Accelerating Tomographic Updates with Physics-Based Deep Learning and Realistic Earth Models

Mark Roberts*, Leandro Gabioli, Olga Brusova, David Brookes and Alejandro Valenciano
TGS

Summary

Velocity Model Building (VMB) remains a significant bottleneck in frontier exploration due to the intensive iterative nature of tomography and the frequent absence of high-quality initial models. We present a pragmatic approach to automating tomographic updates (MLTomo) using advanced neural architectures and physics-based regularization. This study seeks to close the gap between aspirational machine learning research and production-ready workflows that generalize on real-world datasets.

Introduction

Traditionally, VMB requires highly experienced geophysicists to navigate complex sequences involving Full Waveform Inversion (FWI), tomography, and manual geological interpretation. In frontier areas, where models often rely on legacy 2D data, acquiring an accurate macro-velocity model is inherently difficult. While deep learning-based velocity estimation has a long history, recent interest has shifted from the direct conversion of shot gathers to more manageable tasks within the model space.

Following the philosophy established by Crawley *et al.* (2024), our approach leverages the simplicity of mapping within the model space by utilizing migrated gathers as input rather than raw field shot gathers. Migration serves to regularize and filter the data, effectively narrowing the disparity between synthetic training data and complex field data containing noise and irregular acquisition geometry. Because errors in a migration velocity model typically manifest non-locally elsewhere in the image, capturing long-range dependencies is critical for effective Migration Velocity Analysis (MVA). By keeping the mapping within the model space and utilizing residual moveouts and defocused images, we present a more manageable task for the neural network, facilitating better generalization to unseen field datasets.

We evaluate three distinct neural network architectures: Fourier Neural Operators (FNO), the hybrid transformer-CNN RAPUnet, and a ConvNeXt V2-enhanced DuckNet. These models are trained on a novel dataset incorporating lithological trends from over 60,000 global offshore wells and integrated with realistic salt masks (Roberts *et al*., 2025). Our results, detailed in the convergence analysis (Figure 1), demonstrate that a hybrid CNN-transformer architecture (RAPUnet) significantly outperforms pure convolutional or operator-based models. Furthermore, we introduce a cross-gradient regularization term (Gallardo *et al*., 2004) to ensure structural consistency between the predicted velocity model and the seismic reflectivity.

Model Selection

We reviewed and benchmarked three distinct architectures for the tomographic update task:

• **Fourier Neural Operator (FNO):** As implemented in Crawley *et al*. (2024), FNOs utilize global convolutions efficiently computed via Fast Fourier Transforms (FFTs). While mesh-independent and capable of capturing non-local effects, they can struggle with fine-grained local geological details in complex environments.

• **RAPUnet:** A hybrid CNN-transformer model based on Lee *et al*. (2024), utilizing a MetaFormer backbone. We modified the architecture with a convolutional first layer to map 81-channel input gathers and adapted the head for regression. The core "RAPU" (Residual and Atrous Convolution in Parallel Unit) blocks combine residual connections for small details with atrous convolutions to extract multi-scale features.

• **DuckNet-ConvNeXt V2:** Based on the DuckNet architecture, we updated this model to incorporate ConvNeXt V2 blocks (Woo *et al*., 2023). By adding a Global Response Normalization (GRN) layer, we mitigate the "feature collapse" phenomenon often observed in pure ConvNets, promoting channel diversity and inter-channel feature competition.

Observation of the model convergence (Figure 1) indicates that the hybrid architecture of RAPUnet achieves the lowest validation loss and the most stable training curve. We speculate that this superiority stems from the combination of specialized CNN blocks for extracting small-scale, fine-grained features and a Transformer backbone for capturing long-range dependencies. This dual capability is particularly well-suited for velocity estimation from gathers, where the network must simultaneously resolve local RMO details and understand non-local velocity-depth trade-offs.

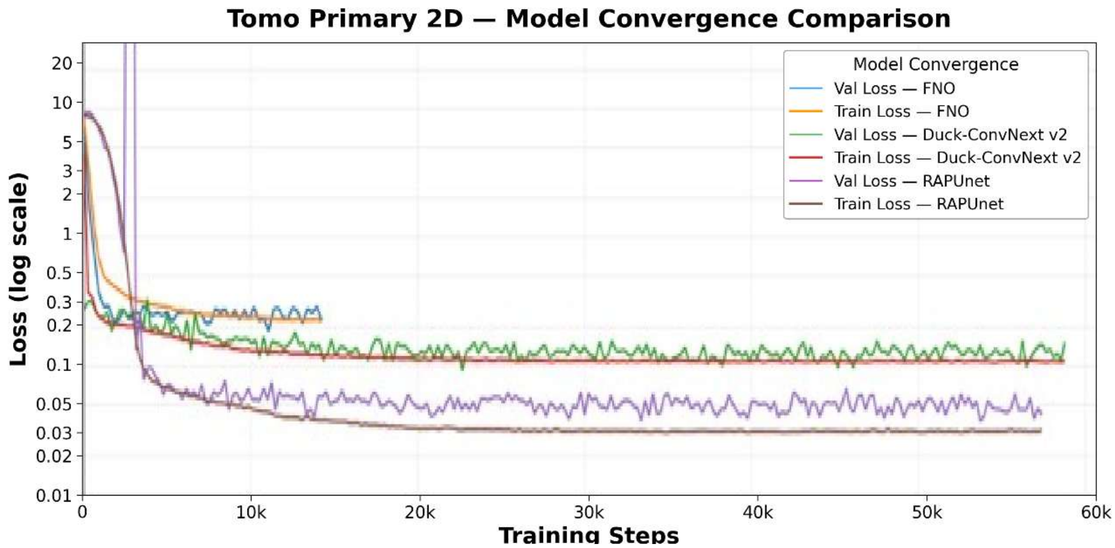

Figure 1: Model convergence analysis demonstrating the improved performance of the Hybrid CNN-Transformer RAPUNet model

Training Data: Realistic Salt and Earth Models

The efficacy of ML-based tomography is critically dependent on training data that adequately represents real-world geological variability. Following Roberts *et al*. (2025), we generated a training dataset leveraging 60,090 offshore wells. This workflow includes ML-driven prediction of missing shear wave logs.

To overcome the limitations of "blob-like" salt models in standard packages, we modified the SynthoSeis package to integrate realistic, interpreted salt masks from a global data library (Roberts *et al*., 2025). Lithological trends from actual well data were integrated into the pseudo-synthetic models to ensure petrophysical diversity. This hybrid approach provides high-quality labels in structurally complex areas, essential for the model to generalize to challenging frontier basins.

Method: Physics-Based Loss and Regularization

To ensure the model learns wave propagation physics rather than just statistical correlations, we utilize a multi-component loss function evaluating residual moveout (RMO). The input Kirchoff Depth Migrated Offset gathers are first converted to time, then RMO is applied to the gathers using the migration velocity and estimated velocity from the network (the velocities are also converted to time and RMS). RMO is only valid in laterally continuous media. As a result, the following Stack-Power loss and DSO loss are only applied with a mask, limiting their impact to areas where this approximation is valid.

Stack Power Loss: Maximizing the coherence of corrected gathers is achieved by maximizing the stack power:

$$J_{SP} = \int \left[\iint I_{rmo}(x,\ t, h; m) dh \right]^2 dx\, dt$$

Where I_{rmo} represents the RMO corrected gathers and h are the offsets.

Differential Semblance Optimization (DSO) Loss: To minimize issues with cycle skipping in the stack power loss and ensure gathers are flattened, we minimize the DSO loss:

$$J_{DSO} = \int \int \int \left[\frac{\partial I_{rmo}(x, t, h; m)}{\partial h} \right]^2 dh\, dx\, dt$$

Cross-Gradient Regularization Structural consistency is enforced by minimizing the cross-gradients between the predicted velocity model and a stack of the input offset gather (Gallardo *et al*., 2004).

$$J_{cg} = \sum |\nabla I_{stack} \times \nabla m_{vel|}^2$$

This term promotes collinearity between the gradients of the velocity field and the seismic image: By minimizing, we ensure that velocity boundaries align with seismic reflectivity, providing a robust structural link that prevents the network from generating geologically implausible velocity updates.

During model training two different loss functions and datasets were used simultaneously. The synthetic data was used in a supervised manner with a combined Huber and SSIM loss function. The read data and gathers was also used but in a self-supervised manner (the correct velocity model is unknown), using the cross-gradient loss function as well as the RMO based Stack Power and Differential Semblance Optimization loss functions targeted in the shallow section where the velocity and reflectivity seem to be predominantly laterally continuous.

Brazil (Equatorial Margins) Streamer Example

The primary dataset utilized here for demonstration of the proposed method is the PAMA 3D survey, a large-scale acquisition located offshore the north coast of Brazil. This survey covers a vast frontier exploration area of approximately 17,000 square kilometers, presenting a highly representative and challenging environment for modern velocity model building. The region is situated on a rifted margin and is characterized by extreme topographic variations on the seafloor, with water depths ranging from 1,000 to 4,000 meters. The bathymetry features a complex landscape composed of intricate channels, steep-walled canyons, seamounts, and reefs, which create significant ray-path distortions and substantial imaging hurdles.

Geologically, the subsurface environment is equally demanding, containing significant igneous rocks in the form of high-velocity volcanic sills and intrusions. These volcanic features, embedded within the sedimentary section, create high-contrast velocity boundaries and complex wave propagation effects that pose substantial obstacles for traditional reflection tomography and initial macro-model construction. The presence of these high-velocity geobodies makes the PAMA 3D dataset an ideal benchmark for testing the MLTomo approach, as accurate velocity estimation is critical for providing a robust starting model for Full Waveform Inversion (FWI) in areas where legacy data is often sparse or unavailable.

The seismic data was acquired using multi-sensor streamer technology. The acquisition configuration utilized a spread of 12 streamer cables, with each cable containing 804 channels at a fine 12.5-meter group interval. This high-density acquisition ensures a high signal-to-noise ratio and broad bandwidth, providing the necessary data quality to drive deep learning-based velocity estimation.

Despite the extensive pre-training on realistic synthetic models, applying a network trained solely on synthetic data to complex field datasets often results in significant artifacts and a lack of generalization. As illustrated in Figure 2, the velocity updates derived from such models frequently exhibit spatial instabilities and localized noise that lack geological continuity. Crucially, the background velocity model predicted by the synthetic-only network often fails to adequately flatten the migrated gathers, leaving visible residual moveout (RMO) that compromises the final image quality. This failure highlights the "domain gap" between idealized synthetic physics and real-world seismic data, which includes complex noise patterns and irregular acquisition geometries not fully captured in the training set. These artifacts demonstrate that synthetic pre-training alone is insufficient for production-level accuracy, making the subsequent physics-based, self-supervised fine-tuning stage essential to ensure the model generalizes to the specific seismic characteristics of the target area and delivers a flattened, structurally consistent result.

Figure 3, shows the results from the same model after fine-tuning using the hybrid approach described in this abstract using both the synthetic data in a supervised and the real gathers using a RMO based loss on the shallower more

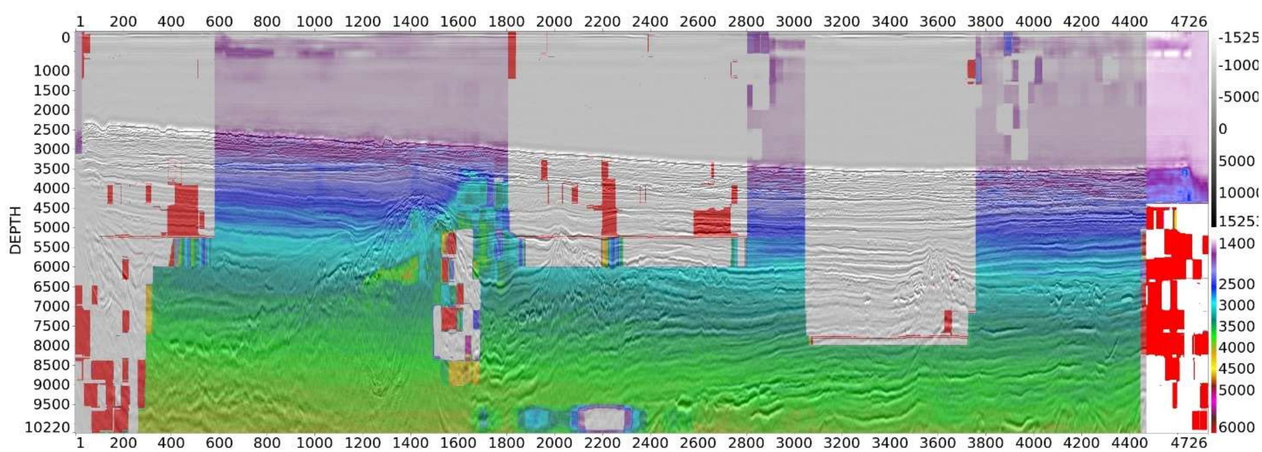

Figure 2 Example of inference on Kirchoff Depth migrated gathers with model trained on synthetics

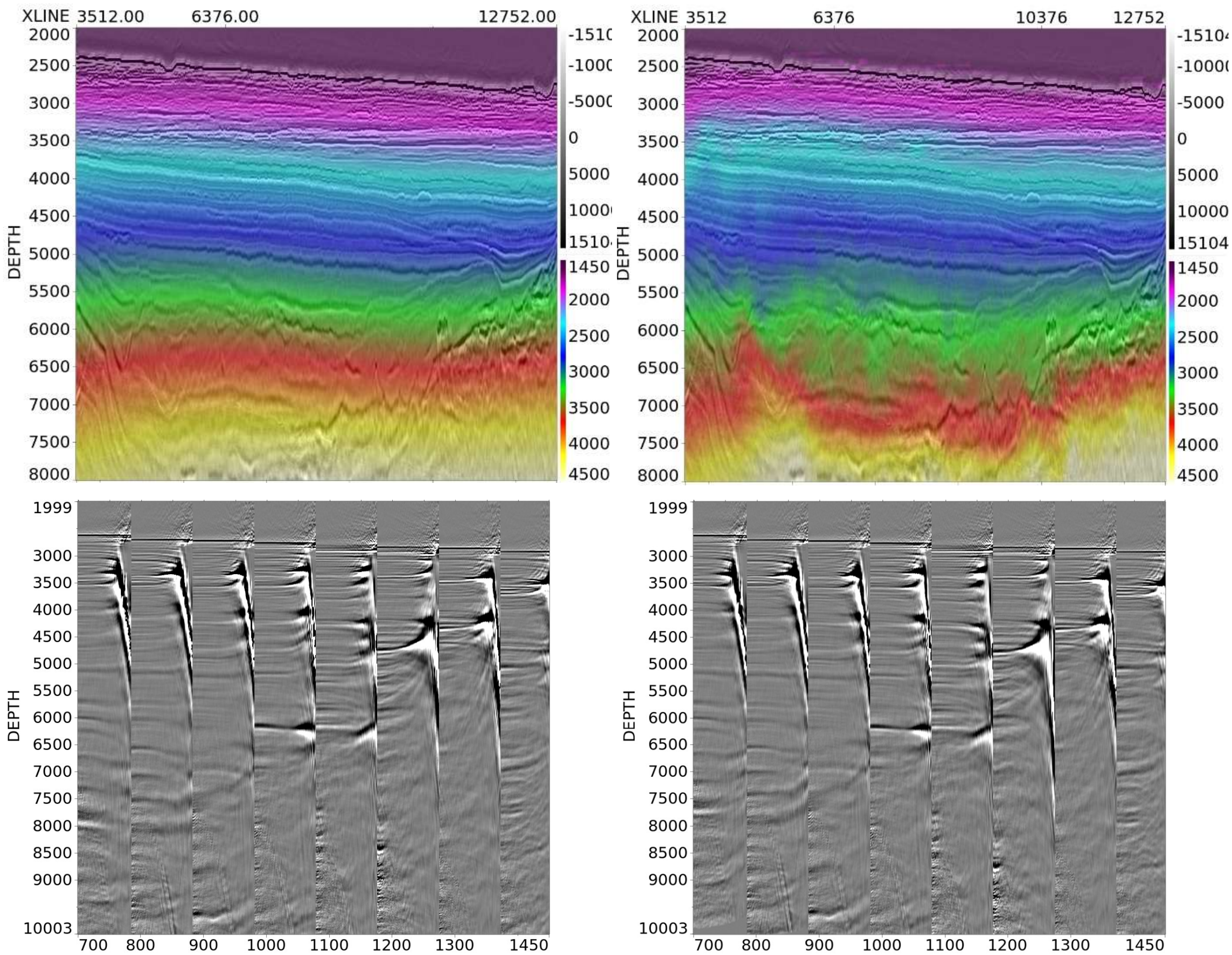

Figure 4 Velocity model (top) and Kirchoff Depth Offset gathers (bottom) before (left) an after (right) ML tomographic update.

laterally continuous section. The resulting model is now stable everywhere and an examination of the gathers shows improved gather flatness, particularly around 7 km in the deeper syn-rift section.

Conclusions

The proposed ML based velocity model updating workflow has the potential to significantly reduces VMB turnaround time by providing high-quality starting models suitable for immediate FWI refinement. We have shown that hybrid CNN-Transformer architectures like RAPUnet can outperform FNO and CNN architectures. Realistic earth models are and synthetic data alone can be challenging to reliably predict velocity updates on unseen real data. Using a RMO based loss in a self-supervised manner in conjunction with cross-gradient regularization, we achieve a level of structural consistency and generalization on real data previously unattainable with simpler approaches.

Acknowledgments

We are thankful to TGS management for permission to show the data. We thank Lean Ge for his help providing the inputs to the work shown here.