Reducing Turnaround Time in Frontier Basins through ML-Accelerated Preprocessing

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

Summary

This study presents a hybrid ML workflow for dual-sensor towed-streamer data that tackles three of the most computationally intensive steps in marine seismic preprocessing: deblending, denoising, and deghosting. Each ML step is designed to minimize manual parameter tuning: denoising models are trained on pseudo-synthetic examples generated from real noise and cleaned records; deblending uses an ML-created seed to accelerate sparse iterative inversion; and deghosting incorporates both hydrophone and geophone inputs through a modified DuckNet architecture. By utilizing the complementary ghost responses of co-located geophones, the deghosting model achieves a validation loss of 0.017, compared to 0.025 for a hydrophone-only model, demonstrating that the geophone supplies information that the network cannot extract from pressure data alone. The workflow is tested on a 3D survey along the Equatorial Margin of Brazil and processes individual sail lines within hours, greatly reducing the parameterization effort compared to traditional processing methods.

Introduction

Building on recent advances in single-sensor machine learning applications (e.g., Roberts *et al*., 2025), this study introduces a hybrid workflow optimized for dual-sensor streamer data, targeting three often time-consuming and often computationally intensive stages of marine seismic preprocessing: denoising, deblending, and deghosting. For denoising, models are trained on pseudo-synthetic examples that combine real noise with cleaned records from both sensor types. For deblending, an ML-derived initial estimate seeds a sparse iterative inversion, reducing the number of iterations needed. For deghosting, the model utilizes the complementary ghost response of co-located hydrophone and geophone sensors, with preliminary results showing that including the geophone component significantly lowers training loss compared to a hydrophone-only model. We apply this physics-integrated ML approach to a 3D survey offshore Brazil. The resulting workflow processes individual sail lines with minimal manual parameter tuning, generating a high-resolution baseline suitable for Velocity Model Building (VMB) and initial interpretation.

ML Denoise

Building on the experience with ML processing of hydrophone-only data (Brusova *et al*., 2021), we have developed four enhanced denoising models tailored for dual-sensor data. These models leverage an expanded training dataset composed of both raw noise recordings and carefully denoised signals from hydrophone and geophone sensors. To train the primary denoising models, pseudo-synthetic training examples are generated by randomly adding real noise to the cleaned records for each sensor type. Combining the data pipeline for training enables robust data augmentation, such as random cropping and random scaling of noise levels, significantly improving model robustness. To further safeguard data fidelity, a secondary model is applied to mitigate potential signal leakage from the denoising process (Roberts *et al*., 2024).

ML Assisted Deblending

Iterative inversion deblending methods that rely on sparsity constraints, such as POCS and FISTA, are known for their effectiveness but are computationally intensive and slow. Instead of replacing these with a standalone ML solution, we have implemented an ML-assisted workflow that accelerates the inversion process while preserving its robustness and builds on the seeded-deblend idea proposed by Udengaard *et al.* (2024). One challenge with sparse-iterative solvers is the time and cost to develop a robust model. In this work, we use the blind-trace model to develop an initial estimate (or seed) for the iterative inversion step. This hybrid approach reduces turnaround time without compromising deblending quality.

ML Deghosting

Wavefield-separation methods applied to dual-sensor towed streamer data present a robust approach to removing the receiver-side deghost notch (Day *et al*., 2013). The key idea behind dual-sensor acquisition is that the notch frequency for

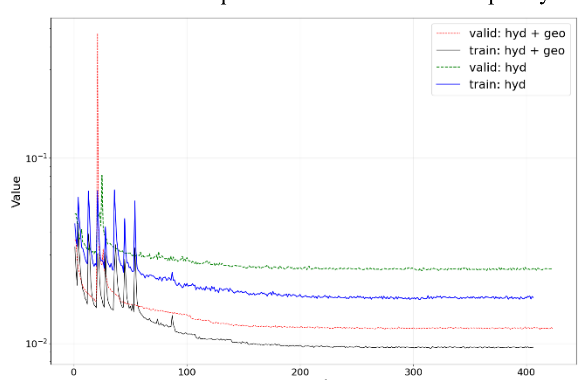

Figure 1: This shows that training and validation losses when training ML models for deghost with hydrophone only (hyd) verses the two-channel hydrophone and geophone (hyd + geo).

the particle's pressure corresponds to a maximum response for particle motion. With estimates of the vertical wavenumber (from calculations of obliquity effects) and instrument responses, the geophone and hydrophone data can be combined to generate a ghost-free upgoing pressure field. One limitation of this method is the geophone's sensitivity to mechanical noise at lower frequencies, which typically limits the usable frequency range and requires a

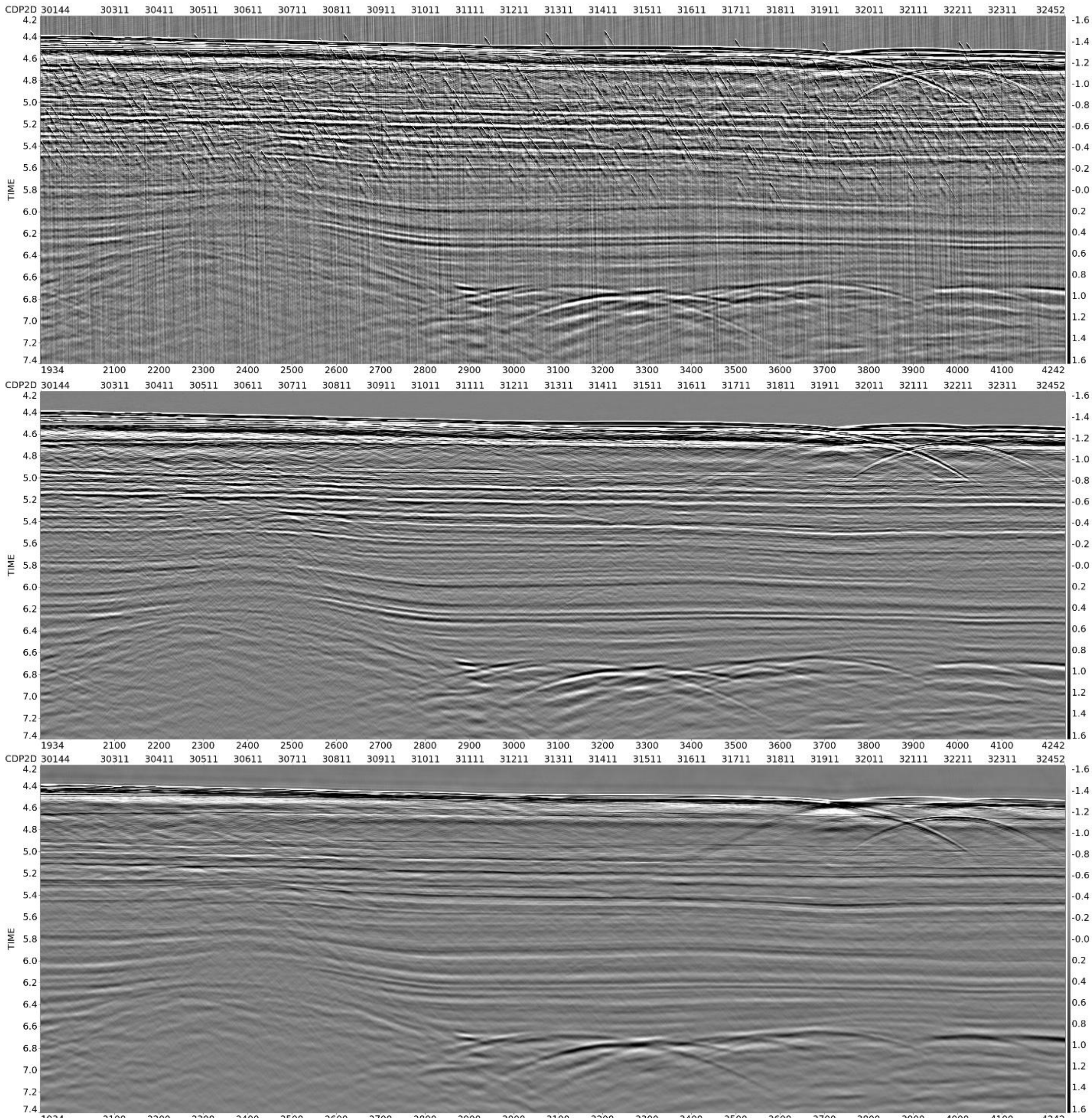

Figure 2: Stacked sections of the raw data (top), data after deblend and deghost (middle) and complete ML accelerated flow (bottom)

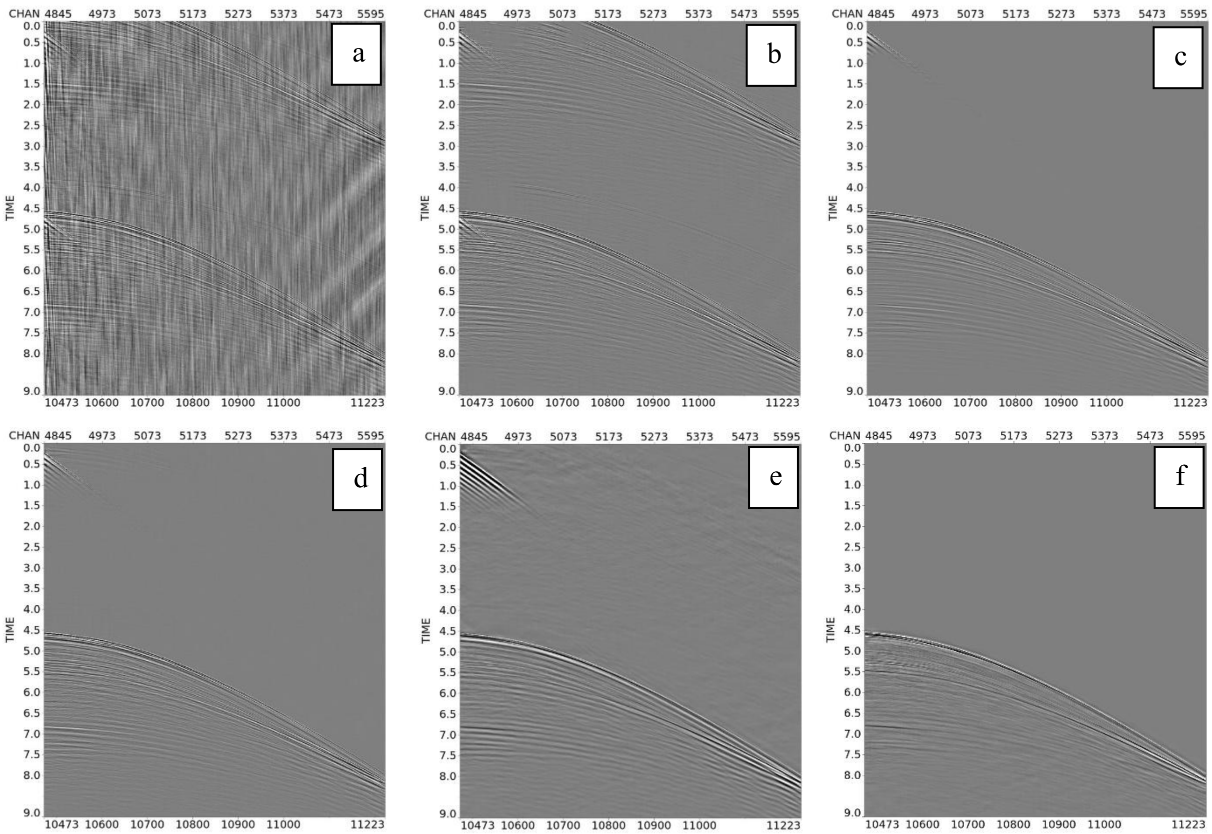

Figure 3: A single shot-gather across the processing flow: (a) raw data, (b) ML denoise (c) ML-seed for deblending, (d) ML seeded deblend, (e) ML multi-sensor deghost, (f) designature and debubble.(a) - (d) are hydrophone only. (e) and (f) are combined mutli-sensor data.

hydrophone-only deghost estimate for low frequencies. The wavefield separation technique is also only applicable to the receiver-side ghost, requiring a separate method for source-side only deghosting.

For conventional single-component streamer acquisition, Roberts *et al*. (2025) proposed an ML approach for deghosting combining a limited amount of real training data with “pseudo-synthetic” training data. The pseudo-synthetic training data were derived from real data through aggressive deghosting and spectral whitening. This data was assumed to be ghost-free and was then “ghosted” using wavefield continuation methods to simulate “ghosted” data at various depths. The idea being that applying a ghosting operator is much simpler than applying a deghosting operator (with issues around inexact water-depth estimates, handling singularities, etc.). This approach is also used here to estimate both the hydrophone and geophone responses. The geophone response is limited to the same low-cut filter as applied to the production data. Unlike conventional wavefield-decomposition, no additional work was done to generate a low-frequency (< 20Hz) response on the geophone. Two additional lines of real data were also used for training that underwent conventional wavefield separation and source-side deghosting.

The ML model used here is an adapted version of the DuckNet model (Dumitru *et al.*, 2023), in which the DuckNet block has been replaced with the ConvNext v2 block. The output layer has also been changed from a sigmoid to a linear activation function. The advantage of this architecture is that it preserves the multiscale approach of the DuckNet model and leverages the highly performant ConvNext v2 block. During training, a combined loss function of the Huber and Structural Similarity Index Measure was used.

To investigate the impact of the geophone on ML-based deghosting, two models were trained. A single-channel model was trained using just the hydrophone component. A second model was trained with two input channels with the geophone component added as the second channel. Figure 1 shows the training and validation losses for the two models using a common target deghosted dataset. The final validation loss using two-channels was 0.017 vs 0.025 for the single channel model. This shows that incorporating data from the geophone provides significant added value and leads to a clear enhancement in deghosting performance, which cannot be achieved by relying solely on the hydrophone.

Debubble and Designature

On a sail line by sail line basis, a wavelet is estimated from near to mid offsets using a minimum phase assumption. That wavelet is used to derive a debubble filter and zero phase filter which are applied to the sail line.

Results

This 3D survey acts as the main dataset to illustrate the proposed machine-learning acceleration of seismic pre-processing workflows. It is a large-scale acquisition located offshore along the northern coast of Brazil, covering a vast frontier exploration area of about 20,000 square kilometers. Positioned on a rifted margin, the region offers a challenging environment with extreme seafloor topographic variations and water depths between 1,500 and 4,000 meters. The bathymetry features a complex landscape of detailed channels, steep-walled canyons, and seamounts, which cause notable wavefront distortions and provide a testing ground for the robustness of processing flows.

Geologically, the subsurface is equally demanding due to the presence of significant igneous rocks, specifically high-velocity volcanic sills and intrusions embedded within the sedimentary section.

The seismic data was acquired using multi-sensor streamer technology to ensure high quality deghost. The configuration involved a spread of 12 streamer cables, with each cable containing 804 channels at a 12.5-meter group interval.

Figure 2 shows the progression of a single shot gather through the processing flow. The impact of the ML Deswell model is clear in Figure 2 (b). The detail captured by the ML deblend seed is clear in 2(c). Figure 3 shows stacked sections corresponding to 2 (a), 2(d) and 2(f). The stacked section corresponding to the complete flow clearly highlights the quality of the denoise and deblending flow along with the much improved focusing from the ML deghost and the automated debubble/designature flow.

Critically, through the careful and targeted use of ML, this complete workflow can be run on a sail-line by sail-line basis, and the total run time for each sail-line is measured in hours.

Conclusions

This study shows that machine learning can be combined with conventional signal-processing techniques to address the most computationally intensive stages of multi-sensor marine seismic preprocessing. The resulting workflow substantially reduces the manual parameterization effort inherent in conventional processing while maintaining high-fidelity output. Comparison tests between single- and multi-component deghosting confirm that dual-sensor input provides complementary information that measurably improves model performance beyond what is achievable from the hydrophone alone. Ongoing work focuses on incorporating additional training data to generalize the geophone denoising and multi-sensor deghosting models across varying geological settings. Applied to the large data volumes typical of frontier exploration, the workflow processes individual sail lines in hours, reducing the timeline for intensive preprocessing stages.

Acknowledgements

We thank TGS for permission to show the data and Leon Ge, Laura Alexander and Mahmoud Soliman for assistance with generating much of the training data.