REAL-DATA-GUIDED SEISMIC FORWARD MODELING IN LATENT SPACE

Anonymous authors
Paper under double-blind review

ABSTRACT

Seismic forward modeling (SFM) predicts wave responses from subsurface velocity fields, but learning this mapping accurately remains challenging because seismic responses are high-dimensional and paired real-world velocity–response data are scarce. Nevertheless, we argue that real earthquake records can still provide informative waveform trends for data-driven SFM. Motivated by this insight, we apply Parallel Symbolic Enumeration (PSE) to event–station records to discover empirical relationships. After offline physics-based screening with assistance from a large language model (LLM), the retained relationships can be transformed into normalized trend templates that regularize training at no additional inference cost. Building on this real-data-derived guidance, we propose Latent Seismic Predictor (LSP), a coarse-to-refined latent framework that separates dominant response prediction from oscillatory detail recovery. LSP first maps the subsurface velocity field into a coarse response latent capturing principal arrivals, reflection geometry, and broad energy distribution, and then applies attention-guided spectral residual refinement to recover unresolved waveform components through tapered sine–cosine basis functions. Moreover, learnable tapering suppresses spurious oscillations before residual synthesis, after which the refined latent representation is decoded into the complete seismic response. Across ten OpenFWI subsets, LSP achieves the best average rank in MAE, MSE, and SSIM, reducing average MAE and MSE on the challenging Style subsets by 11.4% and 26.2%, respectively, over the strongest baseline. When transferred from Style-B to Marmousi and Overthrust without fine-tuning, it reduces average MAE by 24.5%. The leading performance on GlobalTomo also supports its extension to high-dimensional 3D acoustic wave prediction. Codes are attached.

1 INTRODUCTION

SFM predicts seismic wave responses from subsurface velocity fields, supporting massive practical applications, including subsurface imaging, full-waveform inversion, earthquake-hazard analysis, and deep-earth exploration (Tarantola, 1984; Virieux & Operto, 2009). Traditional numerical methods, such as finite-difference and spectral-element solvers, can produce accurate results (Dablain, 1986; Komatitsch et al., 1999). However, their computational cost grows rapidly with spatial resolution, simulation time, geological complexity, and problem dimension. This cost is especially high when forward simulations must be repeated during inversion or uncertainty analysis. Learning-based methods provide a faster alternative by approximating the velocity-to-response mapping (Moseley et al., 2020; Li et al., 2023). Convolutional encoder–decoders, Fourier neural operators (FNOs), and DeepONet-style models provide common architectures for learned operator approximation (Ronneberger et al., 2015; Li et al., 2021; Lu et al., 2021), while recent operator-learning studies have explored low-rank parameterization and frequency-domain scattered-wavefield prediction (Koren et al., 2026; Huang & Alkhalifah, 2025). Nevertheless, complete multi-source time-domain responses remain difficult to predict because they contain complex reflections, diffractions, interference, and oscillations. Neural networks also favor smooth and low-frequency components, motivating progressive and multi-scale strategies for recovering high-frequency information (Rahaman et al., 2019; Kania et al., 2025; Han et al., 2026), while global operator layers may over-smooth localized solution structures (Liu-Schiaffini et al., 2024). Consequently, weak arrivals and localized waveform details can be insufficiently resolved.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

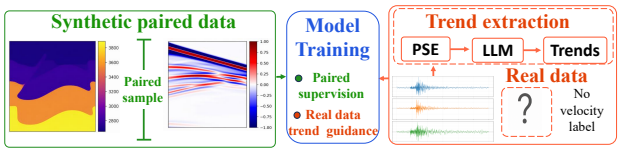


Figure 1: Overview of the two training signals used in this work. Synthetic velocity–response pairs provide direct supervision, while unpaired real earthquake records are processed to extract waveform trends that guide model optimization.

Another challenge is the scarcity of paired real-world training data. As illustrated in Fig. 1, synthetic benchmarks such as OpenFWI provide paired velocity fields and seismic responses (Deng et al., 2022). Real seismic archives contain many waveform records, but the corresponding high-resolution velocity fields are rarely available and remain costly and uncertain to recover (Alkhalifah et al., 2022). Consequently, real records cannot serve as paired training samples, although their source spectra, propagation attenuation, and amplitude variation still provide informative waveform trends.

To address these challenges, we propose LSP, a coarse-to-refined latent framework for seismic forward modeling. LSP first encodes the input velocity field into a compact representation and predicts the main structure of the seismic response. It then applies attention-guided spectral residual refinement with sine–cosine latent basis functions to recover details that are not fully captured by the coarse prediction. A cosine-initialized learnable taper further controls the contribution of different basis components before the refined representation is decoded into the final waveform. Note the sine and cosine functions operate on latent features rather than on physical Fourier modes.

More importantly, LSP also introduces real-data-derived trend regularization without requiring paired real velocity fields. We apply PSE-based symbolic regression (Ruan et al., 2026) to event–station records from the STanford EArthquake Dataset (STEAD) to generate candidate relationships among source scale, propagation distance, spectral characteristics, attenuation, and relative amplitude. On this basis, an LLM assists in checking whether the candidate expressions agree with basic physical knowledge and expected variable relationships. The overall construction of the real-data-derived trend regularization is illustrated in Fig. 2.

The selected expressions are not used to calculate seismic waveforms directly. Instead, they describe general trends observed in real earthquake records. During model optimization, these trends guide the predicted waveforms toward more realistic spectral, attenuation, and amplitude behavior. Recall that this regularization is used only during training and adds no computation during inference. We evaluate LSP across three complementary settings. The primary evaluation covers ten OpenFWI subsets with diverse geological structures. Out-of-distribution (OOD) generalization is examined by training all models only on Style-B and directly evaluating them on the original Marmousi and Overthrust fields, without fine-tuning or test-time adaptation. We further conduct a controlled experiment on the GlobalTomo Acoustic tier to assess the extension of LSP to high-dimensional 3D acoustic wave prediction. The main contributions are:

- We propose LSP, a coarse-to-refined latent predictor that models dominant response structures before recovering unresolved waveform components with tapered sine–cosine residual refinement.
- We derive soft spectral, attenuation, and amplitude constraints from unpaired real earthquake records through PSE and subsequent offline physics-based screening with LLM assistance, at no additional inference cost.
- Experiments on ten OpenFWI subsets, OOD geological structures, and a controlled 3D benchmark demonstrate the effectiveness and generalizability of LSP.

2 Related Work

SFM and learned operators. SFM traditionally solves wave equations using finite-difference, finite-element, or spectral-element schemes (Dablain, 1986; Komatitsch et al., 1999), but repeated

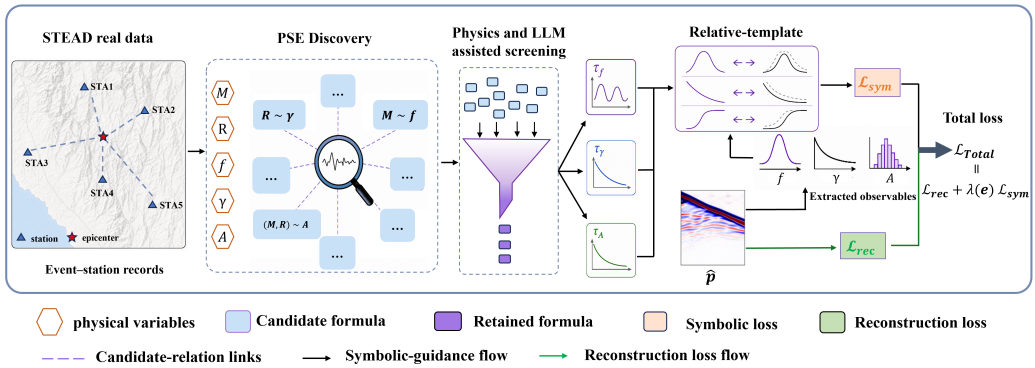


Figure 2: Overview of real-data-derived trend regularization. PSE discovers candidate relations from STEAD records, which are screened using physical criteria and LLM assistance, then converted into relative spectral, attenuation, and amplitude templates for training-time regularization.

simulation remains expensive. Learning-based approaches approximate the velocity-to-response mapping with convolutional networks, FNOs, and DeepONet-style operators (Ronneberger et al., 2015; Li et al., 2021; Lu et al., 2021; Deng et al., 2022; Li et al., 2023). Neural operators provide a general framework for learning mappings between function spaces, with recent work extending them through localized kernels and attention-based operator constructions (Kovachki et al., 2023; Liu-Schiaffini et al., 2024; Xiao et al., 2024; Wang et al., 2025). Other studies investigate recursive neural-operator formulations, low-rank operator parameterization, and frequency-domain scattered-wavefield prediction (Singh et al., 2024; Koren et al., 2026; Huang & Alkhalifah, 2025). These formulations provide relevant operator-learning context, whereas LSP directly predicts complete multi-source time-domain responses by combining coarse latent transport with subsequent structured residual refinement within a unified latent framework.

Symbolic regression and scientific regularization. Symbolic regression discovers compact scientific relationships from observations (Schmidt & Lipson, 2009; Cranmer et al., 2020; Brunton et al., 2016; Udrescu & Tegmark, 2020). Recent work improves formula discovery through structured search, noise robustness, dimension reduction, and control-variable strategies (Xu et al., 2024; Yu et al., 2025). PSE evaluates candidate expressions under accuracy–complexity trade-offs (Ruan et al., 2026). Because relationships extracted from real earthquakes cannot be treated as exact governing equations for synthetic benchmarks, LSP converts them into soft relative waveform constraints rather than explicit PDE residuals (Raissi et al., 2019).

3 Methodology

3.1 Seismic Forward Operator

Under the constant-density acoustic approximation, the wavefield generated by the s-th source satisfies

$$\frac{1}{v^2(\mathbf{x})}\frac{\partial^2 u_s(\mathbf{x},t)}{\partial t^2} - \nabla^2 u_s(\mathbf{x},t) = q_s(t)\delta(\mathbf{x}-\mathbf{x}_s), \tag{1}$$

where $v(\mathbf{x})$ is the subsurface velocity field, $q_s(t)$ is the source wavelet, and $\mathbf{x}_s$ is the source location. The seismic response recorded at receiver $\mathbf{x}_r$ is $p_{s,r}(t) = u_s(\mathbf{x}_r, t)$. For a fixed acquisition geometry, source configuration, and source wavelet, we denote the seismic forward operator by

$$\mathcal{G} : \mathcal{V} \to \mathcal{P}, \qquad p = \mathcal{G}(v), \tag{2}$$

where $\mathcal{V}$ and $\mathcal{P}$ denote the velocity-field and seismic-response spaces, respectively. The operator $\mathcal{G}$ maps a subsurface velocity field $v \in \mathcal{V}$ to its complete multi-source seismic response $p \in \mathcal{P}$. Accordingly, learning-based SFM approximates this velocity-to-response operator from paired samples under a fixed acquisition configuration.

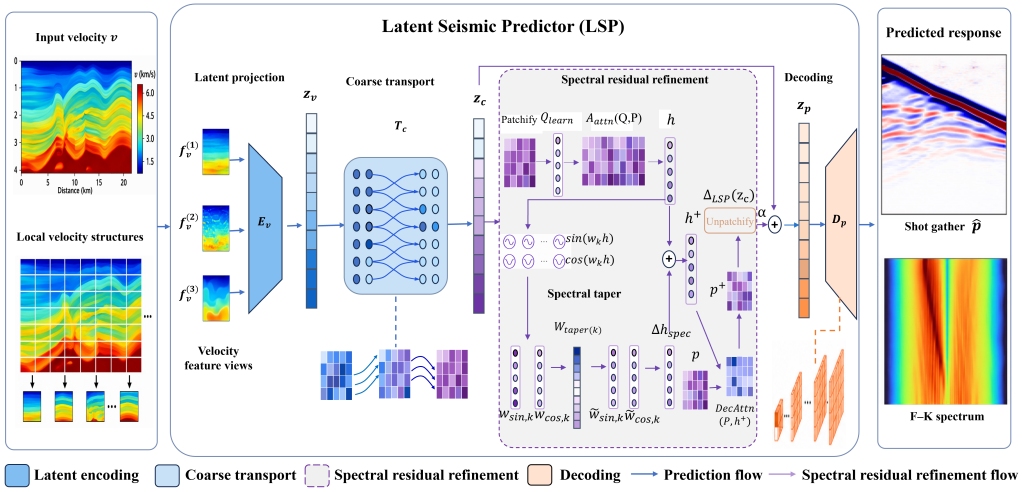


Figure 3: Overview of LSP. The velocity field is encoded into a compact latent representation and mapped to a coarse response latent. Attention-guided spectral residual refinement with learnable tapering then recovers unresolved waveform components before decoding the final seismic response.

In the 2D OpenFWI setting, $v \in \mathbb{R}^{1\times70\times70}$ and $p \in \mathbb{R}^{5\times1000\times70}$, corresponding to five sources, 1,000 time samples, and 70 receivers. This mapping is highly asymmetric: the input describes spatial variations of subsurface properties, whereas the output represents source-dependent wave propagation over the time–receiver domain. Large-scale velocity structures determine the dominant travel paths and reflection geometry, while the recorded responses additionally contain oscillatory arrivals, interference patterns, and amplitude decay. Accurately learning these effects with a single end-to-end neural mapping is therefore challenging.

3.2 Latent Seismic Predictor

Overall framework. To address the asymmetric mapping from spatial velocity structures to source-dependent seismic responses, we decompose the prediction process into velocity projection, coarse latent transport, response-oriented spectral residual refinement, and waveform reconstruction. As shown in Fig. 3, LSP is formulated as

$$z_v = E_v(v), \tag{3}$$
$$z_c = T_c(z_v), \tag{4}$$
$$z_p = z_c + \alpha \odot \Delta_{\mathrm{LSP}}(z_c), \tag{5}$$
$$\hat{p} = D_p(z_p). \tag{6}$$

The encoder E_v converts the velocity field into a compact geological representation z_v, and T_c transports it toward a coarse response latent z_c. Rather than recovering the complete oscillatory response at once, LSP preserves this coarse state and learns the missing waveform components through Δ_{LSP}. The learnable scale α controls the residual contribution, while D_p reconstructs the final shot gather.

This decomposition reflects the organization of seismic propagation. Background velocity variations and large-scale interfaces primarily determine the dominant travel-time and moveout patterns, whereas weaker reflections, diffractions, multipath propagation, and interference introduce finer waveform variations around this response backbone. Accordingly, LSP first models the dominant velocity-to-response transport and then represents the remaining waveform complexity as a structured residual within the response-oriented latent space.

Velocity projection and coarse transport. The velocity encoder extracts three complementary feature views $f_v^{(1)}$, $f_v^{(2)}$, and $f_v^{(3)}$ from different representation depths. The shallow view empha-

sizes local velocity contrasts around interfaces and faults, whereas the deeper views capture broader geological structures that influence propagation paths and travel times. These features are concatenated and aggregated through cross-attention with learnable velocity-domain queries, producing the compact latent representation z_v. This query-based aggregation retains response-relevant geological information while compressing the original velocity grid into a fixed-size latent representation for efficient response prediction under diverse geological settings. The coarse transport module maps z_v to the response-domain latent z_c, which captures the dominant organization across source and receiver dimensions, including principal arrivals, reflection geometry, and broad energy distributions. Fine waveform components that remain unresolved in z_c are subsequently recovered through residual refinement. Appendix A provides the corresponding tensor definitions, attention equations, and detailed encoder configurations.

3.3 Attention-Guided Spectral Residual Refinement

Seismic responses contain multiple arrivals distributed over time and receiver locations. To preserve this local organization, the coarse response latent is rearranged into response tokens and aggregated using learnable queries:

$$P = \mathrm{Patchify}(z_c), \qquad h = \mathcal{A}_{\mathrm{attn}}(Q_{\mathrm{learn}}, P). \tag{7}$$

where P denotes the response tokens and Q_{learn} contains learnable response queries that aggregate correlated latent structures into h. To reconstruct oscillatory variations that are not sufficiently represented by the coarse response latent, we construct complementary sine–cosine basis responses over the aggregated feature h. These basis functions operate only on the response-oriented latent correction rather than directly modeling the complete velocity-to-response mapping. Let $\widetilde{w}_{\sin,k} = W_{\mathrm{taper}}(k) \odot w_{\sin,k}$ and $\widetilde{w}_{\cos,k} = W_{\mathrm{taper}}(k) \odot w_{\cos,k}$. The spectral residual is then computed as

$$\begin{aligned} \Delta h_{\mathrm{spec}} = w_0 + \sum_{k=1}^{K} \Big[& \widetilde{w}_{\sin,k} \odot \sin(\omega_k h) \\ & + \widetilde{w}_{\cos,k} \odot \cos(\omega_k h) \Big]. \end{aligned} \tag{8}$$

$$h^{+} = h + \Delta h_{\mathrm{spec}}. \tag{9}$$

The sine and cosine branches provide complementary nonlinear basis responses, while $w_{\sin,k}$ and $w_{\cos,k}$ determine their data-adaptive contributions. The residual connection preserves the coherent response information already represented in h while incorporating the learned oscillatory correction into the refined latent representation.

The learnable taper $W_{\mathrm{taper}}(k)$ regulates the contribution of different basis components. It is initialized with a smooth cosine profile and optimized jointly with the predictor, suppressing components that may introduce spurious waveform oscillations while retaining useful latent corrections.

Although seismic waveforms are oscillatory, ω_k is not a physical temporal frequency or spatial wavenumber. Unlike Fourier-domain neural operators and spectral-domain learning methods (Li et al., 2021; Du et al., 2024; Cao et al., 2025), the trigonometric functions are applied element-wise to latent features; therefore, *spectral* refers to a latent basis expansion rather than a Fourier transform of the physical wavefield.

Finally, decoder-side attention redistributes h^{+} over the original response tokens, after which unpatchification restores the layout of z_c and produces $\Delta_{\mathrm{LSP}}(z_c)$. This step preserves the response-token organization while injecting the learned correction. The reconstructed correction is then fused with the coarse latent through the outer residual pathway, retaining the dominant propagation structure. The detailed attention equations, tensor dimensions, taper initialization, and patch reconstruction procedure are given in Appendix B. Because the spectral correction is incorporated into the latent representation before decoding, LSP produces the final waveform in an end-to-end manner and requires no Fourier-domain post-processing during inference.

3.4 Real-Data-Derived Trend Regularization

Synthetic simulations provide paired velocity fields and responses but cannot fully reproduce the variability of real source signatures, propagation losses, and diverse acquisition conditions. Con-

versely, real earthquake records contain realistic waveform statistics but generally lack paired high-resolution velocity fields. We therefore use synthetic pairs for direct reconstruction supervision and real records for distribution-level waveform guidance.

As detailed in Fig. 2, each STEAD event–station record is summarized by source scale M, source–receiver distance R, characteristic frequency f_c, attenuation coefficient γ, and relative amplitude A. PSE generates candidate relationships

$$f_c \sim g_f(M), \qquad \gamma \sim g_\gamma(R), \qquad A \sim g_A(M, R). \tag{10}$$

The candidates are ranked according to fitting quality and expression complexity. Physics-based criteria then remove relationships with implausible dependencies, signs, or monotonic behavior, with an LLM used only to assist in offline screening. Recall that neither PSE nor the LLM is included in neural-network training or inference.

The retained expressions are transformed into relative spectral, attenuation, and amplitude templates, τ_f, τ_γ, and τ_A. Correspondingly, the predicted response provides three differentiable observables: the normalized average spectrum, the normalized temporal amplitude evolution, and the receiver-wise root-mean-square (RMS) amplitude. We denote these observables by

$$o_f(\hat{p}) = \mathcal{N}_f(S_{\text{pred}}), \tag{11}$$

$$o_\gamma(\hat{p}) = B_{\text{pred}}, \tag{12}$$

$$o_A(\hat{p}) = \mathcal{C}_r\left[\log_{10}\left(A_{\text{pred}} + \epsilon\right)\right]. \tag{13}$$

Here, $\mathcal{N}_f$ removes absolute spectral energy and $\mathcal{C}_r$ removes the common receiver-wise amplitude offset. These relative transformations are necessary because STEAD and OpenFWI differ substantially in source mechanisms, acquisition geometry, and physical amplitude scale.

The resulting trend regularization loss is

$$\mathcal{L}_{\text{sym}} = \sum_{j \in \{f,\gamma,A\}} w_j \ \text{MSE}\left(o_j(\hat{p}), \tau_j\right). \tag{14}$$

This objective does not prescribe individual seismic arrivals or replace the paired synthetic targets. Instead, it encourages the predicted waveforms to follow relative spectral, attenuation, and receiver-wise amplitude trends observed in real earthquake records. The observable definitions, retained symbolic expressions, normalization operators, and term-wise regularization losses are provided in Appendix C.

3.5 Training Objective

The final loss combines waveform reconstruction from paired synthetic data with real-data-derived trend regularization:

$$\mathcal{L}_{\text{rec}} = \beta_1 \,\text{MAE}(\hat{p}, p) + \beta_2 \,\text{MSE}(\hat{p}, p). \tag{15}$$

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{rec}} + \lambda(e)\mathcal{L}_{\text{sym}}. \tag{16}$$

During the initial warm-up period, $\lambda(e)$ is set to zero so that the predictor first establishes a velocity-to-response mapping from paired simulations. The regularization weight is then increased smoothly, allowing real-data trends to provide auxiliary guidance without overriding the primary waveform reconstruction objective. At inference time, only the neural prediction pathway is evaluated, and the trend-regularization pathway introduces no additional inference cost.

4 Experiments and Analysis

4.1 Experimental Setup

We evaluate LSP on ten OpenFWI subsets (Deng et al., 2022), use Marmousi and Overthrust (Versteeg, 1994; Aminzadeh et al., 1996) for OOD evaluation, and conduct controlled 3D evaluation on GlobalTomo Acoustic (Li et al., 2026). OpenFWI is the primary benchmark across diverse geological structures, while the remaining datasets assess cross-structure generalization and high-dimensional extension. U-Net, FNO, DeepONet, and LSP use identical OpenFWI splits, normalization, reconstruction loss, and a 500-epoch Adam budget initialized at 10^{-3} (Kingma & Ba, 2014).

Style-B-trained models are tested on original and smoothed Marmousi and Overthrust fields without parameter updates or test-time fine-tuning. We report MAE, MSE, and SSIM for 2D experiments and RL2 and Pearson correlation for controlled 3D evaluation. Detailed processing and implementation settings are provided in Appendix D.

4.2 Main Results

Performance on Standard Benchmarks (OpenFWI). Table 1 summarizes the quantitative comparison on the OpenFWI datasets. No single model dominates all metrics across all geological regimes. On the relatively simple Vel subsets, FNO obtains the lowest MAE and MSE, suggesting that Fourier-domain operators are particularly effective when the dominant response patterns are comparatively smooth.

The advantage of LSP becomes more evident on structurally complex datasets, especially in the Fault and Style families. LSP achieves the best average rank for all three metrics and obtains competitive or superior performance on most complex subsets. In particular, its higher SSIM scores on CVA, CVB, FFA, FFB, CFB, STA, and STB are consistent with improved preservation of wavefront structures, scattering patterns, and interference-related features. Overall, LSP improves structural fidelity while retaining competitive pointwise accuracy, as further illustrated in Fig. 4.

Table 1: The forward-modeling results on OpenFWI. Best values are bold; arrows indicate the preferred metric direction.

Metric	Model	FVA	FVB	CVA	CVB	FFA	FFB	CFA	CFB	STA	STB	Avg Rank
MAE ↓	U-Net	0.0398	0.0973	0.1115	0.1598	0.0542	0.1090	0.0959	0.1391	0.0493	0.0674	3.8
	FNO	**0.0077**	**0.0385**	0.0700	0.1405	0.0654	0.0945	0.0776	0.1329	0.0365	0.0551	2.5
	DeepONet	0.0358	0.1198	0.0781	0.1345	**0.0282**	0.0829	0.0650	0.1075	0.0375	0.0424	2.4
	LSP (Ours)	0.0302	0.0903	**0.0683**	**0.1166**	0.0496	**0.0819**	**0.0606**	**0.1032**	**0.0342**	**0.0366**	**1.3**
MSE ↓	U-Net	0.0173	0.0427	0.0301	0.1089	0.0344	0.0675	0.0442	0.0897	0.0153	0.0182	3.5
	FNO	**0.0024**	**0.0166**	0.0391	0.0825	0.0317	0.0679	0.0371	0.0839	0.0102	0.0130	2.8
	DeepONet	0.0302	0.0493	**0.0259**	0.0732	**0.0059**	**0.0358**	0.0251	0.0471	0.0067	0.0074	2.1
	LSP (Ours)	0.0117	0.0451	0.0287	**0.0613**	0.0292	0.0403	**0.0231**	**0.0429**	**0.0046**	**0.0058**	**1.6**
SSIM ↑	U-Net	0.9764	0.9779	0.8981	0.8854	0.9557	0.8805	0.9496	0.8636	0.9691	0.9569	3.3
	FNO	**0.9876**	**0.9788**	0.9237	0.8900	0.9559	0.9283	0.9505	0.8185	0.9657	0.9457	2.8
	DeepONet	0.9588	0.9264	0.9456	0.9024	0.9574	0.9115	**0.9571**	0.9106	0.9719	0.9702	2.4
	LSP (Ours)	0.9607	0.9462	**0.9597**	**0.9229**	**0.9620**	**0.9485**	0.9561	**0.9420**	**0.9782**	**0.9762**	**1.5**

OOD Generalization on Complex Structures. Table 2 reports OOD results from models trained on the OpenFWI Style-B subset. Without fine-tuning, parameter updates, or test-time adaptation, LSP achieves the best MAE, MSE, and SSIM on original Marmousi and Overthrust fields. On the smoothed variants, the advantage is less uniform: FNO obtains the lowest MAE and MSE on Marmousi Smooth, while DeepONet achieves the highest SSIM. LSP performs best across all three metrics on Overthrust Smooth. These results suggest that LSP transfers effectively from Style-B to complex and high-contrast geological structures and remains robust across previously unseen velocity configurations.

Controlled 3D Extension. We evaluate LSP on the GlobalTomo Acoustic tier (Li et al., 2026), where a 405-dimensional structural representation is mapped to seven 3D wavefield snapshots and seismograms at 1,369 stations. As shown in Table 3, LSP achieves the lowest RL2 and the highest R for both outputs. Figure 5 shows coherent wavefield propagation and station-level waveforms. These results support the scalability of the coarse-to-refined latent formulation, although this experiment remains a controlled high-dimensional evaluation rather than field-scale 3D validation. Compared with the two-dimensional OpenFWI setting, the GlobalTomo experiment substantially increases both input-output dimensionality and spatial complexity. The consistent improvement in wavefield RL2, wavefield correlation, and seismogram correlation suggests that the coarse-to-refined latent representation is not restricted to the two-dimensional acquisition geometry used in the primary benchmark. At the same time, the remaining gap to exact numerical simulation indicates that further adaptation is required for more heterogeneous and field-scale three-dimensional settings. Full configurations are provided in Appendix F.

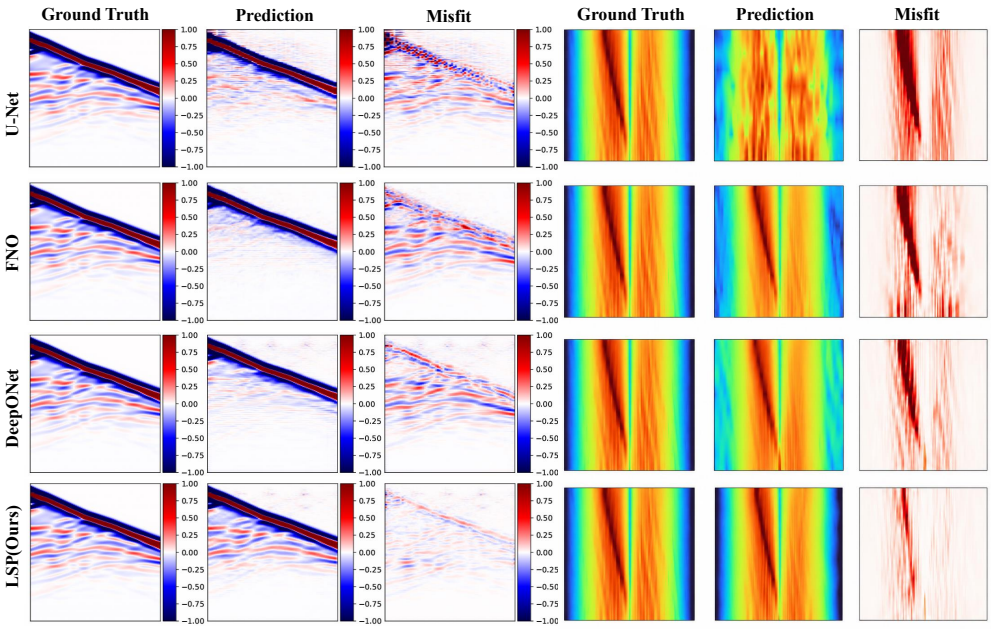


Figure 4: Qualitative comparison of seismic-response predictions on the OpenFWI STB subset in the space–time and frequency–wavenumber domains, including the ground truth, prediction, and corresponding misfit.

Table 2: OOD results on the Marmousi and Overthrust velocity fields using neural models trained only on Style-B. Mar.-S and Over.-S denote their smoothed variants, respectively. Best values are bold.

Metric	Model	Mar.	Over.	Mar.-S	Over.-S
MAE ↓	U-Net	0.2818	0.3018	0.2502	0.2311
	FNO	0.1484	0.2893	**0.1030**	0.2175
	DeepONet	0.1540	0.2792	0.1316	0.1865
	LSP	**0.1264**	**0.2061**	0.1043	**0.1308**
MSE ↓	U-Net	0.2216	0.5403	0.1074	0.4126
	FNO	0.1210	0.5593	**0.0813**	0.3433
	DeepONet	0.1151	0.4479	0.0954	0.3478
	LSP	**0.1128**	**0.3846**	0.0909	**0.3239**
SSIM ↑	U-Net	0.7344	0.7210	0.8712	0.7663
	FNO	0.8148	0.7447	0.9020	0.8426
	DeepONet	0.8195	0.7768	**0.9202**	0.8535
	LSP	**0.8354**	**0.8071**	0.9129	**0.8891**

4.3 Ablation Study

Table 4 evaluates residual replacement and the removal of spectral refinement, tapering, and symbolic guidance on STB. Replacing the harmonic residual with a convolutional block reduces SSIM from 0.9762 to 0.8957; adding symbolic guidance raises it to 0.9134, still well below the full model. Thus, the gain cannot be attributed to an arbitrary residual branch or regularization alone.

This also suggests that symbolic guidance is more effective when paired with a refinement mechanism capable of representing oscillatory waveform components. Removing spectral refinement causes the largest structural degradation, reducing SSIM to 0.9231, while removing tapering or

Table 3: Results on the GlobalTomo Acoustic benchmark. Best values are bold.

Wave	Model	Wavefield		Seismogram	
		RL2 ↓	R ↑	RL2 ↓	R ↑
Acoustic	U-Net	0.3947	0.8977	0.4460	0.8205
	FNO	0.3556	0.9174	0.5947	0.8359
	DeepONet	0.3685	0.9272	0.5034	0.8627
	LSP	**0.2909**	**0.9570**	**0.3765**	**0.9074**

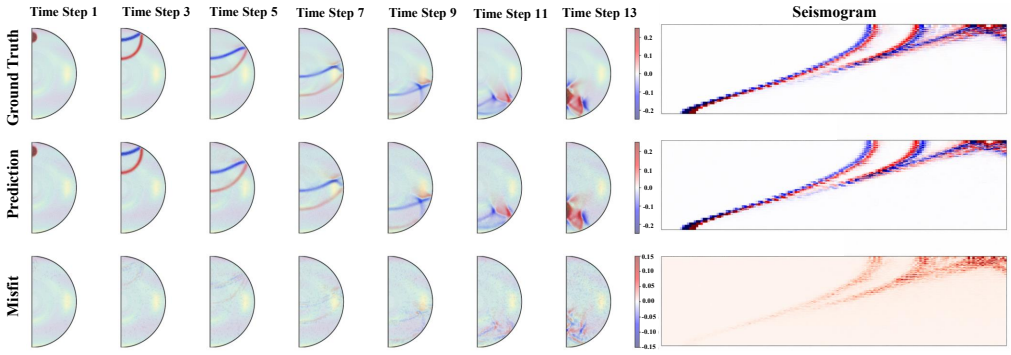


Figure 5: Qualitative comparison of LSP predictions on the GlobalTomo Acoustic benchmark, including three-dimensional acoustic wavefields and surface-station seismograms.

symbolic guidance yields 0.9492 and 0.9513, respectively. These smaller but consistent drops show that tapering and symbolic guidance provide complementary benefits beyond harmonic refinement. Tapering regulates basis contributions, whereas symbolic guidance encourages realistic spectral, attenuation, and amplitude trends.

Together, harmonic refinement recovers unresolved structures, while tapering and symbolic guidance improve stability and consistency, yielding the best overall performance.

Table 4: Ablation study of LSP components. The trend regularizer is used only during training.

Type	Design	Components			Efficiency			Performance		
		Residual	Taper	Symbolic	Params (M)	Memory (GB)	Time (s/epoch)	MAE↓	MSE↓	SSIM↑
Replacement	Conv residual	Conv	–	–	26.36	5.2174	73.2846	0.0425	0.0077	0.8957
	Conv residual + Symbolic	Conv	–	✓	26.36	6.0816	77.5997	0.0421	0.0079	0.9134
Removal	w/o Spectral residual	–	–	✓	25.79	4.8629	68.2187	0.0410	0.0083	0.9231
	w/o Taper	Harmonic	–	✓	26.07	7.7891	85.7063	0.0375	0.0081	0.9492
	w/o Symbolic loss	Harmonic	✓	–	26.14	7.7700	83.7243	0.0372	0.0077	0.9513
Full	LSP	Harmonic	✓	✓	26.14	8.4699	89.6507	0.0366	0.0058	0.9762

5 Conclusion

We presented LSP, a coarse-to-refined latent framework that combines coarse transport, tapered spectral residual refinement, and real-data-derived trend regularization for SFM. LSP achieves the best average rank across OpenFWI, with its clearest gains on the complex Fault and Style subsets, while introducing no additional inference cost. The Style-B-trained model also generalizes effectively to the original Marmousi and Overthrust fields without adaptation, and the GlobalTomo Acoustic results support its extension to high-dimensional 3D prediction. The main limitations are the transfer of empirical real-data trends through synthetic proxy variables and the controlled nature of the current 3D evaluation. Future work will investigate dataset-adaptive trend recalibration and more realistic heterogeneous 3D settings.

AI USE STATEMENT

Generative AI tools were used in this work to assist the offline screening of candidate symbolic relationships produced by Parallel Symbolic Enumeration, where the tool was prompted to check consistency with basic physical knowledge and expected variable relationships. Generative AI tools were also used to assist manuscript language editing and formatting. All AI-assisted outputs were reviewed by the authors; the retained symbolic relationships, experimental results, claims, and final manuscript content were checked by the authors. The authors take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

We provide the model formulation, training objective, evaluation protocol, and main implementation settings in the paper. Additional architectural details, data processing procedures, training configurations, symbolic-guidance construction, and the controlled 3D setup are documented in the appendix. Code and other supplementary material submitted for review are anonymized.

REFERENCES

Tariq Alkhalifah, Hanchen Wang, and Oleg Ovcharenko. MLReal: Bridging the gap between training on synthetic data and real data applications in machine learning. *Artificial Intelligence in Geosciences*, 3:101–114, 2022.

Fred Aminzadeh, N Burkhard, J Long, Tim Kunz, and P Duclos. Three dimensional SEG/EAEG models—an update. *The Leading Edge*, 15(2):131–134, 1996.

Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.

Shuhao Cao, Francesco Brarda, Ruipeng Li, and Yuanzhe Xi. Spectral-refiner: Accurate fine-tuning of spatiotemporal fourier neural operator for turbulent flows. In *International Conference on Learning Representations*, volume 2025, pp. 60348–60384, 2025.

Miles Cranmer, Alvaro Sanchez Gonzalez, Peter Battaglia, Rui Xu, Kyle Cranmer, David Spergel, and Shirley Ho. Discovering symbolic models from deep learning with inductive biases. *Advances in Neural Information Processing Systems*, 33:17429–17442, 2020.

Mark A Dablain. The application of high-order differencing to the scalar wave equation. *Geophysics*, 51(1):54–66, 1986.

Chengyuan Deng, Shihang Feng, Hanchen Wang, Xitong Zhang, Peng Jin, Yinan Feng, Qili Zeng, Yinpeng Chen, and Youzuo Lin. OpenFWI: Large-scale multi-structural benchmark datasets for full waveform inversion. *Advances in Neural Information Processing Systems*, 35:6007–6020, 2022.

Yiheng Du, Nithin Chalapathi, and Aditi Krishnapriyan. Neural spectral methods: Self-supervised learning in the spectral domain. In *International Conference on Learning Representations*, volume 2024, pp. 29197–29214, 2024.

Jufeng Han, Shu Wei, Min Wu, Lina Yu, Weijun Li, Linjun Sun, Hong Qin, and Yan Pang. Implicit neural representation with multi-scale sine activation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 21567–21575, 2026.

Xinquan Huang and Tariq Alkhalifah. Learned frequency-domain scattered wavefield solutions using neural operators. *Geophysical Journal International*, 241(3):1467–1478, 2025.

Adam Kania, Marko Mihajlovic, Sergey Prokudin, Jacek Tabor, and Przemysław Spurek. Fresh: Frequency shifting for accelerated neural representation learning. In *International Conference on Learning Representations*, volume 2025, pp. 76383–76406, 2025.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Dimitri Komatitsch, Jean-Pierre Vilotte, Rossana Vai, José M Castillo-Covarrubias, and Francisco J Sánchez-Sesma. The spectral element method for elastic wave equations—application to 2-D and 3-D seismic problems. *International Journal for Numerical Methods in Engineering*, 45(9): 1139–1164, 1999.

Noam Koren, Ralf JJ Mackenbach, Ruud JG van Sloun, Kira Radinsky, and Daniel Freedman. SVD-NO: Learning PDE solution operators with SVD integral kernels. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 525–533, 2026.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to PDEs. *Journal of Machine Learning Research*, 24(89):1–97, 2023.

Bian Li, Hanchen Wang, Shihang Feng, Xiu Yang, and Youzuo Lin. Solving seismic wave equations on variable velocity models with Fourier neural operator. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–18, 2023.

Shiqian Li, Zhi Li, Zhancun Mu, Shiji Xin, Zhixiang Dai, Kuangdai Leng, Rita Zhang, Xiaodong Song, and Yixin Zhu. GlobalTomo: A global dataset for physics-ML seismic wavefield modeling and FWI. *Advances in Neural Information Processing Systems*, 38, 2026.

Zongyi Li, Nikola Borislavov Kovachki, Kamyar Azizzadenesheli, Burigede liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=c8P9NQVtmnO.

Miguel Liu-Schiaffini, Julius Berner, Boris Bonev, Thorsten Kurth, Kamyar Azizzadenesheli, and Anima Anandkumar. Neural operators with localized integral and differential kernels. In *International Conference on Machine Learning*, pp. 32576–32594. PMLR, 2024.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021.

Ben Moseley, Tarje Nissen-Meyer, and Andrew Markham. Deep learning for fast simulation of seismic waves in complex media. *Solid Earth*, 11(4), 2020.

Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *International Conference on Machine Learning*, pp. 5301–5310. PMLR, 2019.

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

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241. Springer, 2015.

Kai Ruan, Yilong Xu, Ze-Feng Gao, Yang Liu, Yike Guo, Ji-Rong Wen, and Hao Sun. Discovering physical laws with parallel symbolic enumeration. *Nature Computational Science*, 6(1):53–66, 2026.

Michael Schmidt and Hod Lipson. Distilling free-form natural laws from experimental data. *Science*, 324(5923):81–85, 2009.

Ashutosh Singh, Ricardo Borsoi, Deniz Erdogmus, and Tales Imbiriba. Learning semilinear neural operators: A unified recursive framework for prediction and data assimilation. In *International Conference on Learning Representations*, volume 2024, pp. 44092–44109, 2024.

Albert Tarantola. Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49 (8):1259–1266, 1984.

Silviu-Marian Udrescu and Max Tegmark. AI Feynman: A physics-inspired method for symbolic regression. *Science Advances*, 6(16):eaay2631, 2020.

Roelof Versteeg. The Marmousi experience: Velocity model determination on a synthetic complex data set. *The Leading Edge*, 13(9):927–936, 1994.

Jean Virieux and Stéphane Operto. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–online, 2009.

Sifan Wang, Jacob Seidman, Shyam Sankaran, Hanwen Wang, George Pappas, and Paris Perdikaris. Cvit: Continuous vision transformer for operator learning. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 24004–24039, 2025.

Zipeng Xiao, Zhongkai Hao, Bokai Lin, Zhijie Deng, and Hang Su. Improved operator learning by orthogonal attention. In *International Conference on Machine Learning*, pp. 54288–54299. PMLR, 2024.

Yilong Xu, Yang Liu, and Hao Sun. Reinforcement symbolic regression machine. In *International Conference on Learning Representations*, volume 2024, pp. 54416–54440, 2024.

Zihan Yu, Jingtao Ding, Yong Li, and Depeng Jin. Symbolic regression via mdlformer-guided search: from minimizing prediction error to minimizing description length. In *International Conference on Learning Representations*, volume 2025, pp. 65306–65333, 2025.

A LATENT PROJECTION AND COARSE TRANSPORT

The latent prediction pathway follows the same coarse-to-refined organization illustrated in Fig. 3 of the main paper. The velocity field is first encoded into a compact velocity-domain representation, after which coarse latent transport produces a response-domain backbone. Response-token patchification is performed only after this coarse state has been obtained.

Local velocity feature extraction. Given an input velocity field $v \in \mathbb{R}^{1\times H\times W}$, the velocity encoder internally extracts multiple local feature views:

$$F_v = \mathrm{Concat}\left(f_v^{(1)}, f_v^{(2)}, f_v^{(3)}\right) \in \mathbb{R}^{P_v\times d}, \tag{A.1}$$

where $f_v^{(1)}$, $f_v^{(2)}$, and $f_v^{(3)}$ denote complementary representations of the local geological structures, P_v is the number of encoded local features, and d is the latent feature dimension. These views are internal features of the velocity encoder rather than response-domain tokens.

A set of learnable velocity-domain queries $Q_v \in \mathbb{R}^{C_v\times d}$ aggregates the local velocity features through cross-attention:

$$z_v = Q_v + \mathrm{softmax}\left(\frac{(Q_vW_Q)(F_vW_K)^\top}{\sqrt{d}}\right) \cdot F_vW_V. \tag{A.2}$$

where W_Q, W_K, and W_V are the query, key, and value projections. The learnable queries produce a fixed-size latent representation while aggregating information from smooth background structures, interfaces, faults, and localized velocity contrasts.

The resulting velocity latent is denoted by

$$z_v = E_v(v). \tag{A.3}$$

Although the encoder uses local structure extraction internally, no response-token patchification is performed at this stage.

Coarse latent transport. The compact velocity representation is subsequently mapped toward the response domain:

$$z_c = T_c(z_v), \tag{A.4}$$

where T_c denotes the coarse latent transport module. The transport blocks progressively mix the velocity-domain latent features and predict the dominant organization of the seismic response.

The coarse state z_c primarily represents large-scale response characteristics, including principal arrival geometry, broad reflection moveout, and source–receiver energy organization. It does not attempt to reconstruct every weak oscillatory component. Instead, it provides the response-domain backbone retained by the outer latent residual connection.

Response-oriented tokenization. The spectral residual refinement begins only after the coarse latent state has been constructed. Specifically, z_c is rearranged into response-oriented tokens:

$$P = \mathrm{Patchify}(z_c). \tag{A.5}$$

A set of learnable response queries Q_{learn} then aggregates the coarse response tokens:

$$h = \mathcal{A}_{\mathrm{attn}}\left(Q_{\mathrm{learn}}, P\right). \tag{A.6}$$

This attention operation produces a compact response feature h that identifies latent regions requiring further oscillatory or structural correction.

The complete simulation-driven inference pathway is therefore

$$\begin{aligned} v \to \{f_v^{(1)}, f_v^{(2)}, f_v^{(3)}\} \to z_v \to z_c \to P \to h \\ \to h^+ \to P^+ \to \Delta_{\mathrm{LSP}}(z_c) \to z_p \to \hat{p}. \end{aligned} \tag{A.7}$$

The operations from P to $\Delta_{\mathrm{LSP}}(z_c)$ constitute the spectral residual refinement pathway described in Appendix B.

B SPECTRAL RESIDUAL AND TAPER

After coarse latent transport, LSP constructs a structured correction for response components that are not fully represented by z_c. The complete refinement pathway consists of response-token patchification, query-based aggregation, tapered harmonic synthesis, decoder-side attention, unpatchification, and outer latent residual fusion.

As described in Appendix A, the coarse response latent is first converted into response-oriented tokens:

$$P = \mathrm{Patchify}(z_c), \qquad h = \mathcal{A}_{\mathrm{attn}}\left(Q_{\mathrm{learn}}, P\right). \tag{B.1}$$

Coefficient-wise spectral taper. The harmonic residual branch contains learnable sine and cosine coefficient families, denoted by $w_{\sin,k}$ and $w_{\cos,k}$. Following Fig. 3 of the main paper, the learnable taper acts on these coefficient families before harmonic residual synthesis:

$$\begin{aligned} \widetilde{w}_{\sin,k} &= W_{\mathrm{taper}}(k) \odot w_{\sin,k}, \\ \widetilde{w}_{\cos,k} &= W_{\mathrm{taper}}(k) \odot w_{\cos,k}. \end{aligned} \tag{B.2}$$

Thus, the taper is applied coefficient-wise rather than being multiplied with the complete synthesized residual afterward.

The tapered coefficient families are used to construct the latent harmonic correction:

$$\begin{aligned} \Delta h_{\mathrm{spec}} = w_0 + \sum_{k=1}^{K} \Big[& \widetilde{w}_{\sin,k} \odot \sin(\omega_k h) \\ & + \widetilde{w}_{\cos,k} \odot \cos(\omega_k h) \Big]. \end{aligned} \tag{B.3}$$

where K is the number of harmonic basis pairs, ω_k denotes the k-th latent basis scale, and w_0 is a learnable bias term.

The refined attention feature is obtained through an internal residual connection:

$$h^+ = h + \Delta h_{\mathrm{spec}}. \tag{B.4}$$

This internal connection preserves the response information selected from P while introducing a structured correction for weak reflections, diffraction patterns, interference structures, and coda-like variations.

The sine and cosine functions are applied element-wise to latent activations. Consequently, ω_k should not be interpreted as a physical temporal frequency or spatial wavenumber. The harmonic functions instead provide complementary nonlinear basis responses over the compact latent representation.

Cosine initialization of the taper. For $k = 1, \ldots, K$, the taper is initialized with a smooth cosine profile:

$$W_{\mathrm{taper}}^{(0)}(k) = \frac{1}{2}\left[1 - \cos\left(\frac{2\pi(k-1)}{K-1}\right)\right]. \tag{B.5}$$

The initialization provides a smooth modulation across harmonic orders at the beginning of training. After initialization, W_{taper} remains fully learnable and is optimized jointly with the other network parameters.

The taper does not operate on waveform time samples, spatial coordinates, physical Fourier coefficients, or the complete latent residual. Its only role is to modulate the learnable harmonic coefficient families before residual synthesis.

Decoder-side response-token reconstruction. The refined feature h^+ is returned to the response-token representation through decoder-side attention:

$$P^+ = \mathrm{DecAttn}\left(P, h^+\right). \tag{B.6}$$

Decoder-side attention preserves the response-token organization contained in P while injecting the harmonic correction encoded by h^+.

The refined token sequence is then restored to the latent layout:

$$\Delta_{\mathrm{LSP}}(z_c) = \mathrm{Unpatchify}(P^+). \tag{B.7}$$

Importantly, the output of the unpatchification operation is the structured correction $\Delta_{\mathrm{LSP}}(z_c)$ rather than the final prediction latent z_p.

Outer latent residual fusion. The structured correction is combined with the coarse response state through the outer latent residual connection:

$$z_p = z_c + \alpha \odot \Delta_{\mathrm{LSP}}(z_c), \tag{B.8}$$

where α is a learnable residual scale. This fusion retains the dominant propagation structure represented by z_c and controls the contribution of the response-oriented spectral correction.

The final waveform prediction is reconstructed by

$$\hat{p} = D_p(z_p). \tag{B.9}$$

All attention, harmonic synthesis, tapering, and residual fusion operations are performed in the latent prediction pathway. No explicit Fourier transform or waveform-domain post-processing is applied during inference.

C Symbolic Discovery and Regularization

STEAD observable extraction and quality control. We processed all available local-earthquake records in STEAD and extracted five quantities from each record: source magnitude M, source–receiver distance R, characteristic frequency f_c, attenuation coefficient γ, and amplitude A. For all waveform-derived quantities, the vertical component was used with a sampling interval of 0.01 s, and the waveform was detrended and mean-centered when required.

The characteristic frequency was estimated by fitting a Brune-type spectrum over 0.5–20 Hz to a window from the P arrival to 10 s after the S arrival, or to 10 s after the P arrival when the S arrival was unavailable. The attenuation coefficient was estimated after fourth-order 1–10 Hz Butterworth filtering by linearly fitting the logarithm of the Hilbert envelope from 3 to 15 s after the S arrival. The amplitude was defined as the peak absolute amplitude of the vertical component.

After extraction, task-specific quality control was applied. For the spectral task, records with invalid f_c or fitting $R^2 < 0.5$ were removed, and records satisfying either $M > 3$ with SNR > 10 dB or $M < 2$ with SNR > 30 dB were retained. For attenuation, we retained records with $0.001 < \gamma < 1$, $5 < R < 300$ km, and fitting $R^2 > 0.6$. For relative amplitude, records were restricted to $5 < R < 300$ km and SNR > 20 dB, and at most 100 records were retained per magnitude–distance bin using a fixed random seed of 42.

The resulting valid records define three symbolic regression tasks:

$$M \to \log_{10} f_c, \tag{C.1}$$

$$R \to \gamma, \tag{C.2}$$

and

$$(M, \log_{10} R) \to \log_{10} A. \tag{C.3}$$

PSE search and physics screening. PSE searched expressions constructed from addition, subtraction, multiplication, division, identity, sine, cosine, exponential, and logarithmic operators. The search used a fixed random seed of 0, at most 100 down-sampled observations per evaluation, a one-hour search budget, and returned the top 6 expressions according to fitting error and expression complexity. Pareto-dominated candidates were first removed. The remaining expressions were then checked against predefined variable-dependency, sign, finiteness, and monotonicity requirements. Qwen-Turbo, accessed through the DashScope API, was used only to provide an auxiliary plausibility assessment of these candidates. Final retention was determined offline from the PSE fitting error, expression complexity, Pareto status, and the predefined physical screening criteria; Qwen was not involved in model training, evaluation, or inference. The selected symbolic relationships are:

$$\log_{10} f_c = 1 - \frac{M}{8}, \tag{C.4}$$

$$\gamma(R) = \frac{113}{512} + \frac{1}{16R}, \tag{C.5}$$

$$\log_{10} A = M - \log_{10} R + 2 \tag{C.6}$$

The complete symbolic loss is evaluated from the final predicted waveform $\hat{p}$. It is activated only after the reconstruction warm-up stage and is removed during inference. Consequently, the real-data-derived symbolic pathway provides output-level training guidance without changing the simulation-driven inference flow from v to $\hat{p}$.

Transfer to synthetic seismic prediction. The discovered symbolic relationships are not treated as governing equations of OpenFWI. Instead, they are converted into soft training-time constraints that transfer only relative waveform trends.

For OpenFWI, the source-scale proxy is fixed as $M = 3$, and the source-receiver offset proxy is defined as:

$$R_{s,r} = |x_r^{rec} - x_s^{src}| + 1. \tag{C.7}$$

The spectral constraint is computed from:

$$S_{\text{pred}}(f) = \text{Mean}_{s,r} \left| \text{rFFT}_t(\hat{p}) \right|. \tag{C.8}$$

The predicted spectrum and symbolic target are converted into relative log-spectral representations before computing the spectral loss. The relative log-spectral normalization and receiver-wise centering operators are defined as

$$\begin{aligned} \mathcal{N}_f(S) &= \log_{10}(S + \epsilon) - \max_f \log_{10}(S + \epsilon), \\ \mathcal{C}_r(a)_{s,r} &= a_{s,r} - \text{Mean}_r(a_{s,r}). \end{aligned} \tag{C.9}$$

Here, $\epsilon = 10^{-8}$ is used as a numerical-stability constant in logarithmic, square-root, normalization, and division operations throughout the symbolic regularization terms. Following the implementation, the discrete frequency coordinates of the real-valued Fourier transform are

$$f_m = \frac{m}{N_t \Delta t}, \qquad m = 0, \ldots, \left\lfloor \frac{N_t}{2} \right\rfloor, \tag{C.10}$$

where $\Delta t = 0.001\,\text{s}$ for OpenFWI. The retained symbolic relationship determines the corner-frequency proxy

$$f_c(M) = 10^{\,1 - M/8}. \tag{C.11}$$

A Brune-type relative spectral profile is then constructed as

$$S_{\text{tar}}(f_m; M) = \frac{1}{1 + \left(\dfrac{f_m}{f_c(M) + \epsilon} \right)^2}, \tag{C.12}$$

and the normalized spectral template used by the source-spectrum loss is

$$\tau_f(f_m; M) = \mathcal{N}_f \left(S_{\text{tar}}(f_m; M) \right). \tag{C.13}$$

This construction removes absolute spectral energy and transfers only the relative spectral shape implied by the retained corner-frequency relation. The relative-amplitude observable is computed using temporal RMS:

$$A_{\mathrm{pred}}^{s,r} = \sqrt{\frac{1}{N_t}\sum_{t=1}^{N_t} \hat{p}_{s,t,r}^2 + \epsilon}, \tag{C.14}$$

where N_t denotes the number of temporal samples. Consistent with the implementation, the receiver-wise relative-amplitude template is

$$\tau_A(s,r) = \mathcal{C}_r \left[M - \log_{10}\left(R_{s,r} + \epsilon\right) + 2 \right]. \tag{C.15}$$

Both the predicted and target log-amplitudes are centered over receivers for each source, so τ_A transfers only the relative receiver-wise amplitude trend rather than an absolute amplitude scale.

Offset and temporal coordinates. Here, x_r^{rec} and x_s^{src} denote the receiver and source positions along the fixed acquisition line, respectively. The quantity $R_{s,r}$ defined in Eq. (C.7) serves as a relative source–receiver distance proxy for transferring the retained distance-dependent symbolic trends to OpenFWI. The additive constant prevents a zero distance and keeps both $1/R_{s,r}$ and $\log_{10} R_{s,r}$ well defined.

Consistent with the normalized temporal amplitude evolution introduced in the main paper, the temporal coordinate is defined as

$$t_n = \frac{n-1}{N_t - 1}, \qquad n = 1, \ldots, N_t, \tag{C.16}$$

where N_t denotes the number of temporal samples and $t_n \in [0, 1]$.

For attenuation modeling, we construct the normalized absolute-amplitude observable

$$B_{\mathrm{pred}}(t_n, s, r) = \frac{|\hat{p}_{s,n,r}|}{\max_{n'} |\hat{p}_{s,n',r}| + \epsilon}. \tag{C.17}$$

The corresponding symbolic attenuation template is

$$\tau_\gamma(t_n, s, r) = \exp\left[-\gamma(R_{s,r}) t_n\right]. \tag{C.18}$$

This template describes only the relative temporal decay tendency and does not prescribe individual arrivals, reflections, or scattering events.

Term-wise symbolic losses. Following the observable definitions in the main paper, the predicted response provides three differentiable observables:

$$\begin{aligned} o_f(\hat{p}) &= \mathcal{N}_f\left(S_{\mathrm{pred}}\right), \\ o_\gamma(\hat{p}) &= B_{\mathrm{pred}}, \\ o_A(\hat{p}) &= \mathcal{C}_r\left[\log_{10}\left(A_{\mathrm{pred}} + \epsilon\right)\right]. \end{aligned} \tag{C.19}$$

The three symbolic-loss terms are

$$\mathcal{L}_{\mathrm{src}} = \mathrm{MSE}\left(o_f(\hat{p}), \tau_f\right), \tag{C.20}$$

$$\mathcal{L}_{\mathrm{decay}} = \mathrm{MSE}\left(o_\gamma(\hat{p}), \tau_\gamma\right), \tag{C.21}$$

and

$$\mathcal{L}_{\mathrm{amp}} = \mathrm{MSE}\left(o_A(\hat{p}), \tau_A\right). \tag{C.22}$$

Here, τ_f, τ_γ, and τ_A denote the relative spectral, attenuation, and receiver-wise amplitude templates derived from the retained symbolic relationships, respectively.

The complete symbolic regularizer is

$$\mathcal{L}_{\mathrm{sym}} = w_{\mathrm{src}}\mathcal{L}_{\mathrm{src}} + w_{\mathrm{decay}}\mathcal{L}_{\mathrm{decay}} + w_{\mathrm{amp}}\mathcal{L}_{\mathrm{amp}}. \tag{C.23}$$

The symbolic loss is activated only after waveform reconstruction warm-up and is removed during inference. In the reported ablation, removing the symbolic loss primarily degrades MSE and SSIM, suggesting that these relative waveform trends provide complementary structural supervision.

D Dataset and Implementation Details

Datasets and evaluation protocol. The 2D experiments are conducted on OpenFWI, including ten subsets from the Vel, CurveVel, Fault, and Style families: FVA, FVB, CVA, CVB, FFA, FFB, CFA, CFB, STA, and STB. All experiments follow the official OpenFWI training and testing protocol. Velocity fields and seismic responses are normalized using statistics computed only from the training split. No information from validation or test samples is used during preprocessing.

The input velocity and seismic response have shapes $v \in \mathbb{R}^{1\times 70\times 70}$ and $p \in \mathbb{R}^{5\times 1000\times 70}$, respectively, where the output dimensions correspond to seismic sources, temporal samples, and receivers.

For OOD evaluation, models are trained on the OpenFWI Style-B subset and directly evaluated on the original and smoothed Marmousi and Overthrust velocity fields. No fine-tuning, parameter updating, or test-time adaptation is performed. All compared models use identical data preprocessing and acquisition configurations. This protocol isolates cross-structure generalization under a fixed preprocessing and acquisition setting.

Table D.1: Architecture configuration of LSP.

Configuration	Value
Input velocity shape	$1 \times 70 \times 70$
Output response shape	$5 \times 1000 \times 70$
Local structure window	5×5
Latent dimension	70
Velocity latent queries	4
Coarse transport blocks	4
Harmonic order K	16
Taper initialization	Cosine profile
Outer residual scale	Learnable

Optimization details. For the 2D OpenFWI experiments, all compared models were trained for 500 epochs using Adam with an initial learning rate of 10^{-3} and a global batch size of 50. During multi-GPU training, the batch was evenly distributed across the two GPUs, corresponding to 25 samples per GPU. The reconstruction loss is

$$\mathcal{L}_{\mathrm{rec}} = \beta_1 \,\mathrm{MAE}(\hat{p}, p) + \beta_2 \,\mathrm{MSE}(\hat{p}, p), \tag{D.1}$$

where $\beta_1 = \beta_2 = 1$.

For LSP, the complete optimization objective is:

$$\mathcal{L}_{total} = \mathcal{L}_{rec} + \lambda(e)\mathcal{L}_{sym}. \tag{D.2}$$

The symbolic regularizer is activated after a 100-epoch reconstruction warm-up. Its weight follows

$$\lambda(e) = \begin{cases} 0, & e < E_{\mathrm{warmup}}, \\ \lambda_{\max} \sin^2\big(\eta(e)\big), & e \geq E_{\mathrm{warmup}}, \end{cases} \tag{D.3}$$

where

$$\eta(e) = \frac{\pi}{2} \frac{e - E_{\mathrm{warmup}}}{E_{\mathrm{total}} - E_{\mathrm{warmup}}}.$$

The complete optimization settings and symbolic-term weights are listed in Table D.2.

Computing environment. All experiments were conducted on a Linux server running Ubuntu 20.04.6 LTS, equipped with an Intel Xeon Silver 4210R CPU, 187 GiB of system memory, and two NVIDIA GeForce RTX 3090 GPUs with 24 GB of memory each. Multi-GPU training was implemented with Python 3.10, PyTorch 2.10, CUDA 12.8, and cuDNN 9.10.2.

Table D.2: Training configuration of LSP.

Configuration	**Value**
Optimizer	Adam
Learning rate	10^{-3}
Global batch size	50
Epochs	500
MAE weight	1.0
MSE weight	1.0
Warm-up epochs	100
Maximum symbolic weight	10^{-4}
Source-spectrum weight	1.0
Amplitude proxy weight	0.3
Decay proxy weight	0.3

Table E.1: Detailed configurations of the LSP ablation variants.

Variant	Modification
Conv residual	Preserve response-token patchification, query aggregation, decoder-side attention, unpatchification, and outer latent fusion. Replace the sine–cosine harmonic synthesis with a convolutional residual transformation. Disable both the coefficient taper and symbolic regularization.
Conv residual + Symbolic	Use the same convolutional replacement while retaining the complete symbolic regularization objective. Response-token aggregation, reconstruction, and outer latent fusion remain unchanged.
w/o Spectral residual	Remove the complete response-oriented refinement pathway, including Patchify, response-query attention, harmonic synthesis, coefficient tapering, decoder-side attention, and Unpatchify. The coarse latent is directly used as the prediction latent, i.e., $z_p = z_c$.
w/o Taper	Retain Patchify, response-query attention, sine–cosine harmonic synthesis, decoder-side attention, unpatchification, and outer latent fusion, but remove coefficient-wise modulation by setting $W_{\mathrm{taper}}(k) = 1$ for all harmonic orders.
w/o Symbolic loss	Retain the complete simulation-driven neural architecture, including coefficient-wise tapering and both residual connections, but optimize the model using only $\mathcal{L}_{\mathrm{rec}}$ by setting $\lambda(e) = 0$ throughout training.
Full LSP	Use the complete pathway: $z_c \to P \to h \to h^+ \to P^+ \to \Delta_{\mathrm{LSP}}(z_c) \to z_p$, together with coefficient-wise tapering and soft symbolic regularization.

E Ablation Configuration Details

The ablation study isolates the contributions of response-oriented harmonic refinement, coefficient-wise latent modulation, and real-data-derived symbolic guidance. All variants use the same velocity encoder E_v, coarse transport module T_c, waveform decoder D_p, data split, optimization schedule, and training budget. Except for the component modified in Table E.1, the remaining prediction pathway is unchanged. The convolutional variants test whether a generic residual transformation provides comparable benefits, whereas the removal variants separately isolate harmonic refinement, coefficient-wise tapering, and symbolic guidance. This design separates gains from the residual representation from those due to additional capacity or training-only regularization, ensuring performance differences reflect the modified component rather than the experimental protocol.

F GLOBALTOMO ACOUSTIC 3D EXTENSION

Dataset and evaluation protocol. We evaluate the three-dimensional extension of LSP on the official Acoustic tier of GlobalTomo. The 10,000 acoustic simulations are randomly divided into 80% for training and 20% for evaluation using a fixed random seed of 42. The split is performed at the simulation level, so the structures in the evaluation subset are not included in training. The split is performed according to velocity structures so that all evaluation samples correspond to geological configurations unseen during training. The same split, input representation, output coordinates, and training budget are used for all compared models.

Table F.1: Configuration of the GlobalTomo Acoustic-tier experiment.

Configuration	Value
Dataset tier	GlobalTomo Acoustic
Number of simulations	10,000
Physical domain	1-km-radius fluid sphere
Simulation interval	0–3 s
Structural representation	405 parameters
Source configuration	Fixed
Training/evaluation	80% / 20% by structure
Evaluated wavefield steps	1, 3, 5, 7, 9, 11, 13
Slices per wavefield step	16
Points per slice	3,648
Wavefield output size	$7 \times 16 \times 3{,}648$
Seismogram length	150
Number of stations	1,369
Seismogram output size	$150 \times 1{,}369$
Evaluation metrics	RL2 and Pearson R

Wavefield and seismogram targets. The surface-seismogram target records acoustic displacement at 1,369 predefined surface stations. Each station contains a time series of length 150, giving

$$d \in \mathbb{R}^{150\times 1369},$$

or 205,350 scalar values per simulation. Wavefield and seismogram prediction are evaluated as two complementary forward-modeling outputs: the former measures reconstruction of the spatial propagation process, while the latter measures the accuracy of observable surface responses.

Three-dimensional adaptation of LSP. The 405-dimensional structural representation is embedded into a fixed-size latent input. LSP then retains the same coarse-to-refined organization defined in Appendices A and B, including coarse transport, response-token aggregation, tapered harmonic refinement, decoder-side attention, unpatchification, and outer latent residual fusion. Only the input and output projections are adapted to the GlobalTomo targets. The wavefield and surface-response outputs have shapes $\widehat{\Phi} \in \mathbb{R}^{7\times 16\times 3648}$ and $\widehat{d} \in \mathbb{R}^{150\times 1369}$, respectively. The same target dimensions are used for all compared baselines.

Training and metrics. All compared models are independently trained under the same data split and optimization budget. For the GlobalTomo experiments, we use AdamW with an initial learning rate of 3×10^{-4} and train each model for 500 epochs. Metrics are reported using the final checkpoint on the fixed 20% evaluation split.

Performance is measured using RL2 and the Pearson correlation coefficient R. Lower RL2 and higher R indicate better performance. Both metrics are computed separately for the complete wavefield output and the surface-station seismogram output and then averaged over the evaluation samples.

When surface-station waveforms are used during LSP training, the source-spectrum, relative attenuation, and relative-amplitude constraints are evaluated on the predicted station traces after the reconstruction warm-up stage. These auxiliary symbolic constraints are disabled during inference. No symbolic processing or station-level post-processing is applied after neural prediction. The latent spectral residual refinement remains part of the LSP neural inference pathway.

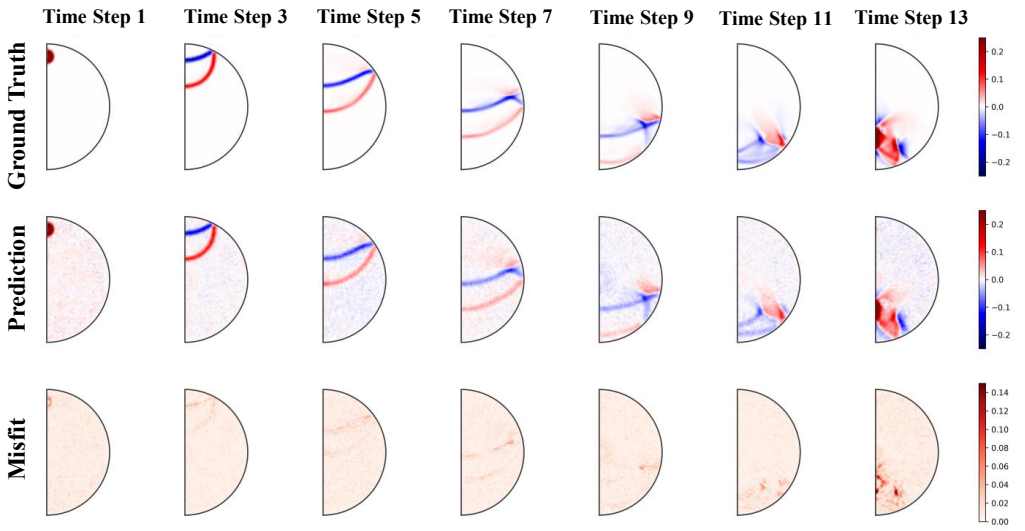


Figure F.1: Qualitative wavefield prediction on the GlobalTomo Acoustic benchmark at the seven evaluated timesteps. Columns correspond to time-step indices $1, 3, 5, 7, 9, 11$, and 13. The first and second rows show the ground-truth and LSP-predicted wavefields, respectively, while the third row shows the pointwise absolute prediction error.

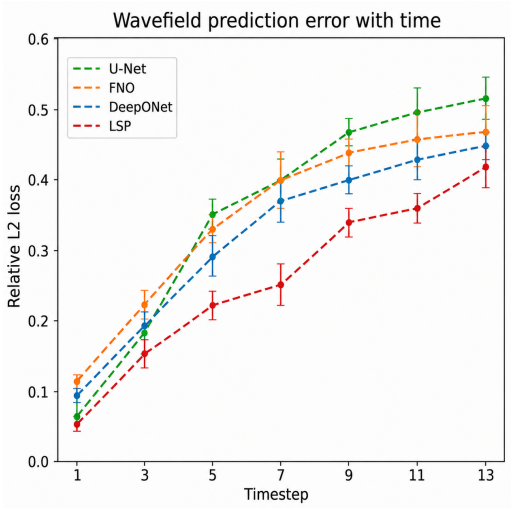


Figure F.2: Relative L_2 error at the seven evaluated wavefield timesteps on the GlobalTomo Acoustic benchmark. Curves show the mean over the evaluation samples, and error bars indicate one standard deviation.

This experiment provides a controlled evaluation of LSP for high-dimensional 3D acoustic-wave prediction rather than field-scale seismic simulation.

G Additional Discussion and Limitations

The symbolic relationships extracted from STEAD are transferred to OpenFWI through relative source-scale, source–receiver distance, spectral, attenuation, and amplitude proxies rather than being treated as governing equations. Their effectiveness may therefore depend on cross-domain similarity. Moreover, latent harmonic refinement is a learned nonlinear basis expansion rather than a phys-

ical Fourier solver, and the current GlobalTomo experiment remains a controlled high-dimensional evaluation rather than field-scale validation.

H ADDITIONAL TRAINING DYNAMICS

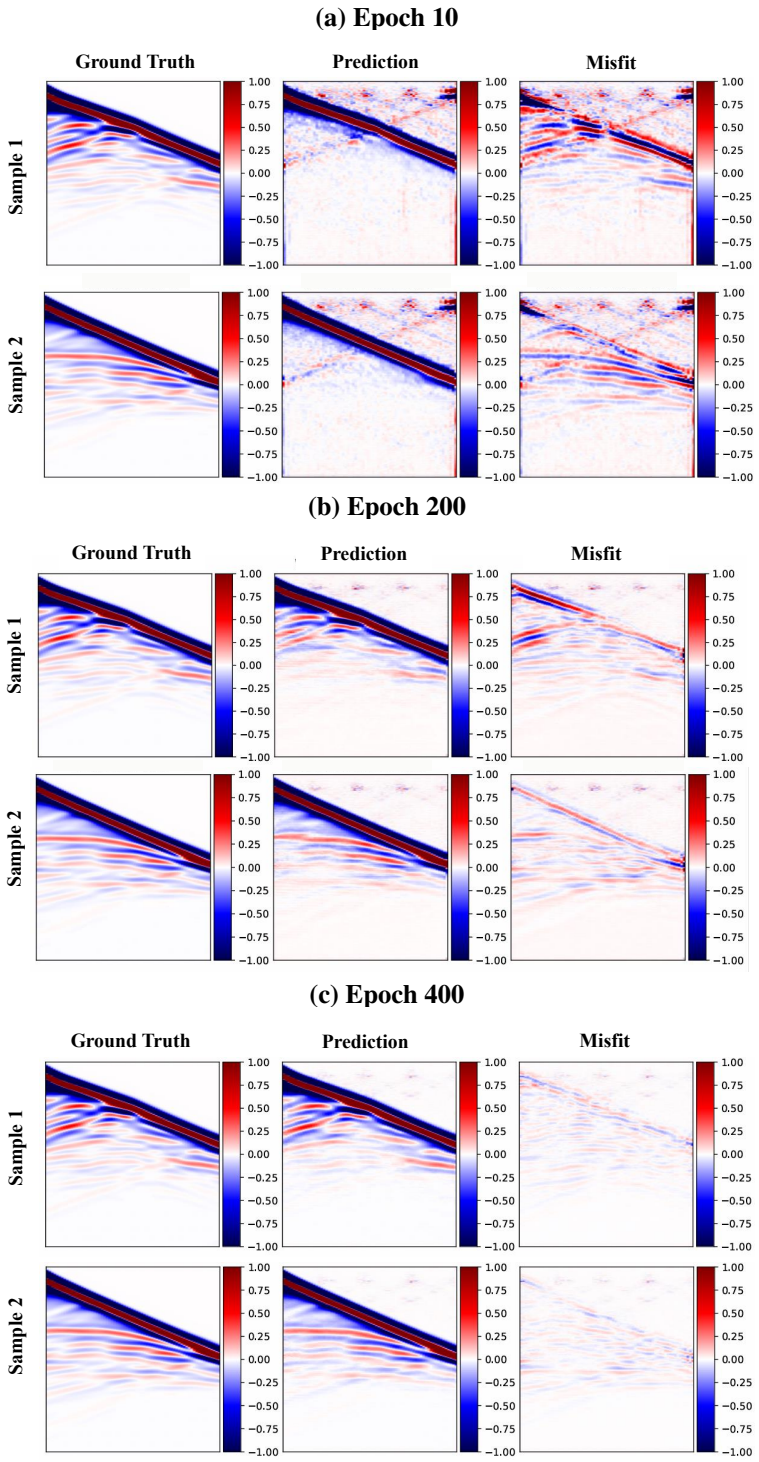


Figure H.1: Reconstruction dynamics on two held-out samples at epochs 10, 200, and 400. Each panel shows the ground truth, prediction, and signed misfit $p - \hat{p}$.