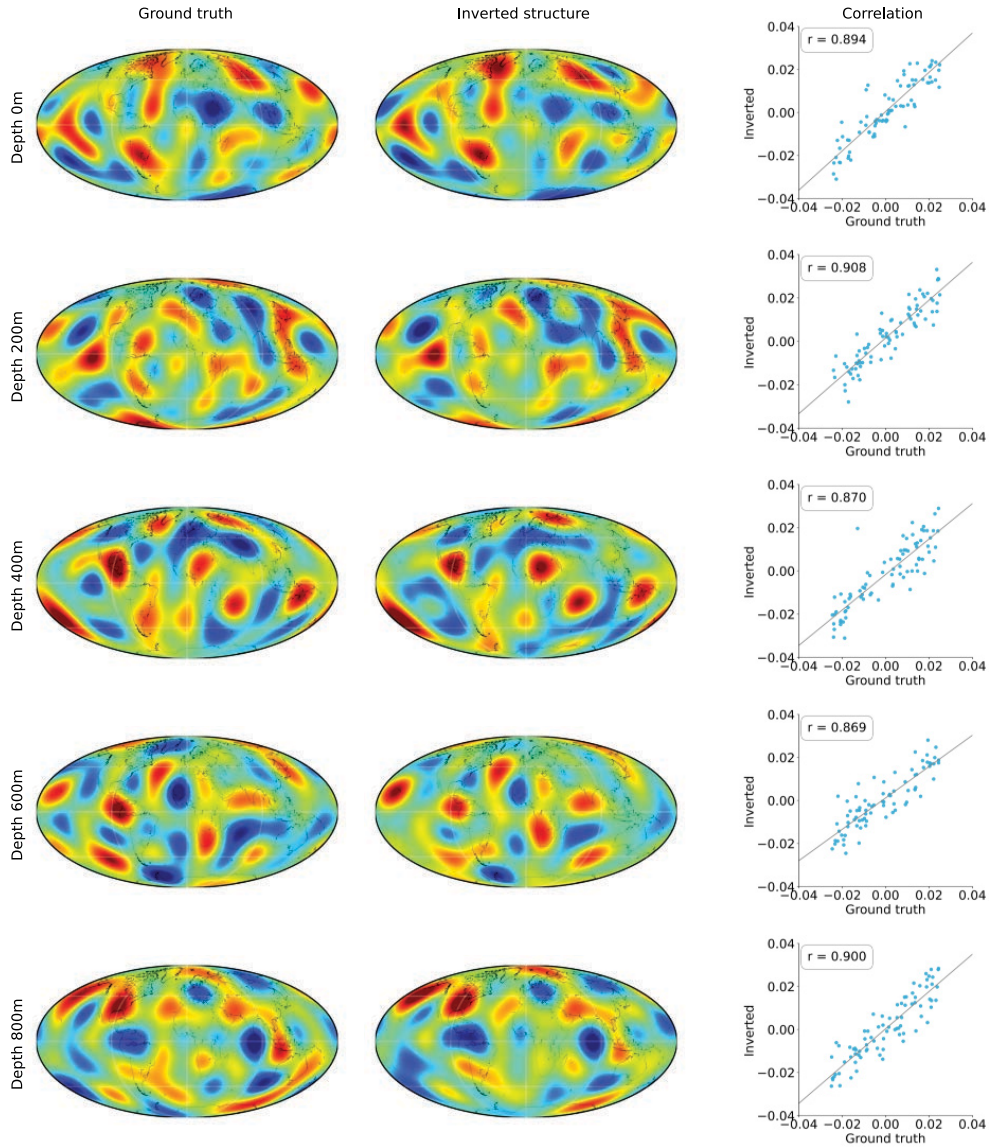


Figure A16: **Example 6 of acoustic inversion.** The colors range from blue to red, representing velocity perturbations from negative to positive. There are five rows, each depicting the velocity structure at different depths: 0m, 200m, 400m, 600m, and 800m. The first column displays the actual velocity structure, the second column shows the structure as derived from direct inverse mapping, and the third column illustrates the correlation between the spherical harmonics of the actual structure and those of the inverted structure.

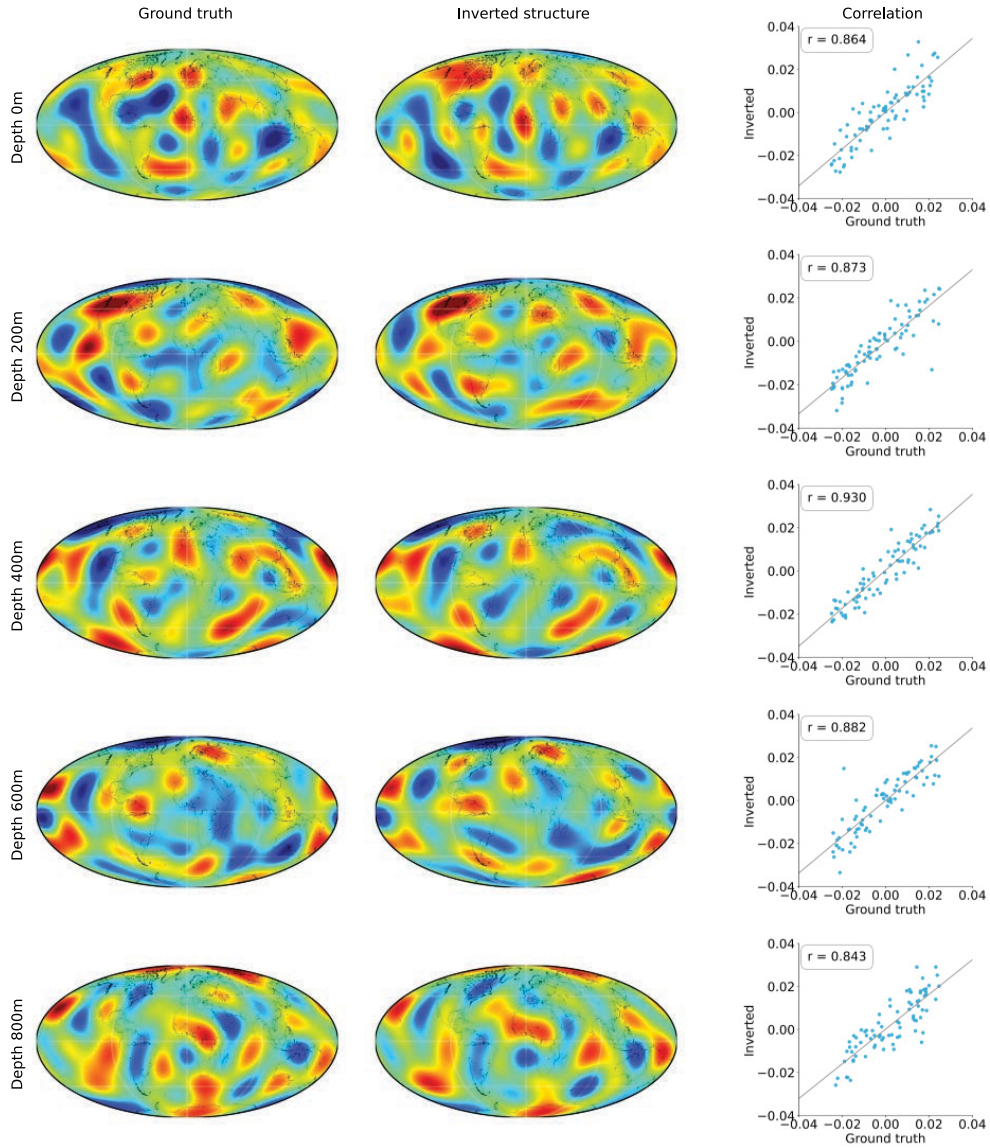


Figure A17: **Example 7 of acoustic inversion.** The colors range from blue to red, representing velocity perturbations from negative to positive. There are five rows, each depicting the velocity structure at different depths: 0m, 200m, 400m, 600m, and 800m. The first column displays the actual velocity structure, the second column shows the structure as derived from direct inverse mapping, and the third column illustrates the correlation between the spherical harmonics of the actual structure and those of the inverted structure.

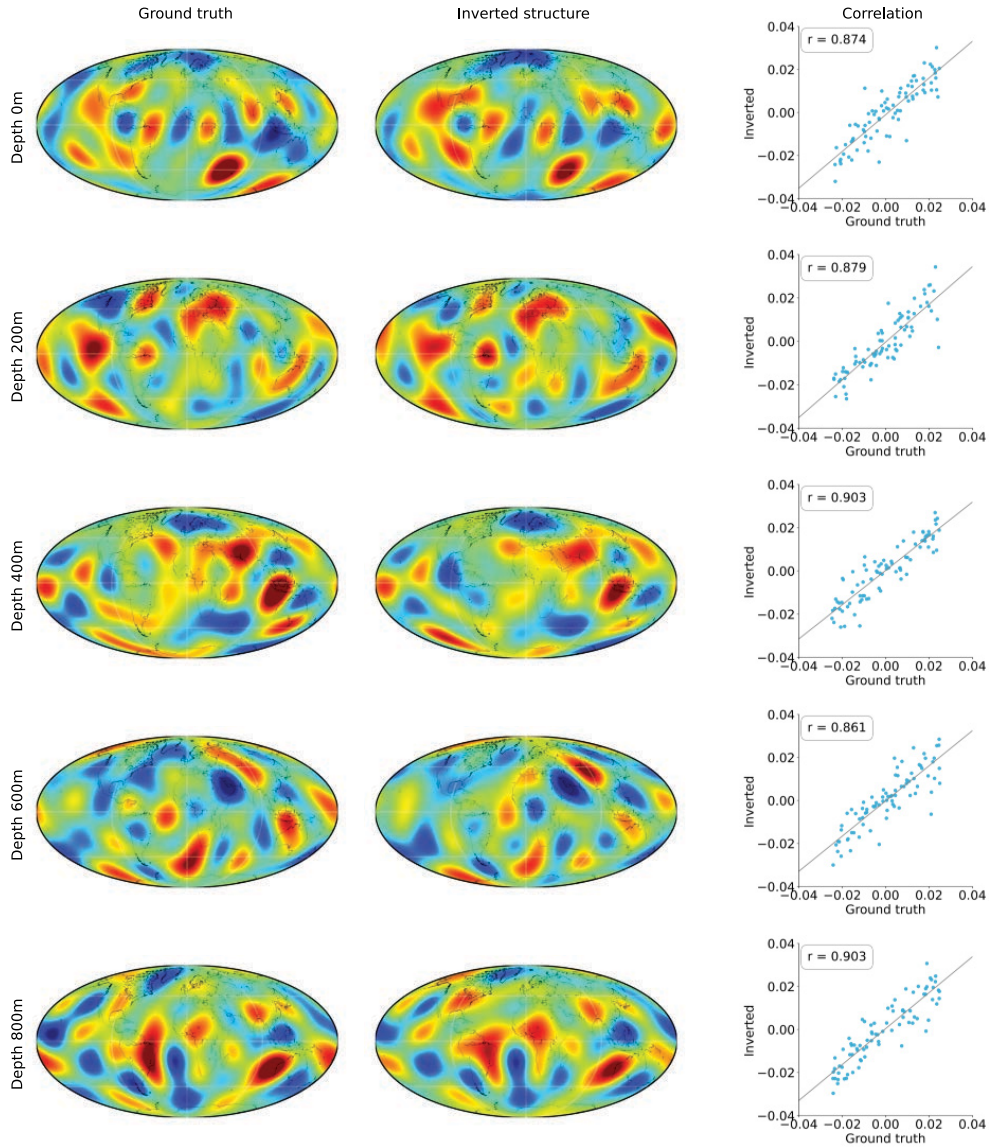


Figure A18: **Example 8 of acoustic inversion.** The colors range from blue to red, representing velocity perturbations from negative to positive. There are five rows, each depicting the velocity structure at different depths: 0m, 200m, 400m, 600m, and 800m. The first column displays the actual velocity structure, the second column shows the structure as derived from direct inverse mapping, and the third column illustrates the correlation between the spherical harmonics of the actual structure and those of the inverted structure.

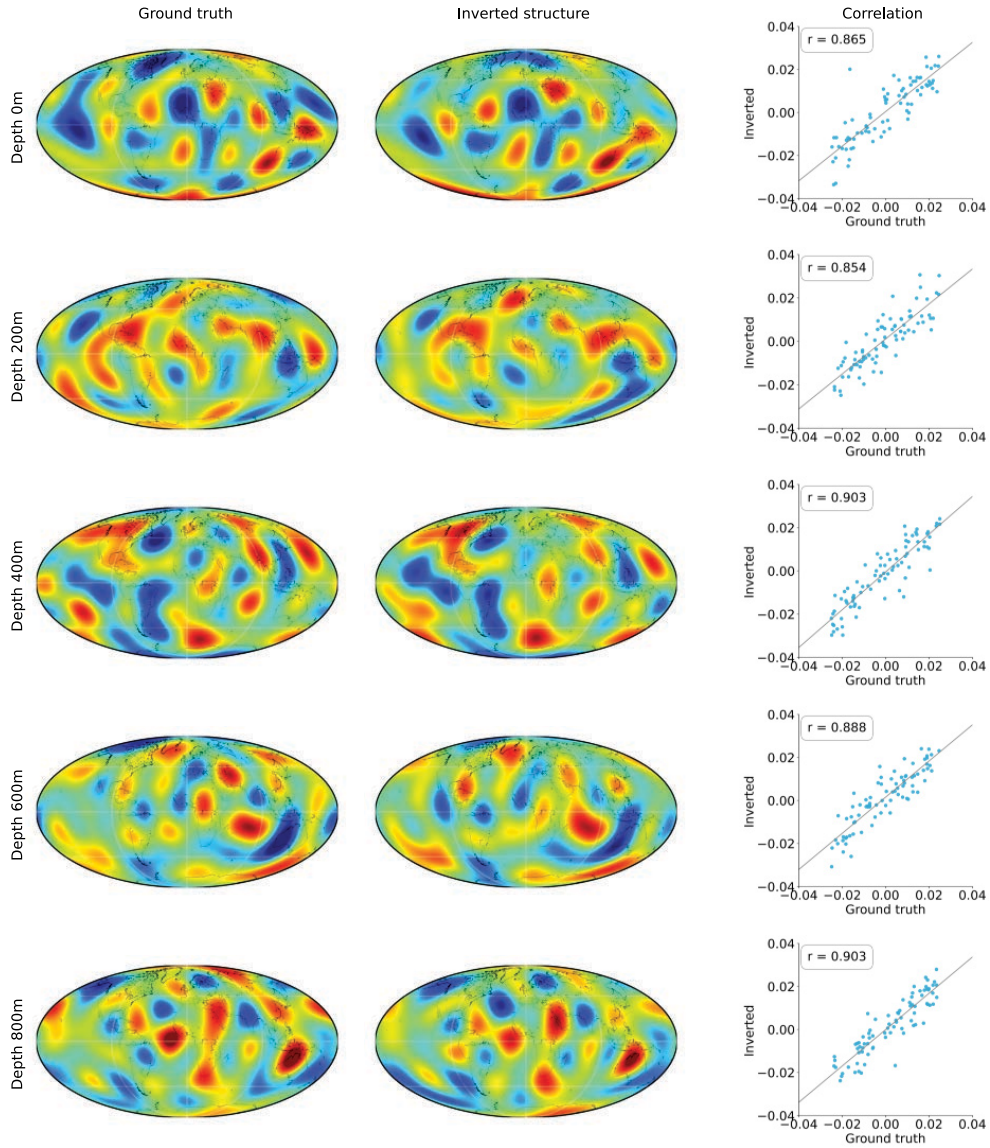


Figure A19: **Example 9 of acoustic inversion.** The colors range from blue to red, representing velocity perturbations from negative to positive. There are five rows, each depicting the velocity structure at different depths: 0m, 200m, 400m, 600m, and 800m. The first column displays the actual velocity structure, the second column shows the structure as derived from direct inverse mapping, and the third column illustrates the correlation between the spherical harmonics of the actual structure and those of the inverted structure.

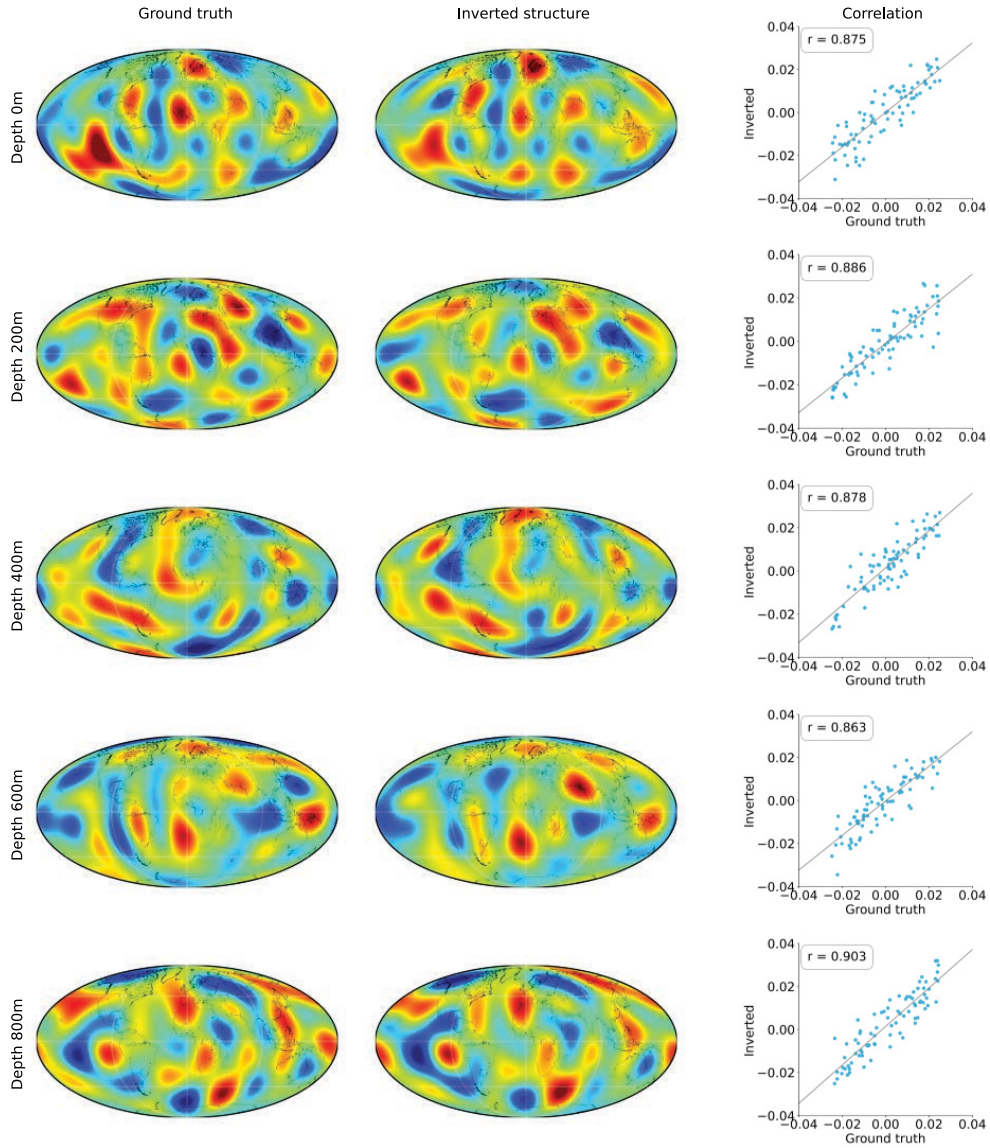


Figure A20: **Example 10 of acoustic inversion.** The colors range from blue to red, representing velocity perturbations from negative to positive. There are five rows, each depicting the velocity structure at different depths: 0m, 200m, 400m, 600m, and 800m. The first column displays the actual velocity structure, the second column shows the structure as derived from direct inverse mapping, and the third column illustrates the correlation between the spherical harmonics of the actual structure and those of the inverted structure.

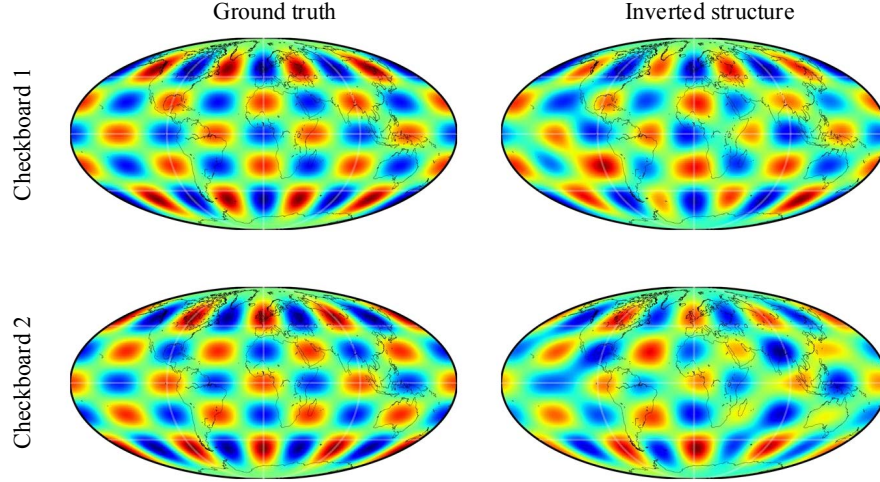


Figure A21: A **checkboard test on our ML-based inversion**. This test demonstrates the robustness and generalization of our inversion workflow.

I.3 A Checkboard Test

To validate the robustness of our inversion methods, we conducted a checkerboard test as a standard diagnostic tool in geophysical tomography [52, 65]. The test involves simulating a known subsurface model with alternating high and low-velocity regions arranged in a checkerboard pattern, which allows us to evaluate the algorithm’s ability to recover fine-scale variations in the subsurface. We evaluate the checkboard inversion using our pre-trained InversionMLP. The inversion results demonstrated a high correlation of 0.89 with the ground truth structures, with the algorithm successfully resolving the checkerboard structure and accurately recovering the modeled velocity contrasts; see [Figure A21](#) for visualization. This robust performance indicates that our approach is capable of handling complex geophysical inversion tasks and offers significant potential for real-world applications in seismic tomography and other geophysical imaging techniques.