LEARNING SEISMIC WAVE PROPAGATION: FROM LOCAL GEOMETRY TO GLOBAL PHYSICS

Anonymous authors
Paper under double-blind review

ABSTRACT

Seismic forward modeling is computationally expensive because a wave equation must be solved repeatedly for each source and, possibly, several velocity models. Learned surrogates could reduce this cost, but they must preserve both the local influence of heterogeneous geologic structures and the long-range coherence of propagating wavefields. We address this problem using such machine learning models as Fourier Neural Operators (FNOs), MeshGraphNets (MGNs), and Transolver. FNO provides global spectral mixing on a structured grid; MGN explicitly uses local mesh connectivity; and Transolver communicates through learned global physical states. Here, we introduce SeisGeoSolver, which composes local message passing, global physics attention, and local refinement in a shared latent state. In 2D simulations for the GeoFWI acoustic model, edge-aware models are substantially more stable in autoregressive rollout than FNO and global-only physics attention. SeisGeoSolver matches the rollout accuracy of a deeper MGN while using fewer parameters and less training time. These results confirm that explicit local geometry is essential in this setting, while global physical interaction can replace part of the local processing budget.

1 INTRODUCTION

Seismic imaging and full-waveform inversion (FWI) require multiple numerical solutions of the wave equation. This work is a step toward a foundation-model solver for seismic wave propagation (Choi et al., 2025; Ashton et al., 2025), where a model is trained on broad and diverse simulation data that can produce high-resolution, accurate wavefields across different geologic settings. Such a program requires a surrogate that is inexpensive enough to train and evaluate at scale, yet sufficiently robust to serve as a reusable solver component. Spectral-element and finite-difference solvers remain accurate reference tools (Komatitsch & Vilotte, 1998), but their repeated use makes both large-scale training-data generation and downstream modeling expensive.

This motivates data-driven surrogates in addition to physics-informed neural networks (PINNs). PINNs can encode the governing partial differential equations (PDEs) directly, but residual training for wave equations requires repeated automatic differentiation and can become costly and difficult to scale for high-dimensional, high-order problems (He et al., 2022; Rasht-Behesht et al., 2021; Sethi et al., 2023). Instead, data-driven surrogate converts high-fidelity simulations into a fast learned time-stepper. The main challenge is then representational because a seismic wavefield evolves through *local* interactions—mesh connectivity, material contrasts, boundaries, and source neighbourhoods—but its phase and amplitude must remain coherent over the *entire* domain. Learning with graph connectivity and physics-aware attention provides a route to capturing these complementary structures from data. FNOs provide efficient global spectral mixing on regular grids (Li et al., 2021; Yang et al., 2021; 2023); MeshGraphNets (MGNs) operate on meshes and encode local geometric interactions with message passing (Pfaff et al., 2021; Sanchez-Gonzalez et al., 2020; Nabian et al., 2024); and Transolver uses learned physical states to make global interactions practical for general geometries (Wu et al., 2024; Luo et al., 2025). None of these representations, however, explicitly combines edge-conditioned local dynamics with global state-based communication.

The question is whether local geometric and global physical representations can be combined to develop a more efficient seismic wave-propagation surrogate? Our hypothesis is not that one representation supersedes the others, but that all of them encode complementary information. We test

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.

this hypothesis with SeisGeoSolver, a sequential hybrid in which local message passing first encodes mesh and material relations, then physics attention communicates globally, and the final local stage restores edge-conditioned consistency.

Our contributions are:

- Controlled comparison of grid-spectral, local mesh, and global physical-state representations for autoregressive acoustic wave propagation.
- SeisGeoSolver, a shared-latent local $\rightarrow$ global $\rightarrow$ local architecture that combines MGN message passing with Transolver-style physics attention.
- Evaluation of wavefield fidelity, autoregressive stability, computational cost, and validation on SEG salt patches.
- Proof that local edge information plays a key role for heterogeneous media, whereas global attention can reduce the depth required for local processing

The main text presents this argument and its evidence first. Equations, data-generation details, complete diagnostics, and secondary architectural analyses are relegated to the appendix.

2 Related Work

Neural operators for PDE surrogates. Neural operators learn mappings between fields rather than a solution for one fixed discretization. FNO performs global convolution in the Fourier space and provides a efficient baseline on regular grids (Li et al., 2021; Yang et al., 2021; 2023). Its grid-based spectral representation, however, does not directly expose mesh connectivity or edge-wise material relations.

Geometric physical simulation. Graph networks represent a discretized physical system through nodes and edges. MeshGraphNet couples neighboring mesh entities by learned message passing, providing a natural local geometric bias and support for irregular meshes (Pfaff et al., 2021; Sanchez-Gonzalez et al., 2020; Nabian et al., 2024; Iparraguirre et al., 2026). The cost of this locality is that distant communication requires many graph hops.

Global operators for general geometries. Transolver avoids all-pairs attention by pooling points into learned physical states, attending among those states, and projecting the result back to points (Wu et al., 2024; Luo et al., 2025). Recent geometry-aware extensions similarly emphasize persistent geometric context and multiscale interaction on irregular domains (Adams et al., 2026). These approaches motivate global state-based communication, but do not by themselves make graph-edge quantities available to the attention processor.

Seismic surrogate modeling and gap. Neural operators have been explored for seismic modeling and inversion, mostly on structured fields (Yang et al., 2021; 2023; Moseley et al., 2020; Zou et al., 2026). A foundation-model solver for seismic propagation (Choi et al., 2025; Ashton et al., 2025) requires a surrogate that is robust beyond a prescribed short time range because it must preserve wavefields when repeatedly queried as a time-stepper. Therefore, we train and evaluate all models autoregressively rather than treating wave propagation as a fixed-time field-regression task (Brandstetter et al., 2023). For heterogeneous media, the question to be addressed is not simply which standalone network scores highest, but whether explicit local connectivity and learned global physical interaction should coexist. We answer that question by comparing FNO, Transolver, MGN, and SeisGeoSolver under one seismic time-stepping protocol.

3 Problem Formulation

We model two-dimensional wave propagation in a heterogeneous acoustic medium. The reference simulation operate with pressure $p(\mathbf{x}, t)$, density $\rho(\mathbf{x})$, velocity $v_p(\mathbf{x})$, and source s that satisfy the following equation:

$$\rho(\mathbf{x})^{-1} v_p(\mathbf{x})^{-2} \partial_{tt} p - \nabla \cdot \left(\rho(\mathbf{x})^{-1} \nabla p\right) = s(\mathbf{x}, t). \tag{1}$$

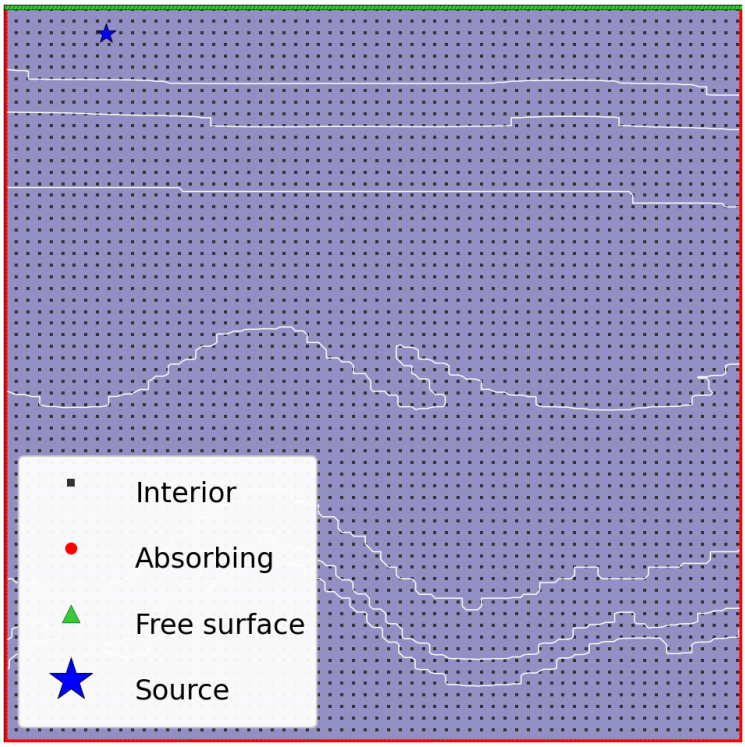

Figure 1: Mesh-graph node types on a GeoFWI salt sections include interior, absorbing (CPML) boundary, free surface, and source nodes.

The displacement wavefields $\mathbf{u}(\mathbf{x}, t)$ are generated by SPECFEM2D (Komatitsch & Vilotte, 1998); the network is a data-driven time-stepper.

After resampling, the domain in Fig. 1 is represented by nodes V at positions $\mathbf{x}_i$ and, for graph models, edges E between local neighbors. At time t_k, each node receives the current displacement and temporal history together with velocity, density-derived quantities, boundary masks, source information, coordinates, and time (Fig. 1). Graph models additionally receive relative position, distance, local traveltime, and velocity-contrast edge features. The exact feature specification is given in Appendix A.

The surrogate predicts a bundled forward displacement velocity,

$$f_\theta(\mathbf{q}(t_k), \mathbf{e}, \mathcal{G}) = \widehat{\dot{\mathbf{u}}}(t_k), \qquad \widehat{\mathbf{u}}_{k+1} = \widehat{\mathbf{u}}_k + \Delta t\, \widehat{\dot{\mathbf{u}}}_k, \tag{2}$$

where $f_\theta(\cdot)$ is the learned surrogate neural operator, $\mathbf{q}(t_k)$ are node features, $\mathbf{e}$ are edge features, $\mathcal{G}$ is the computational graph, $\widehat{\dot{\mathbf{u}}}_k$ is the predicted forward velocity increment at time t_k for all nodes, $\widehat{\mathbf{u}}_k$ is the displacement field at time t_k, Δt is the simulation time step size, and $\widehat{\mathbf{u}}_{k+1}$ is the predicted displacement field at time t_{k+1}.

The surrogate is rolled out autoregressively. Training supervises velocity increments and their integrated displacement fields, with a phase-sensitive spectral term (Brandstetter et al., 2023; Sanchez-Gonzalez et al., 2020). Thus, the primary outcome is rollout accuracy rather than one-step regression error alone.

4 Representation Baselines

The baselines solve the same autoregressive mapping in Eq. (2) but differ in how the spatial field exchanges information. This controlled progression allows the experiment to test an architectural hypothesis rather than merely rank unrelated models.

Table 1: Compared representations.

Model	Primary interaction	Main limitation for this study
FNO	global Fourier modes on a grid	no explicit mesh edges
Transolver	global learned physical states	no edge-conditioned messages
MGN	local graph message passing	long-range coupling needs depth
SeisGeoSolver	local + global + local	tested only for 2D acoustic media

FNO. FNO lifts the regular-grid field, alternates pointwise maps with truncated spectral convolutions, and decodes future velocity increments (Li et al., 2021; Yang et al., 2021; 2023). Its global receptive field makes it an important structured-grid reference, but neither connectivity nor material contrast along an edge is explicit.

MeshGraphNet. MGN encodes node and edge features, repeatedly updates edge messages and node states, and decodes the nodewise increment (Pfaff et al., 2021; Sanchez-Gonzalez et al., 2020). One message-passing block has the form

$$\mathbf{e}_{ij}^{\ell+1} = \phi_e(\mathbf{e}_{ij}^{\ell}, \mathbf{h}_i^{\ell}, \mathbf{h}_j^{\ell}), \quad \mathbf{h}_i^{\ell+1} = \phi_v\left(\mathbf{h}_i^{\ell}, \sum_{j\in\mathcal{N}(i)} \mathbf{e}_{ij}^{\ell+1}\right). \tag{3}$$

where $\mathbf{e}_{ij}^{\ell}$ is the edge feature at the l-th message-passing block, $\mathbf{h}_i^{\ell}$ is the node feature at the l-th block, and $\mathcal{N}(i)$ is the neighborhood of node i. This directly represents local geometric and material relationships, but a distant node affects $\mathbf{h}_i$ only after a sequence of graph hops.

Transolver. Transolver maps point features to a small set of learned physical states, attends among those states, and maps the result back to points (Wu et al., 2024; Luo et al., 2025). Its slice–attention–deslice mechanism provides each block with global communication without quadratic attention over all spatial nodes. In contrast to MGN, it receives no edge-level traveltime or velocity-contrast features.

These baselines motivate the development of SeisGeoSolver: local graph processing supplies what global physical states omit, while physics attention supplies domain-wide communication without a deeper graph stack. The baseline equations and configurations are described in Appendix B.

5 SeisGeoSolver: Local Geometry with Global Physics

SeisGeoSolver implements the paper's hypothesis using one shared latent representation:

$$\boxed{\text{local geometry} \longrightarrow \text{global physics} \longrightarrow \text{local geometry.}} \tag{4}$$

A local context stage encodes mesh connectivity, relative geometry, and material contrast into node states. A Transolver-style core then communicates among learned physical states (Wu et al., 2024; Luo et al., 2025; Adams et al., 2026). A local refinement stage uses the preserved edge representation to convert the globally informed state into an edge-consistent prediction. The three stages act successively on one latent state.

Let $(\mathbf{h}^0, \mathbf{e}^0)$ be node and edge encodings. The processor is

$$(\mathbf{h}^{\text{out}}, \mathbf{e}^{\text{out}}) = \mathrm{MP}_{\text{post}}^{P}\left(\mathrm{PA}^{L_a}\left(\mathrm{MP}_{\text{pre}}^{P}(\mathbf{h}^0, \mathbf{e}^0)\right), \mathbf{e}^{P}\right), \tag{5}$$

where MP is MGN message passing (Pfaff et al., 2021) and PA is physics attention (Wu et al., 2024). Only the local stages update edge states. The default model uses three local blocks before and after eight global attention blocks, then decodes four forward velocity increments.

This ordering has a physical interpretation. Before global mixing, edge features make the latent state sensitive to local traveltime and impedance contrast. Global mixing can then coordinate spatially separated parts of the field. After mixing, message passing reconnects that global context to the local mesh before the next increment is predicted.

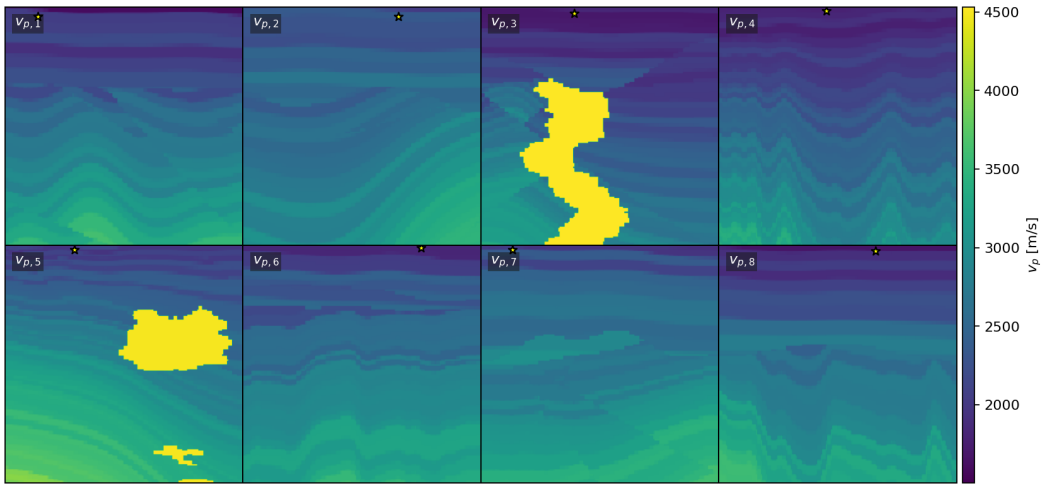

Figure 2: P-velocity sections from different GeoFWI cubes on a shared color scale.

Table 2: Primary model configurations.

Model	Processor	Parameters	Best epoch
FNO	4 Fourier layers, 16 modes	9.45M	220
Transolver	12 blocks, 256 states	8.97M	250
MGN	15 message-passing blocks	9.04M	200
SeisGeoSolver	3 MP + 8 PA + 3 MP	5.38M	260

SeisGeoSolver is trained and evaluated autoregressively using Eq. (2). Therefore, the comparison with MGN asks whether global state-based communication can substitute for a portion of a deep local stack. The comparison with Transolver asks whether global interaction becomes useful once it is grounded in explicit local geometry. The processor equations and description of rollout training (Brandstetter et al., 2023) and complexity can be found in Appendix B.

6 Experimental Setup

We evaluate all models as autoregressive surrogates for wavefields generated in the 2D acoustic GeoFWI salt model with SPECFEM2D (Swaminadhan et al., 2026; Komatitsch & Vilotte, 1998). The data are split in 3D velocity cubes, so held-out sections come from unseen geologic structures rather than adjacent slices of training models (Fig. 2). Each model receives the same physical state, source, material, and boundary information wherever its native representation permits, e.g., only MGN and SeisGeoSolver use graph edges.

The primary data set contains 2,000 simulations from 1,000 GeoFWI salt cubes. We use 1,800/100/100 train/validation/test sections, a 96×96 training grid, and 156 autoregressive steps at $\Delta t = 6.75$ ms. Training predicts bundles of four velocity increments and unrolls two bundles (Brandstetter et al., 2023). The complete source, preprocessing, and solver protocol is in Appendix C.

Our headline metric is the space–time relative L^2 error of the integrated displacement over a full rollout. We also measure active-frame energy ratio, qualitative wavefield, shot gathers, and wiggle-trace fidelity, validation on SEG salt patches, parameter count, and complete-rollout time. Checkpoints are selected by validation rollout rather than one-step error. Appendix C.3 defines every metric and lists the training and hardware settings.

Table 3: Autoregressive performance on three unseen GeoFWI test simulations. Errors are space–time relative L^2 (%) for u_x/u_z and E is the active-frame energy ratio.

Model	Rollout error	E
FNO	100.5/98.2	0.65
Transolver	43.2/33.8	0.94
MGN	**11.2**/11.9	0.98
SeisGeoSolver	11.8/**11.3**	0.99

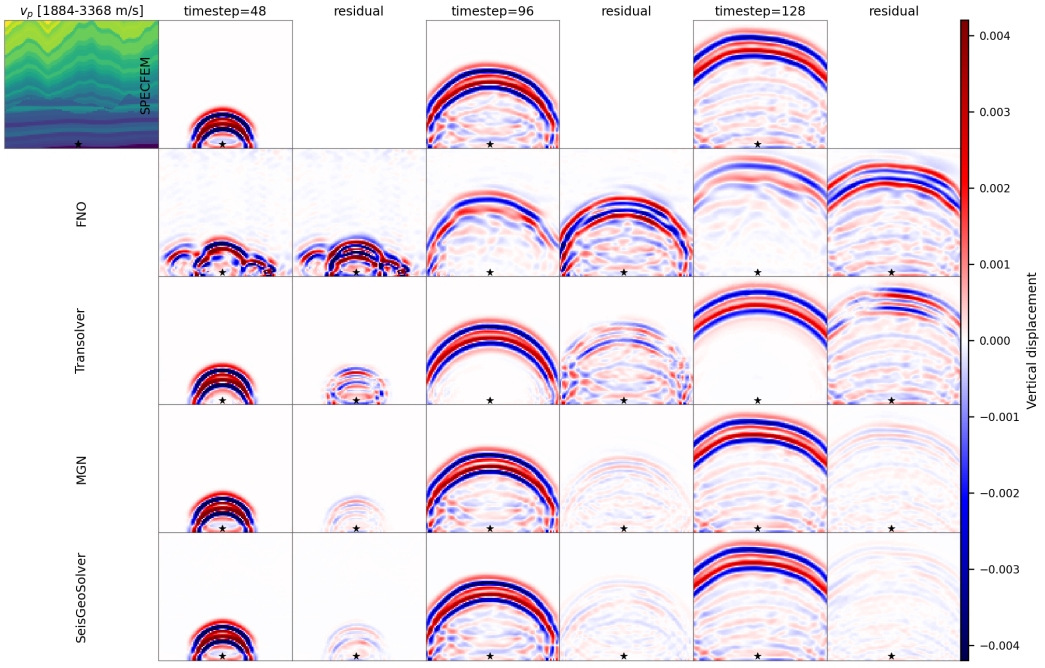

Figure 3: In-distribution rollout on GeoFWI test simulation (sim 3, u_z). The rows after SPECFEM are FNO, Transolver, MGN, and SeisGeoSolver. The residual columns use the same colour scale as the wavefield. Edge-aware models keep the reflected wavefront; FNO collapses and Transolver smears the salt-side reflection.

7 Results

Models with explicit local edge processing remain accurate under rollout, whereas FNO loses amplitude and physics attention drifts in phase. On three GeoFWI sections (Swaminadhan et al., 2026), SeisGeoSolver reaches the same $\approx$ 11–12% rollout-error band as a substantially deeper MGN, while using 40% fewer parameters. Figure 3 shows the test simulation (sim 3); sims 1 and 4 are in Appendix D.

Local geometry determines stable propagation. MGN and SeisGeoSolver retain near-unit energy and sharp wavefronts (Figs. 3–5). Their residuals are concentrated near high-contrast interfaces and the free surface. By contrast, Transolver retains energy but accumulates the spatial and phase errors, showing that amplitude preservation is not sufficient for robust modeling. In particular, the reflected and refracted wavefronts around strong velocity contrasts are not reliably reproduced by the global-only model. FNO approaches 100% rollout error, reinforcing that one-step regression is an inadequate proxy for a wave propagator.

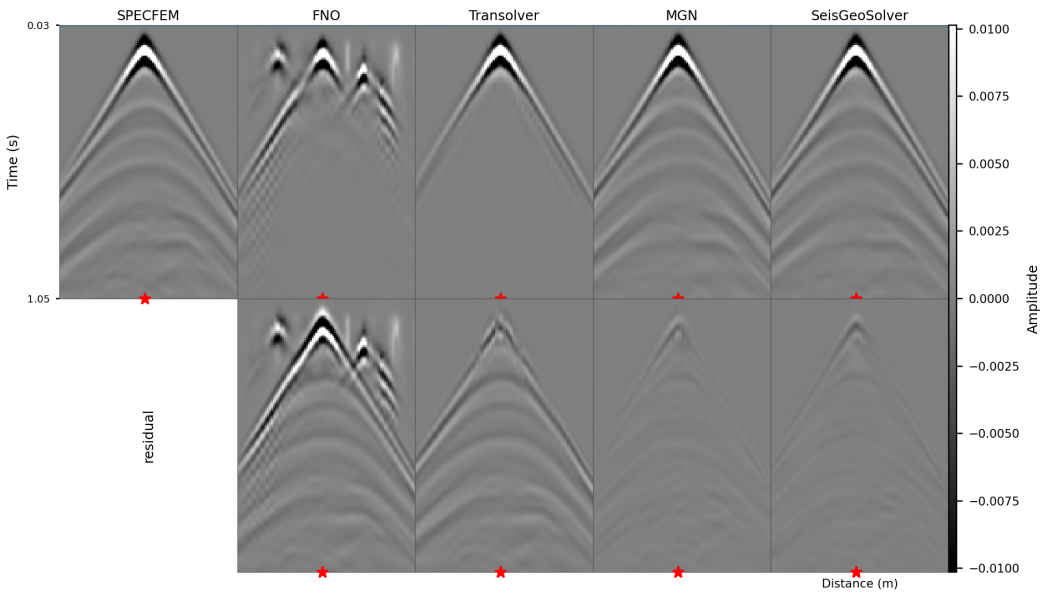

Figure 4: Surface gathers for the same test simulation (sim 3) rollout. Top row: the predicted u_z traces; bottom row: the residuals relative to SPECFEM. MGN and SeisGeoSolver preserve moveout, whereas FNO and Transolver do not.

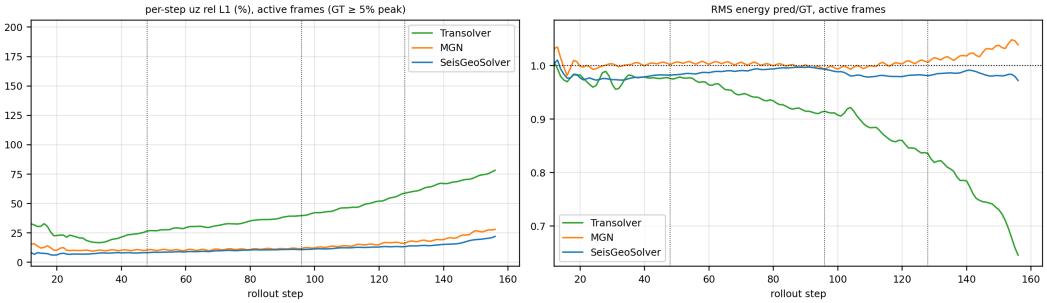

Figure 5: Per-step relative error in the vertical displacement u_z and energy ratio on test simulation (sim 3).

Validation on SEG salt patches remains difficult. On four unseen SEG 2D salt windows, edge-aware models remain substantially better than FNO and Transolver, but errors increase with geologic shift. The results can be found in the Appendix D.

The hybrid provides an efficiency gain but not yet a higher accuracy than the deep MGN. SeisGeoSolver is not more accurate than the 15-block MGN on the current in-distribution subset. However, its six local blocks plus eight global blocks reproduce the MGN accuracy regime with 5.38M rather than 9.04M parameters and 534 rather than 759 seconds per training epoch (Fig. 6). A 156-step SeisGeoSolver rollout takes 1.94 seconds on one A100, compared with 39.2 ± 3.2 seconds for one SPECFEM2D simulation (Komatitsch & Vilotte, 1998) (Fig. 7).

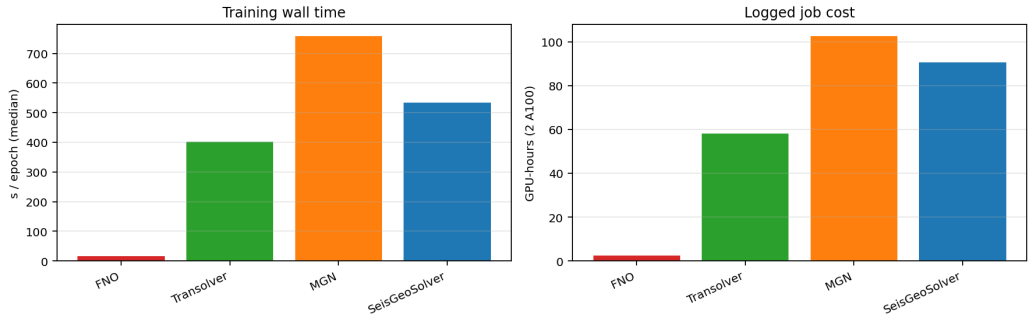

Figure 6: Training cost for the GeoFWI salt protocol (two A100s). SeisGeoSolver is cheaper per epoch than the 15-block MGN and uses fewer parameters (Table 2).

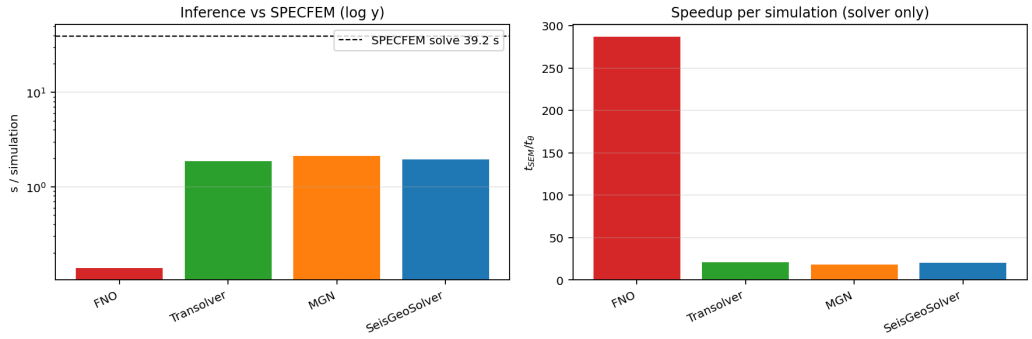

Figure 7: Single-section inference versus the SPECFEM2D solver-only wall time.

8 Ablations

Removing local graph processing yields the global-only Transolver baseline, whose test rollout error rises to $43.2/33.8\%$ despite near-unit energy. Thus, learned global states alone do not preserve wavefield geometry in these heterogeneous geologic models. In particular, they do not reliably represent reflected and refracted arrivals generated at strong velocity contrasts. Conversely, the 15-block MGN matches SeisGeoSolver's error band but has 68% more parameters and requires 42% more training time per epoch.

These results lead to two conclusions. First, explicit local geometric processing is necessary for the reported level of rollout fidelity. Second, global physical processing can replace part of the local-depth budget. These tests, however, do *not* establish that global attention improves accuracy over a fully tuned MGN. The detailed ablation record is given in Appendix E.

9 Discussion

Seismic propagation in heterogeneous media requires local information about where and across which material contrast a wavefront moves. MGN supplies that information directly through edge messages. Physics attention provides a different capability: domain-wide communication among learned physical states. In SeisGeoSolver, the local context stage makes the global states edge-aware indirectly, and the local refinement stage reconnects their global context to the mesh. This explains why the hybrid model is useful without implying that attention replaces geometry.

The present comparison also clarifies the role of global attention. Transolver accurately retains energy but loses spatial alignment; visually, it does not reliably learn the reflected and refracted wavefronts produced by strong velocity contrasts (Fig. 3). These complex arrivals depend on local propagation paths and interface conditions, which motivates explicit local message passing. A sufficiently deep MGN already achieves the best current accuracy band. SeisGeoSolver's demonstrated

Table 4: Comparison on the three GeoFWI test simulations.

Variant	Parameters	Rollout error (u_x/u_z, %)	Interpretation
Transolver	8.97M	43.2/33.8	no local edges
MGN	9.04M	**11.2**/11.9	no global attention; deeper MP
SeisGeoSolver	5.38M	11.8/**11.3**	local–global composition

advantage is efficiency because it obtains comparable rollout accuracy with fewer parameters, fewer local message-passing blocks, and lower training time (Figs. 6–7). It has not provided yet a higher accuracy than a well-tuned MGN.

This study has several serious limitations. The model is 2D acoustic, uses fixed-resolution 96×96 lattices, one interior source. Since we trained the model on a GeoFWI salt distribution (Swaminadhan et al., 2026), errors increase on validation SEG salt patches, especially around the salt (Fig. 11).

For FWI, this work establishes only the first link—an edge-aware learned forward stepper can be fast and stable enough to warrant further study. Before it can serve as a modeling engine for FWI (Yang et al., 2021; 2023; Rasht-Behesht et al., 2021), its derivatives with respect to velocity, multi-source behaviour, and interface-localized bias must be evaluated. Next, the surrogate can be used as a low-frequency or preconditioning component in FWI, whereas, numerical modeling (e.g., finite elements) is employed where inversion is most sensitive. Detailed limitations of the method and implications for FWI are discussed in Appendix F.

10 Conclusions

We framed seismic surrogate modeling as a representation problem; in which wavefields require both local geometric interactions and global physical coordination. The experiments with FNO (Li et al., 2021; Yang et al., 2021; 2023), Transolver (Wu et al., 2024; Luo et al., 2025), MGN (Pfaff et al., 2021; Sanchez-Gonzalez et al., 2020), and SeisGeoSolver show that explicit local edge information is essential for stable long rollout propagation in 2D heterogeneous media. Global physics attention alone does not produce phase-accurate wavefronts.

SeisGeoSolver combines the two representations with a local $\rightarrow$ global $\rightarrow$ local processor. It matches the rollout accuracy band of a 15-block MeshGraphNet while significantly reducing the model size and lowering the reported training cost. The results, however, do not demonstrate yet that the hybrid method is more accurate than a sufficiently deep MGN.

SeisGeoSolver should be viewed as a compact, edge-aware forward surrogate with promising but so far limited evidence for use in seismic inversion workflows.

AI use statement

Generative AI tools were used to assist with software implementation and LaTeX editing. They were not used to invent experimental results, citations, or claims. All model training, SPECFEM simulations, metrics, and scientific conclusions were produced and checked by the authors, who take full responsibility for the final content of the paper.

Reproducibility statement

Configs, training code, and the SPECFEM2D generation pipeline are described in Appendix C. Architecture cards, splits, and metrics can be found in Sections 6–7 and Appendix C.3.

References

Corey Adams, Rishikesh Ranade, Sheel Nidhan, Mohammad Amin Nabian, Deepak Akhare, Ram Cherukuri, and Sanjay Choudhry. Geotransolver: Learning physics on irregular domains using multi-scale geometry aware physics attention transformer. *arXiv preprint arXiv:2512.20399*, 2026.

Neil Ashton, Johannes Brandstetter, and Siddhartha Mishra. Fluid Intelligence: A Forward Look on AI Foundation Models in Computational Fluid Dynamics, November 2025. URL http://arxiv.org/abs/2511.20455. arXiv:2511.20455 [physics.flu-dyn].

Johannes Brandstetter, Daniel Worrall, and Max Welling. Message Passing Neural PDE Solvers, March 2023. URL http://arxiv.org/abs/2202.03376. arXiv:2202.03376 [cs.LG].

Youngsoo Choi, Siu Wun Cheung, Youngkyu Kim, Ping-Hsuan Tsai, Alejandro N. Diaz, Ivan Zanardi, Seung Whan Chung, Dylan Matthew Copeland, Coleman Kendrick, William Anderson, Traian Iliescu, and Matthias Heinkenschloss. Defining Foundation Models for Computational Science: A Call for Clarity and Rigor, May 2025. URL http://arxiv.org/abs/2505.22904. arXiv:2505.22904 [cs.LG].

Di He, Shanda Li, Wenlei Shi, Xiaotian Gao, Jia Zhang, Jiang Bian, Liwei Wang, and Tie-Yan Liu. Learning physics-informed neural networks without stacked back-propagation. *International Conference on Machine Learning*, 2022.

Mikel M. Iparraguirre, Iciar Alfaro, David Gonzalez, and Elias Cueto. MeshGraphNet-Transformer: Scalable Mesh-based Learned Simulation for Solid Mechanics, February 2026. URL http://arxiv.org/abs/2601.23177. arXiv:2601.23177 [cs.LG].

Dimitri Komatitsch and Jean-Pierre Vilotte. The spectral element method: An efficient tool to simulate the seismic response of 2D and 3D geological structures. *Bulletin of the Seismological Society of America*, 88(2):368–392, April 1998. ISSN 1943-3573, 0037-1106. doi: 10.1785/BSSA0880020368. URL https://pubs.geoscienceworld.org/bssa/article/88/2/368/120304/The-spectral-element-method-An-efficient-tool-to.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *International Conference on Learning Representations*, 2021.

Huakun Luo, Haixu Wu, Hang Zhou, Lanxiang Xing, Yichen Di, Jianmin Wang, and Mingsheng Long. Transolver++: An Accurate Neural Solver for PDEs on Million-Scale Geometries, February 2025. URL http://arxiv.org/abs/2502.02414. arXiv:2502.02414 [cs.LG].

Benjamin Moseley, Andrew Markham, and Tarje Nissen-Meyer. Fast approximate simulation of seismic waves with deep learning. *Solid Earth*, 11(4):1527–1549, August 2020. ISSN 1869-9529. doi: 10.5194/se-11-1527-2020. URL http://arxiv.org/abs/1807.06873. arXiv:1807.06873 [physics.geo-ph].

Mohammad Amin Nabian, Chang Liu, Rishikesh Ranade, and Sanjay Choudhry. X-MeshGraphNet: Scalable Multi-Scale Graph Neural Networks for Physics Simulation, December 2024. URL http://arxiv.org/abs/2411.17164. arXiv:2411.17164 [cs.LG].

Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, and Peter Battaglia. Learning mesh-based simulation with graph networks. *International Conference on Learning Representations*, 2021.

Majid Rasht-Behesht, Christian Huber, Khemraj Shukla, and George Em Karniadakis. Physics-informed Neural Networks (PINNs) for Wave Propagation and Full Waveform Inversions. 2021. doi: 10.48550/ARXIV.2108.12035. URL https://arxiv.org/abs/2108.12035.

Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter W. Battaglia. Learning to Simulate Complex Physics with Graph Networks, September 2020. URL http://arxiv.org/abs/2002.09405. arXiv:2002.09405 [cs.LG].

Harpreet Sethi, Doris Pan, Pavel Dimitrov, Jeffrey Shragge, Gunter Roth, and Ken Hester. Hard enforcement of physics-informed neural network solutions of acoustic wave propagation. *Computational Geosciences*, 27(5):737–751, October 2023. ISSN 1420-0597, 1573-1499. doi: 10.1007/s10596-023-10232-3. URL https://link.springer.com/10.1007/s10596-023-10232-3.

Sujith Swaminadhan, Yiran Shen, Chao Li, Kai Gao, Ting Chen, Sergey Fomel, Tolulope Agbaje, Yang Cui, Liuqing Yang, Jaewook Lee, Robin Dommisse, Umair bin Waheed, Mrinal Sen, and Yangkang Chen. GeoFWI3D: Large-scale 3D Velocity Model Dataset for Deep Learning-assisted Seismic Imaging, 2026. URL https://doi.org/10.5281/zenodo.20148778.

Haixu Wu, Huaiqian Luo, Haoyang Wang, and Mingsheng Long. Transolver: A fast transformer solver for PDEs on general geometries. *International Conference on Learning Representations*, 2024.

Yan Yang, Angela F. Gao, Jorge C. Castellanos, Zachary E. Ross, Kamyar Azizzadenesheli, and Robert W. Clayton. Seismic Wave Propagation and Inversion with Neural Operators. *The Seismic Record*, 1(3):126–134, October 2021. ISSN 2694-4006. doi: 10.1785/0320210026. URL https://pubs.geoscienceworld.org/tsr/article/1/3/126/609317/Seismic-Wave-Propagation-and-Inversion-with-Neural.

Yan Yang, Angela F. Gao, Kamyar Azizzadenesheli, Robert W. Clayton, and Zachary E. Ross. Rapid Seismic Waveform Modeling and Inversion with Neural Operators. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–12, 2023. ISSN 0196-2892, 1558-0644. doi: 10.1109/TGRS.2023.3264210. URL http://arxiv.org/abs/2209.11955. arXiv:2209.11955 [physics.geo-ph].

Caifeng Zou, Yaozhong Shi, Zachary E. Ross, Robert W. Clayton, and Kamyar Azizzadenesheli. Enforcing Reciprocity in Operator Learning for Seismic Wave Propagation, April 2026. URL http://arxiv.org/abs/2602.11631. arXiv:2602.11631 [physics.geo-ph].

A Detailed Problem Formulation

A.1 Reference dynamics and temporal output

The reference is 2D acoustic wave propagation with a Ricker source wavelet. SPECFEM2D (Komatitsch & Vilotte, 1998) advances the internal solver at $\delta t = 4.5 \times 10^{-4}$ s and exports snapshots every 15 solver steps, hence $\Delta t = 6.75$ ms. There are 160 stored snapshots; the first three near-silent frames are omitted, leaving 156 predicted updates. The model predicts forward velocity increments and updates displacement with explicit Euler integration. It emits $B = 4$ increments per call and is unrolled for $K = 2$ calls during training.

A.2 Node and edge representation

At each node, the 16-dimensional feature vector comprises current displacement (u_x, u_z), backward difference velocity $(\dot{u}_x^-, \dot{u}_z^-)$, v_p, density, $|\nabla v_p|$, free-surface and absorbing-boundary masks, source coordinates, spatial coordinates, normalized source frequency, instantaneous Ricker amplitude, and normalized time (Fig. 8). A directed graph edge stores relative coordinates, Euclidean distance, local travel-time proxy, log velocity contrast, and average log velocity. The default graph is an eight-neighbour mesh-graph; four-neighbour and Delaunay alternatives are reserved for ablations.

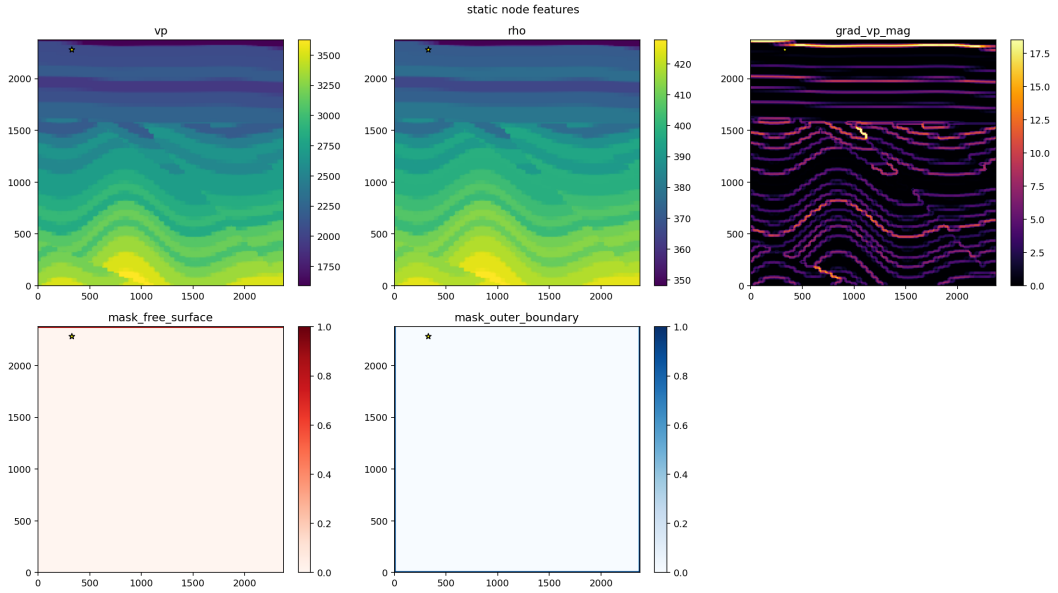

Figure 8: Static node channels on one GeoFWI section: v_p, Gardner density, $|\nabla v_p|$, free-surface mask, and absorbing-boundary mask.

A.3 Objective

For supervised rollout positions k, the loss combines normalized forward-velocity error, integrated displacement error, and a phase-sensitive spatial-Fourier loss (Brandstetter et al., 2023; Sanchez-Gonzalez et al., 2020):

$$\mathcal{L} = \frac{1}{Z} \sum_k w_k \left(\mathcal{L}_{\dot{u}}^{(k)} + 0.5 \mathcal{L}_u^{(k)} + \mathcal{L}_\phi^{(k)} \right). \tag{6}$$

where $\mathcal{L}_{\dot{u}}^{(k)}$ is the normalized forward-velocity error, $\mathcal{L}_u^{(k)}$ is the integrated displacement error, and $\mathcal{L}_\phi^{(k)}$ is the phase-sensitive spatial-Fourier loss. The phase term compares the normalized complex Fourier coefficients after masking the absorbing boundary. The default rollout weights are $(1, 2)$ at the first and final unrolled positions.

B Architecture and Rollout Details

B.1 Baseline operators

FNO lifts the 16 grid channels, applies four Fourier layers retaining 16 modes, and decodes eight bundled-output channels (Li et al., 2021; Yang et al., 2021; 2023). MGN uses an encoder–processor–decoder with 15 residual edge–node message-passing blocks (Pfaff et al., 2021; Sanchez-Gonzalez et al., 2020; Nabian et al., 2024). Transolver uses a 12-block node-only processor with 256 learned physical states (Wu et al., 2024; Luo et al., 2025). In each physics-attention block, nodes are softly assigned to states, state tokens attend globally, and the attended states are projected back to the nodes.

B.2 SeisGeoSolver processor and complexity

SeisGeoSolver has latent width 256, eight attention heads, 256 physical states, and a three-block message-passing prologue and epilogue around eight physics-attention blocks. Its local cost is proportional to $2P|E|d^2$ and its global cost to $L_a H d(NG + G^2)$, where N, E, d, G, and H denote nodes, directed edges, latent width, states, and heads.

B.3 Autoregressive rollout

After each bundled prediction, the four Euler updates are applied sequentially and the resulting state becomes the input to the next network call. The training uses the model's predicted state for the second bundle so that the model encounters its own rollout errors (Brandstetter et al., 2023). At test time this procedure continues from the first retained frame through the complete 156-step trajectory.

C Data set, Training, and Metrics

C.1 Data set and split

The primary data consist of 2,000 2D simulations extracted from 1,000 3D velocity cubes from the GeoFWI3D salt model (Swaminadhan et al., 2026). Two sections per cube are used and splitting occurs at the cube level: 1,800/100/100 sections for training, validation, and test. The physical domain is exported at 12.5 m and strided to a 96×96 (25 m) training grid. Reference simulations use the free-surface boundary conditions and CPML on the other boundaries. One interior Ricker source wavelet is sampled per section; its frequency is selected from local velocity and lies approximately between 8 and 11 Hz.

C.2 Optimization

Models train for 300 epochs with Adam optimizer, learning rate 10^{-4}, weight decay 10^{-5}, five-epoch warmup, cosine decay, gradient clipping at 1, and mixed precision on two 80-GB A100 GPUs. Input displacement is perturbed by Gaussian noise with standard deviation 3×10^{-3} (Sanchez-Gonzalez et al., 2020; Pfaff et al., 2021). Batch size is 24 except for MGN (20). Validation checkpoint selection uses rollout error subject to an active-frame energy ratio of at least 0.3.

C.3 Metrics

For each displacement component, the headline metric is

$$\varepsilon_\alpha = 100 \, \frac{\|\widehat{u}_\alpha - u^\star_\alpha\|_2}{\|u^\star_\alpha\|_2}. \tag{7}$$

where $\widehat{u}_\alpha$ is the predicted displacement and $u^\star_\alpha$ is the reference displacement.

The norm spans all predicted frames and nodes. We additionally report active-frame error, final-frame relative L^1, and active-frame energy ratio $E = \|\widehat{\mathbf{u}}\|_2 / \|\mathbf{u}^\star\|_2$. These distinguish amplitude collapse from phase or spatial drift.

D ADDITIONAL RESULTS

D.1 FURTHER GEOFWI MODEL SIMULATIONS

Table 5 lists per-simulation rollout error for the three visualized GeoFWI cases. Sim 3 is in the main text (Figs. 3–5); sims 1 and 4 are shown in Figs. 9–10.

Table 5: Per-simulation GeoFWI rollout error (%) for u_x/u_z.

Model	sim 1	sim 3	sim 4
FNO	97.2/103.4	95.8/89.1	108.4/102.3
Transolver	36.4/34.4	31.7/34.5	61.5/32.3
MGN	14.4/17.1	7.9/9.9	11.3/8.7
SeisGeoSolver	16.6/15.7	8.1/8.3	10.6/10.0

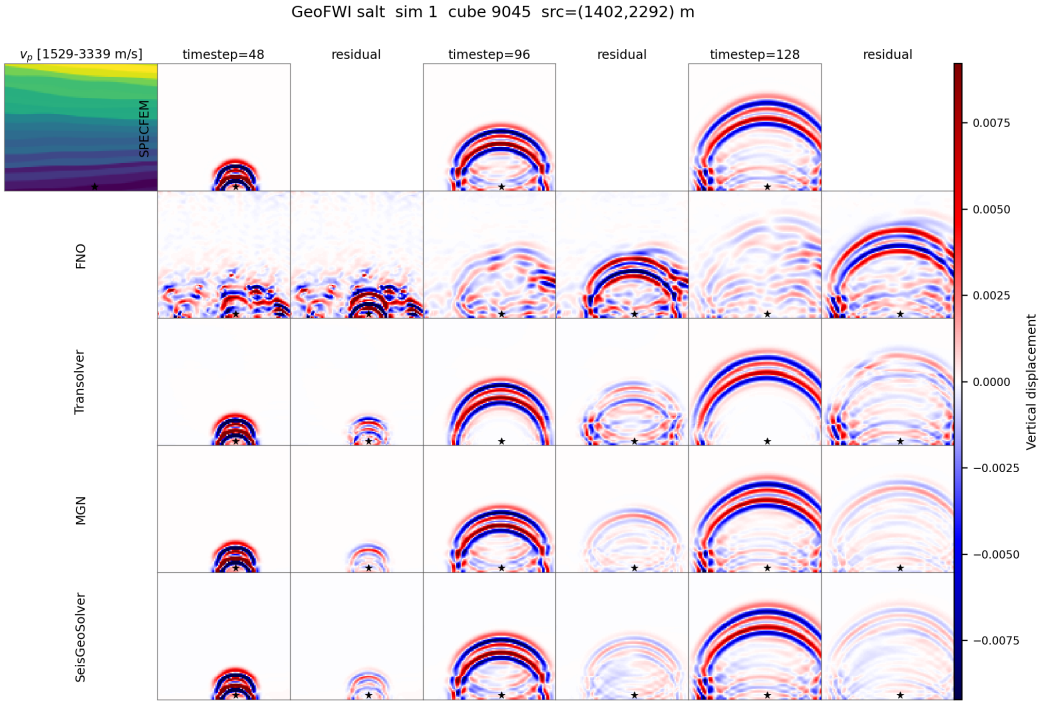

Figure 9: Additional GeoFWI rollout: test sim 1.

The late-horizon diagnostics identify different failure modes. FNO collapses amplitude ($E \approx 0.65$ on the three-sim subset). Transolver keeps the energy closer to one but accumulates spatial and phase errors. MGN and SeisGeoSolver have residuals concentrated near salt and the free surface.

D.2 SEG SALT MODEL

Table 6 reports zero-shot error on four SEG windows. The results are shown below.

D.3 COMPUTE DIAGNOSTICS

Figure 15 records epoch wall time, including validation and the occasional three-sim rollout monitor. Figure 16 is the SPECFEM2D solver-time distribution used for t_{SEM} in Fig. 7. Table 7 summarises the same catalog.

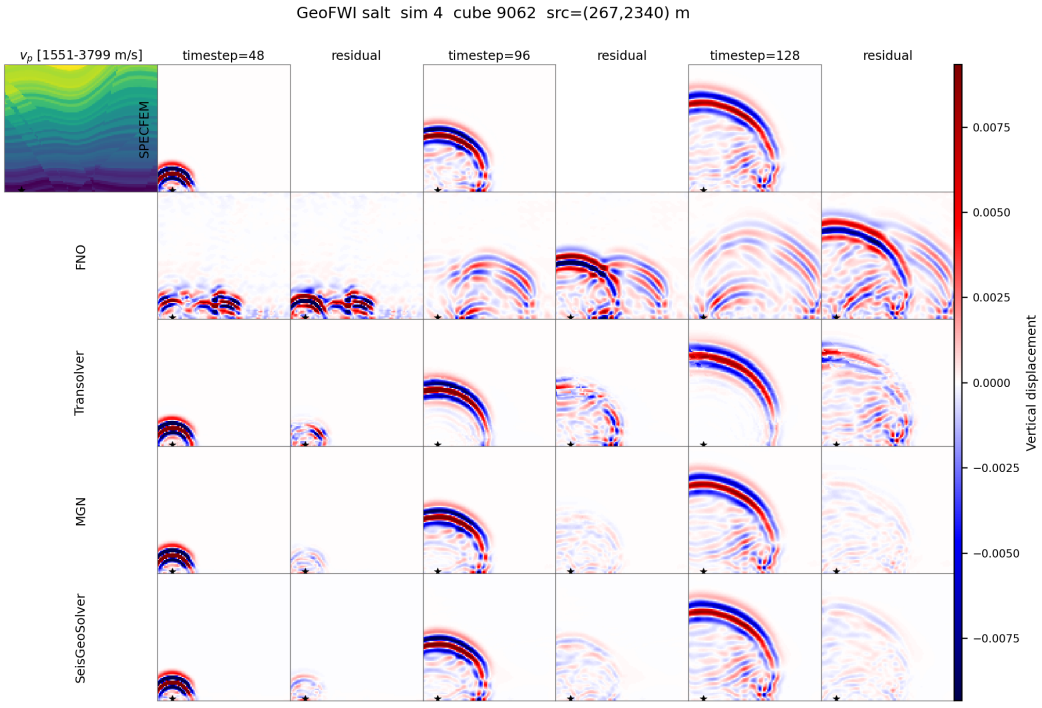

Figure 10: Additional GeoFWI rollout: test sim 4.

Table 6: Zero-shot rollout error (%) for u_x/u_z in the SEG salt model.

Model	w0 (0% salt)	w1 (27%)	w2 (55%)	w3 (30%)
FNO	104.0/118.7	139.4/140.5	97.0/81.0	118.4/142.6
Transolver	76.3/45.4	90.7/84.7	81.0/74.0	53.7/59.1
MGN	15.7/13.9	34.0/41.0	30.6/28.7	46.6/45.6
SeisGeoSolver	11.8/14.3	40.0/45.5	31.2/29.1	42.6/40.7

E Ablation Record and Limitations

The trained-catalog ablation compares Transolver (no local edges), MGN (no global attention but 15 local blocks), and SeisGeoSolver. It supports that local geometry is necessary and global attention is an efficient complement.

F Expanded Discussion: Scope for FWI

The tests were limited to 2D wave propagation in acoustic fixed-resolution salt sections with one source and masks that encode the free surface and CPML. It does not establish resolution transfer, 3D scalability, elastic or anisotropic dynamics, unknown wavelets, or multi-source generalization. The remaining error is concentrated near strong interfaces—precisely where FWI is sensitive to cycle skipping. Before using the surrogate as a replacement forward operator, future work should test the velocity derivatives, inversion trajectories, and hybrid workflows in which the surrogate provides low-frequency propagation or preconditioning while numerical modeling is employed for high-fidelity parameter updates.

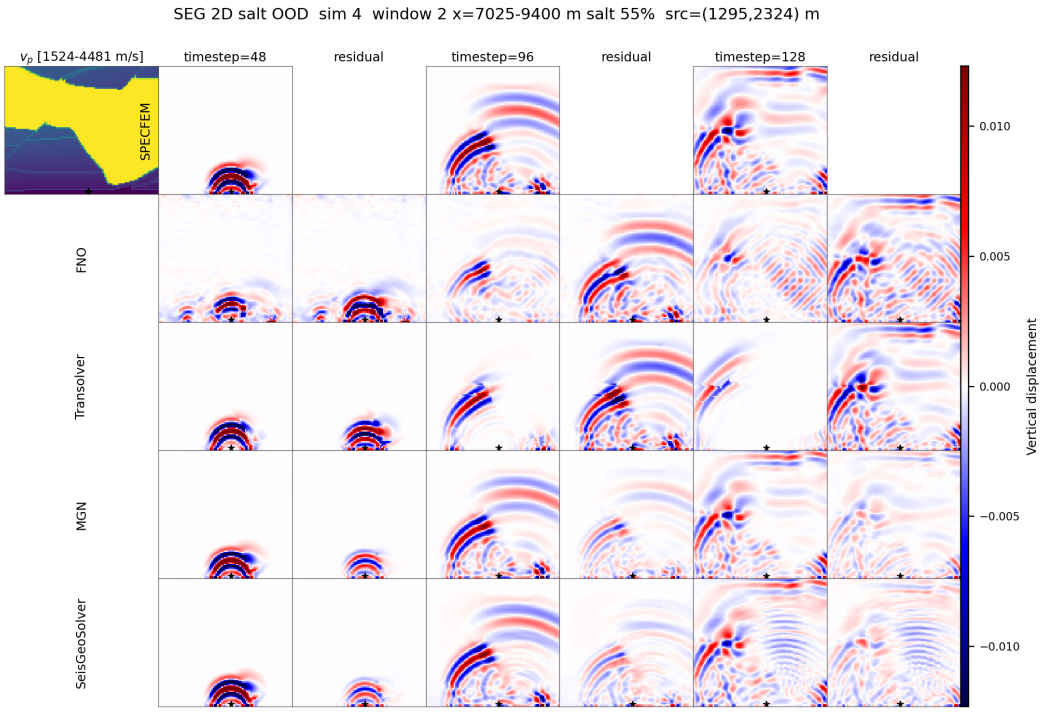

Figure 11: Zero-shot rollout for the SEG salt model (sim 4, 55% salt).

Table 7: Training and inference cost. Inference is one A100, 156 Euler steps, GeoFWI sim 3. Speedup is $t_{\mathrm{SEM}}/t_\theta$ using the mean SPECFEM solve time 39.2 s.

Model	Parameters	s/epoch	s/sim	Speedup
FNO	9.45M	15.3	0.14	287×
Transolver	8.97M	401	1.85	21×
MGN	9.04M	759	2.12	18×
SeisGeoSolver	5.38M	534	1.94	20×

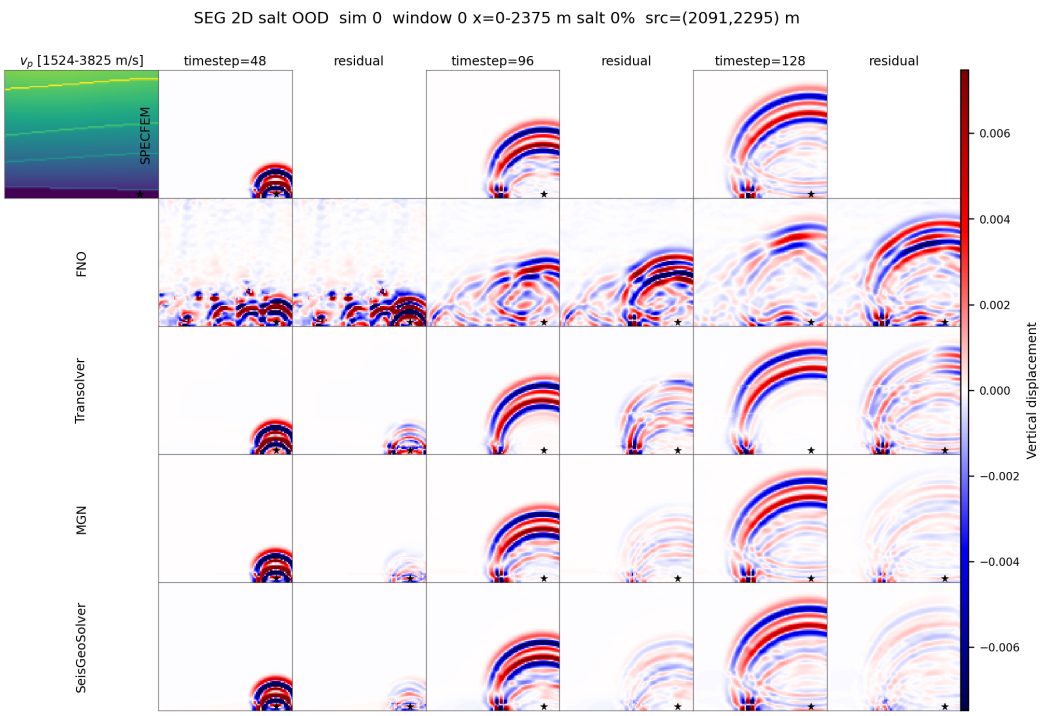

Figure 12: Additional SEG window: sim 0 (0% salt).

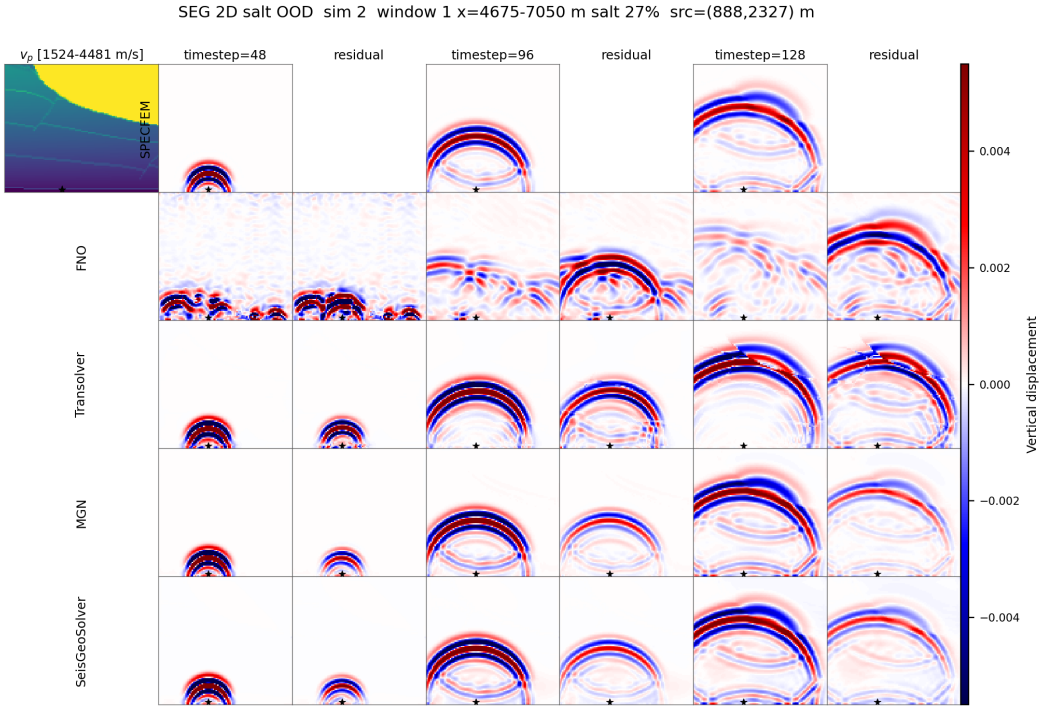

Figure 13: Additional SEG window: sim 2 (27% salt).

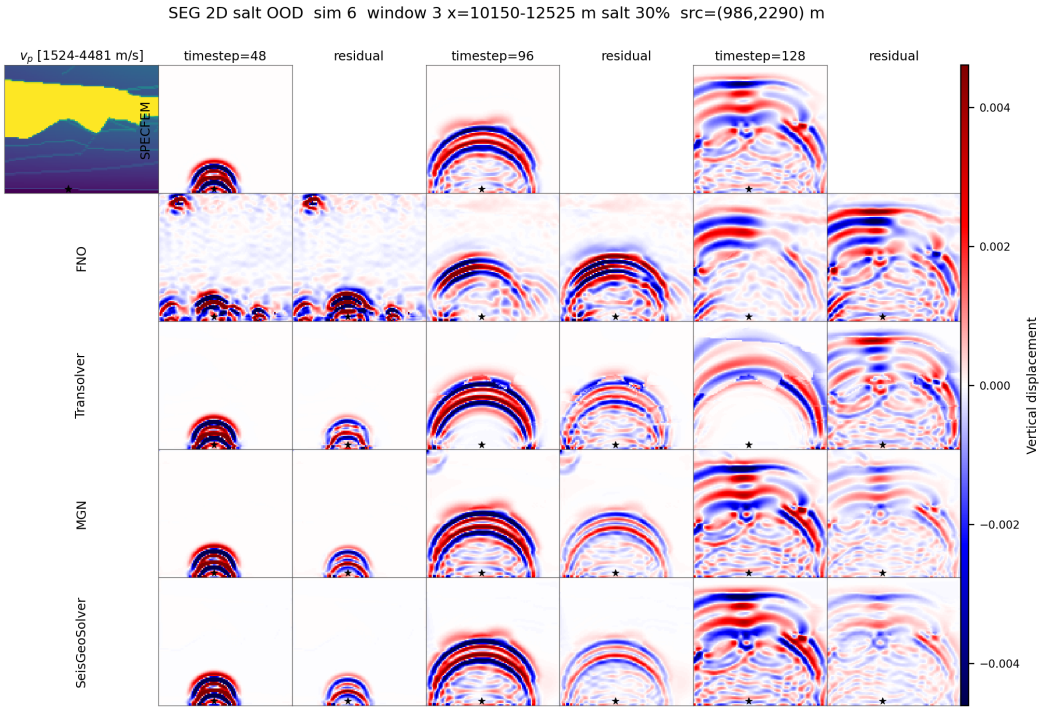

Figure 14: Additional SEG window: sim 6 (30% salt).

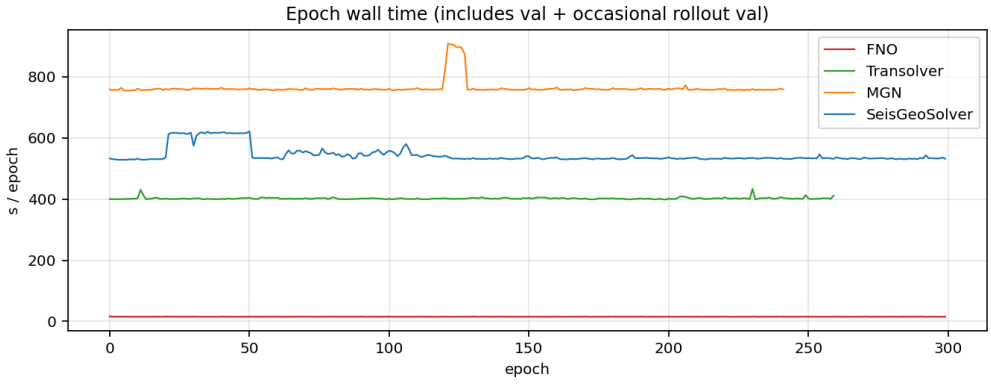

Figure 15: Logged seconds per training epoch. Occasional spikes coincide with rollout validation.

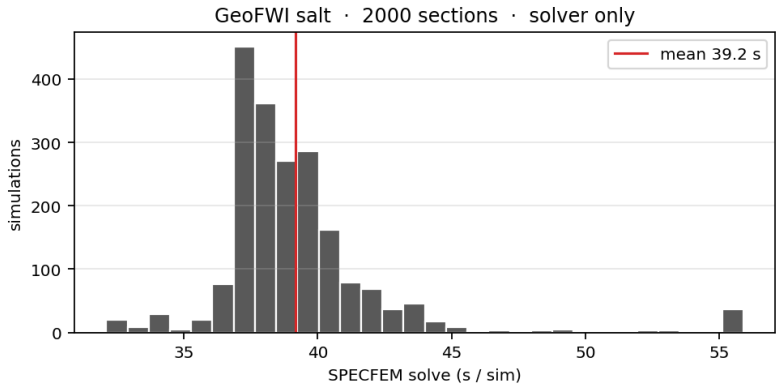

Figure 16: SPECFEM2D solver-only wall time over 2,000 GeoFWI sections (mean 39.2 ± 3.2 s). Mesh generation and zarr export are excluded.