ML deghosting for multi vintage marine seismic data: an augmented learning approach from the Krishna Godavari Basin, India

Graeme Stock, Olga Brusova, Mark Roberts, David Brookes, Amy Knowles, Sahil Mannick, Rebecca Yates, TGS*

Summary

Reprocessing multi-vintage marine seismic data to a consistent broadband standard remains a major technical and operational challenge, particularly when acquisition parameters vary significantly between surveys and when reliable acquisition auxiliary data are incomplete or unavailable. One of the most critical steps in modern marine seismic reprocessing is deghosting, where surface-related ghost reflections generated at the sea surface are removed to recover low- and high-frequency bandwidth, stabilize amplitudes, and improve temporal resolution. This abstract presents an augmented machine-learning (ML) strategy for deghosting legacy marine seismic data, demonstrated on a large multi-survey reprocessing project from the Krishna–Godavari (KG) Basin offshore India. The proposed workflow combines the robustness and efficiency of a generalized ML deghosting model with survey-specific adaptation derived from limited, high-quality deterministic deghosting results. The augmented approach delivers stable broadband spectra, improved seismic image quality, and a high degree of consistency across surveys acquired with differing tow depths and acquisition geometries, while maintaining favorable turnaround times for large reprocessing projects.

Introduction (Optional)

Large-scale reprocessing and merging of legacy marine seismic surveys has become increasingly common as exploration and development workflows demand broadband seismic datasets suitable for basin-scale interpretation, quantitative analysis, and prospect maturation. Such multi-vintage projects typically involve surveys acquired over long time periods, often spanning decades, during which acquisition technologies, operational practices, and survey design philosophies have evolved considerably. As a result, significant variations are commonly observed in source and receiver tow depths, streamer lengths, group intervals, recording bandwidth, and sampling rates.

The Krishna–Godavari Basin mega-merge project involved the reprocessing and integration of several legacy 3D marine seismic surveys acquired between 2004 and 2021 in the Bay of Bengal offshore India (Figure 1). The primary objective of the project was to generate a continuous, broadband seismic volume that maximized the value of the existing vintage datasets and enabled consistent basin-scale interpretation across fields, discoveries, and undrilled prospects. Achieving this objective required careful handling of spectral inconsistencies introduced by varying acquisition configurations, with deghosting identified as a key enabling process.

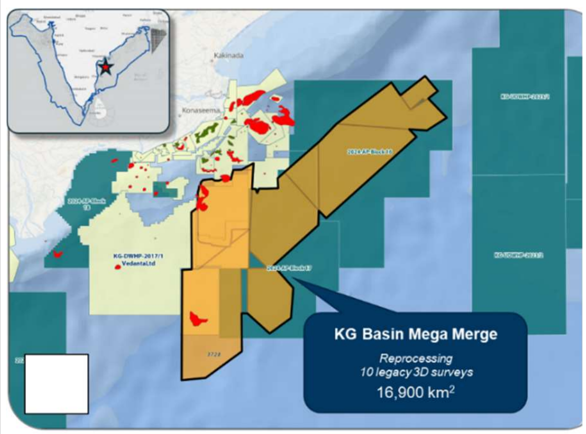

Figure 1: Map of the KG basin mega merge survey

Sea-surface ghosts introduce characteristic spectral notches and bandwidth loss that depend directly on source and receiver tow depths. When these depths vary between surveys, the resulting ghost notch frequencies differ, leading to pronounced spectral discontinuities in merged datasets if not adequately corrected. Conventional deterministic deghosting techniques, including sparse Radon or tau-p based approaches (Seher *et al*., 2021) can produce high-quality results when acquisition parameters are well known and stable. However, for vintage surveys, accurate depth information is often incomplete, uncertain, or spatially inconsistent, increasing the risk of residual ghosts, ringing artefacts, and spectral instability. These limitations motivate the exploration of alternative or complementary approaches that are less sensitive to auxiliary data uncertainty while remaining physically meaningful.

Augmented ML Deghosting Methodology

To address the challenges associated with deterministic deghosting of multi-vintage data, a ML based deghosting workflow was adopted (Roberts *et al*., 2025). The baseline

ML model was trained using a large library of pseudo-synthetic seismic data generated across a broad range of source and receiver depths, acquisition geometries, and noise conditions. This generalized training strategy allows the model to learn the underlying relationship between ghosted and deghosted wavefields without requiring precise knowledge of the true acquisition parameters at inference time. As a result, the model is inherently robust to uncertainties in auxiliary data and can be applied efficiently across multiple surveys.

While the general ML model delivers consistent and operationally robust deghosting results, experience shows that a single global model may not fully capture subtle survey-specific characteristics, particularly when acquisition configurations differ markedly. In such cases, the general model can exhibit spectral imbalances, most commonly in the form of high-frequency over-boosting or incomplete correction of ghost notch frequencies. To mitigate these effects, an augmentation stage is introduced into the workflow.

For each survey, a small and carefully selected subset of data is identified where deterministic 3D sparse tau-p deghosting produces physically consistent and high-quality results. These subsets are typically chosen in areas where acquisition conditions are close to nominal survey parameters and where deterministic deghosting is demonstrably stable. The deterministic outputs from these subsets are then used to further train, or fine-tune, the general ML model, resulting in a survey-specific augmented model. This process allows the ML model to adapt to the unique spectral and kinematic characteristics of each dataset while preserving the generalization capability learned from the broader training set.

The augmentation strategy is computationally efficient, as it requires deterministic processing of only a limited volume of data rather than full-survey application. In the KG Basin project, all ML deghosting was performed at a 4 ms sample rate to ensure consistent frequency content and facilitate inter-survey comparison. The resulting augmented models were then applied to the full extent of each survey prior to subsequent processing and merging steps.

Deghosting Results

The augmented ML deghosting workflow was evaluated on four representative surveys selected from the KG Basin mega-merge project. These surveys span a wide range of acquisition vintages and configurations, with source depths typically between 5 and 6 m and receiver depths ranging from approximately 6 to 10 m. This variability provides a robust test of the workflow's ability to deliver consistent broadband results across differing ghost characteristics.

For surveys acquired with relatively shallow source and receiver tow depths, ghost notch frequencies lie above the usable bandwidth at a 4 ms sampling rate, and clear notches are not readily visible in the input amplitude spectra. In these cases, both the general ML model and the augmented survey-specific models produce clear improvements in bandwidth, event continuity, and temporal resolution on shot gathers and stacked sections. However, spectral analysis reveals that the general model can introduce mild high-frequency over-enhancement, whereas the augmented models produce smoother and more physically plausible spectra. Figure 2 shows the results of the ML deghosting for a shallow tow survey. Input shot gathers (2a) are compared to the results using the general model (2b) and the augmented model (2c). Similar comparisons are made for stack sections, (2e, 2f and 2g). Amplitude spectra for the shots and stacks are shown in Figures 2d & 2h.

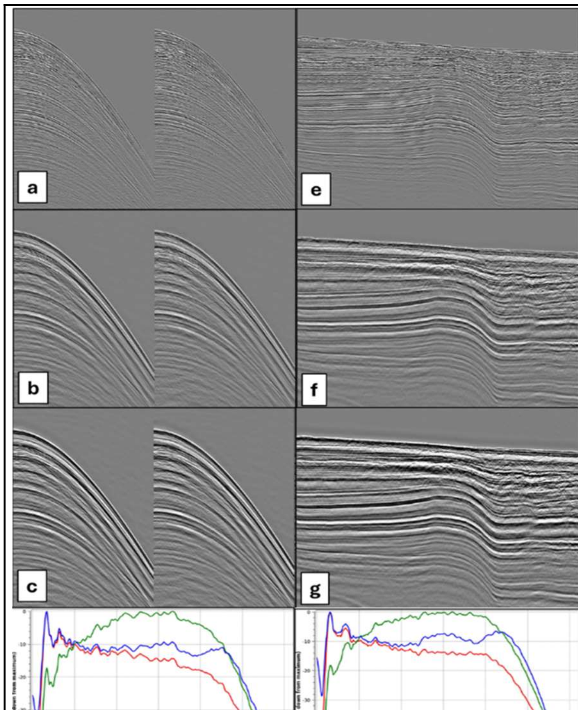

Figure 2: Deghost results for shallow tow survey 2: 2a - Input shot gathers. 2b - deghosted shots (general model). 2c - deghosted shots (augmented model). 2d - shot gather spectra. 2e - Input stack. 2f - deghosted stack (general model). 2g - deghosted stack (augmented model). 2h - stack spectra.

ML Deghosting for Multi-Vintage Marine Seismic Data: An Augmented Learning Approach from the Krishna–Godavari Basin, India

For surveys acquired with deeper receiver tow depths, distinct ghost notches are visible in the input spectra, typically around 70–80 Hz. Application of the general ML model partially corrects these notches but often leaves residual spectral imbalance and increased high-frequency instability. In contrast, the augmented ML models successfully reconstruct energy at the notch frequencies while maintaining stable spectral behavior across the full bandwidth. These improvements are reflected not only in the amplitude spectra but also in clearer reflector continuity and more consistent seismic character on stacked sections. Figure 3 shows the deghosting results for a deeper tow survey. Input shot gathers and stacks (3a and 3e) are compared to the results of the general model (3b and 3f) and the augmented model (3c and 3g). Amplitude spectra are shown in Figures 3d and 3h.

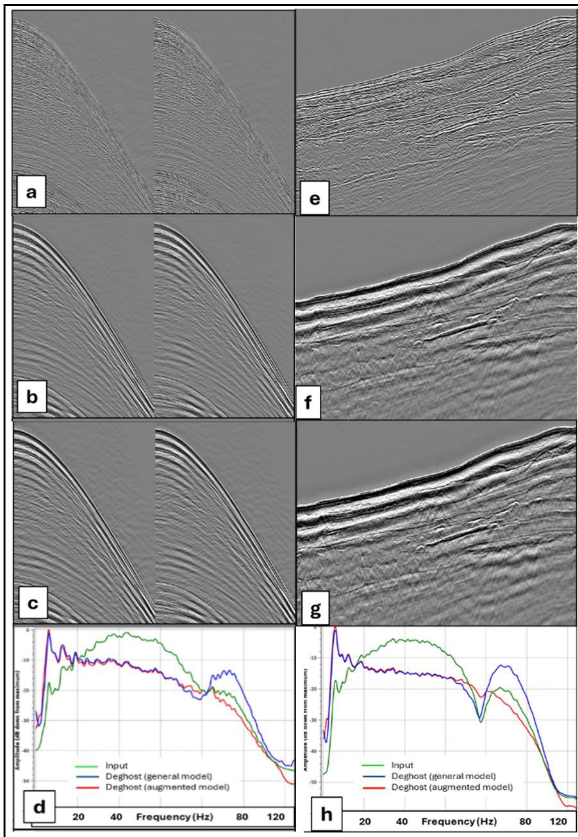

Figure 3: Deghost results for deeper tow survey 3: 3a - Input shot gathers. 3b - deghosted shots (general model). 3c - deghosted shots (augmented model). 3d - shot gather spectra. 3e - Input stack. 3f - deghosted stack (general model). 3g - deghosted stack (augmented model). 3h - stack spectra.

Survey Consistency and Quality Control

A critical requirement for multivintage merge projects is achieving spectral and amplitude consistency across all surveys. To assess this, an RGB (Red-Green-Blue) spectral quality control (QC) attribute (Stock *et al*., 2025) was used and derived from frequency decomposed stacked data. Three overlapping frequency bands, typically centered around the expected ghost notch frequency, are mapped to the red, green, and blue channels of an RGB image. Balanced spectra appear as neutral grey or white colors, while spectral bias manifests as dominant red or blue tones corresponding to low or highfrequency overrepresentation.

Figure 4 shows RGB maps created from 3D stacked data after merging the four surveys from each survey (1 to 4). Frequency intervals of 50-70 Hz, 60-80 Hz and 70-90 Hz were used in the spectral decomposition. The maps are shown for the input, deghosting by the general and augmented models with the positions of the four surveys shown. Prior to deghosting, the RGB maps of the merged KG Basin dataset exhibit strong color contrasts, particularly associated with the deeper towed survey (survey 4), indicating pronounced spectral inconsistency. Application of the general ML model improves overall consistency across the merged volume but reveals a tendency toward high frequency dominance, visible as blue or purple hues in the RGB display. The augmented ML deghosting results show the highest level of consistency, with largely neutral RGB responses both within individual surveys and across survey boundaries. This visual assessment is corroborated by averaged amplitude spectra, which demonstrate well balanced broadband behavior for all surveys following augmented deghosting. Figure 5 shows the average spectra of each survey.

Conclusions

Deghosting is a fundamental component of modern marine seismic reprocessing workflows, but it remains particularly challenging for multi-vintage projects where acquisition parameters vary, and auxiliary data quality is uncertain. Purely deterministic approaches can be sensitive to parameter errors and computationally expensive, while purely generalized ML approaches may suffer from spectral instability when applied across diverse datasets.

The augmented machine-learning strategy presented here combines the strengths of both approaches. By using limited, high-quality deterministic deghosting results to adapt a general ML model on a per-survey basis, it is possible to achieve stable, broadband deghosting that is consistent across surveys with differing acquisition characteristics. Applied to the Krishna–Godavari Basin mega-merge project, this workflow delivered improved spectral balance,

ML Deghosting for Multi-Vintage Marine Seismic Data: An Augmented Learning Approach from the Krishna–Godavari Basin, India

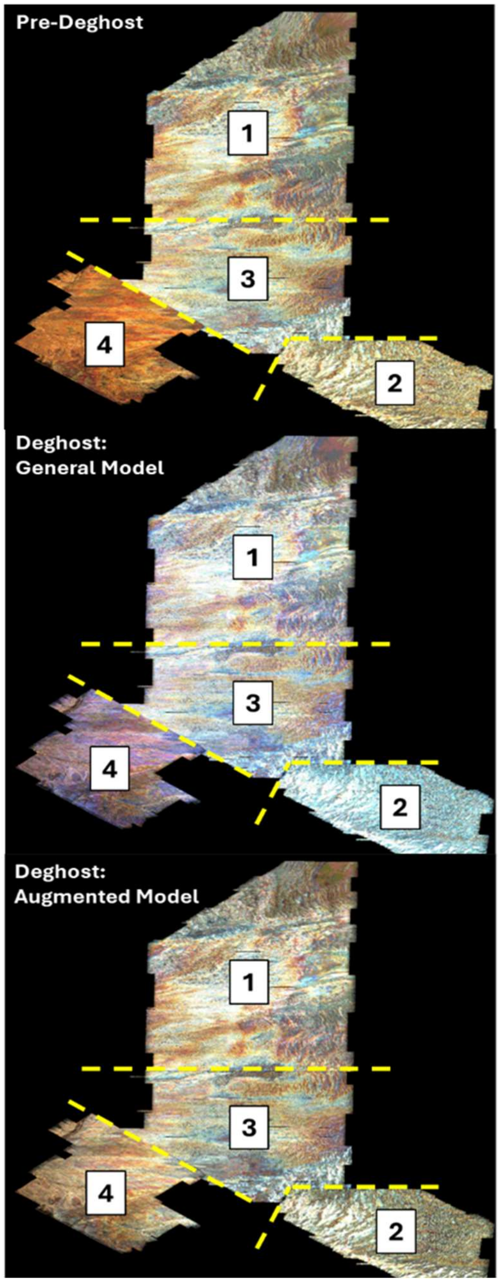

Figure 4: RGB maps of merged 3D stacks of the four surveys, (pre-deghosting, deghosting with general model and deghosting with augmented model)

enhanced seismic image quality, and efficient turnaround times suitable for large-scale reprocessing. Furthermore, the augmented training data generated during production contribute to the ongoing refinement of the baseline ML model, providing a continually improving foundation for future multi-vintage applications.

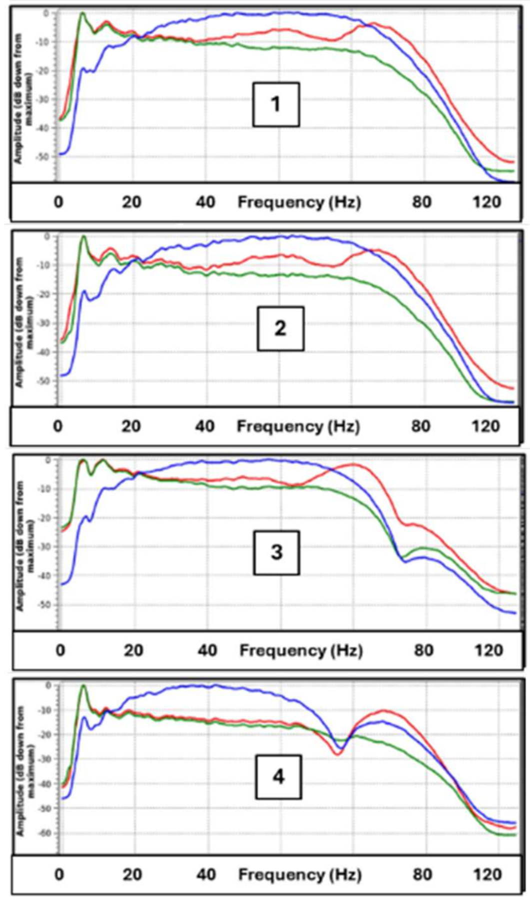

Figure 5: Average spectra of each survey (blue = pre-deghosting, red = deghosting with general model, green = deghosting with augmented model).

Acknowledgements

The authors thank the TGS Imaging team in Weybridge, UK, for their contributions to the processing, quality control, and delivery of the Krishna–Godavari Basin reprocessing project.