ML deghosting applied to multi-vintage datasets from the KG Basin, India – A Case Study

Introduction

The Krishna Godavari (KG) Basin mega merge survey involved the reprocessing of several vintage surveys through a modern processing sequence. The basin is located in the Bay of Bengal, offshore India. The survey area is shown in Figure 1a. The aim of the reprocessing was to create a continuous broadband seismic volume which maximised the value of the vintage surveys and allowed for existing fields and prospects to be tied on a basin wide scale.

A key aspect of the processing was deghosting vintage datasets acquired with varying and relatively shallow tow configurations. Deghosting is a key step in a modern processing sequence where source and receiver side down going reflections from the sea surface, which distort the image and decrease the bandwidth, are removed. Traditional methods use multidimensional transforms such as sparse Tau-p (Seher et al., 2021) and are largely deterministic requiring reasonably accurate source and receiver depth information. Such information can be challenging to acquire or extract from vintage surveys and inaccuracies can lead to artefacts such as ringing.

We describe an augmented machine learning approach to deghosting where a generalised model is updated for each survey from a deterministic sparse tau-p result. Such a method allows for a faster turnaround compared to deterministic approaches and facilitates consistent results across multi-survey projects. This consistency is demonstrated using an RGB QC attribute. The method is illustrated for a sample of four of the surveys, with varying acquisition parameters, from the KG Basin merge project as shown in Figure 1b. These will be referred to as surveys 1,2,3, and 4.

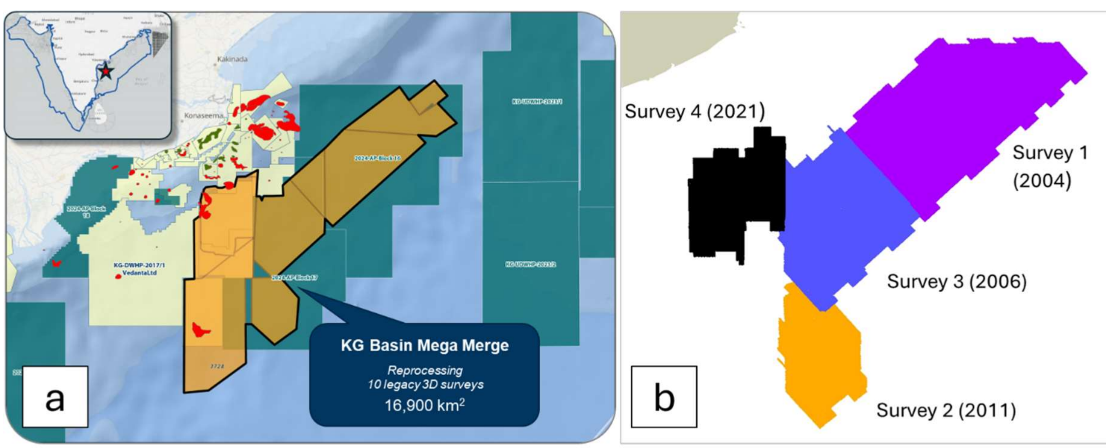

Figure 1a *Map of the KG basin mega merge survey.* ***1b*** *The four surveys used in the case study.*

Augmented Machine Learning Deghosting

Figure 2 shows the acquisition configurations of the four different surveys used in the case study and shows the variation in parameters, notably with source and receiver depths ranging from 5 – 6 m and 6 – 10 m respectively. There are further variations in parameters such as group interval and cable length.

A machine learning technique was employed (Roberts et al., 2025) which utilised a model trained on a large dataset of pseudo synthetically ghosted data generated with a wide range of source and receiver depths. Such a model can effectively eliminate the requirement for the precise knowledge of acquired source and receiver depths. However, this model was augmented by further training using 3D deghosting results obtained from each of the vintage data themselves. A specific augmented model is created for each survey in turn. The deghosting results were achieved using a deterministic 3D sparse tau-p algorithm applied to a small subset of data from each survey where results indicated source and receiver depths were close to the nominal survey values and the deghosting was considered optimal. Deghosting was performed at 4 ms sample rate.

Survey	1	2	3	4
Year	2004	2011	2005/2006	2021
Shooting Direction	137 / 317 degrees	138 / 318 degrees	137 / 317 degrees	0 / 180 degrees
Shot Point Interval	25m	37.5m	25m	25m
Number of gun arrays	2	2	2	2
Total gun array separation	50m	62.5m	50m	37.5m
Source depth	5m	6m	5m	6m
Cable depth	6m	8m	6m	10m
Cable length	6000m	8100m	6000m	8100m
Number of cables	8	10	8	10
Distance between cables	100m	125m	100m	100m
Number of channels per cable	478	648	480	648
Group Interval	12.5m	12.5m	12.5m	12.5m
Acquisition Trace Length	10240ms	13200ms	9216ms	10240ms

Figure 2 *Table showing the acquisition parameters of the four surveys.*

Deghosting results

Figure 3 shows the results of the ML deghosting for survey 3. Input shot gathers (3a) are compared to the results using the general model (3b) and the augmented model (3c). Similar comparisons are made for stack sections, (3e, 3f and 3g). Amplitude spectra for the shots and stacks are shown in Figures 3d & 3h. This survey was acquired with shallow source and receiver depths of 5m and 6m respectively, so the ghost notches are not seen in the spectra at 4 ms sample rate. Both ML models produce good deghosting results in the seismic displays however with the general model the spectra reveal some overboosting of high frequencies contrasting with the augmented model results which show stable spectra.

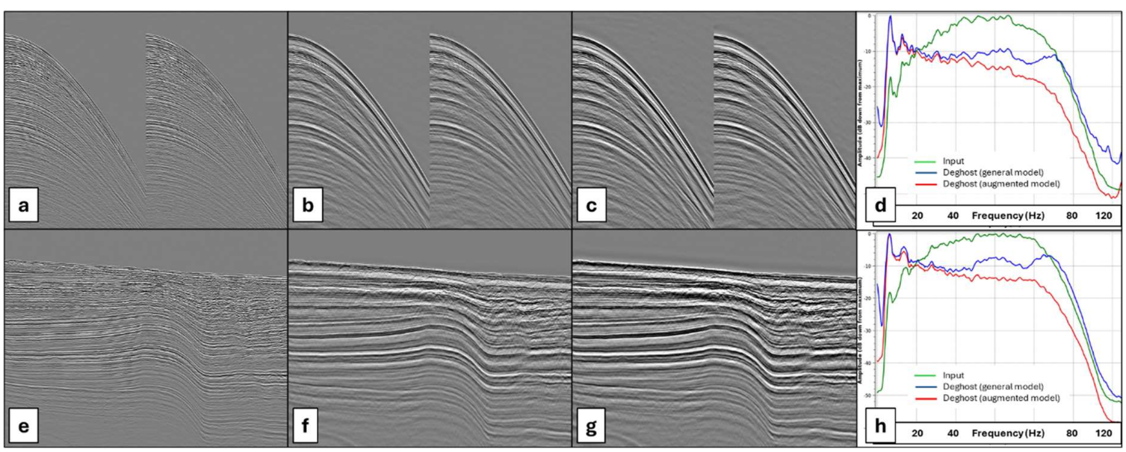

Figure 3 *Deghost results for survey 3: 3a - Input shot gathers. 3b - deghosted shots (general model). 3c - deghosted shots (augmented model). 3d - shot gather spectra. 3e - Input stack. 3f - deghosted stack (general model). 3g - deghosted stack (augmented model). 3h - stack spectra.*

Figure 4 shows the survey 4 deghosting results. Input shot gathers and stacks (4a and 4e) are compared to the results of the general model (4b and 4f) and the augmented model (4c and 4g). Amplitude spectra are shown in Figures 4d and 4h. This survey was acquired with a deeper receiver depth (10 m) so the notch is now visible in the spectra at ~ 75 Hz. Again, whilst the general model provides a reasonable deghosting result it doesn't fill the notch, and some over boosting of higher frequencies is observed above 80 Hz. The augmented model provides good results with stable spectra and the notch frequencies corrected. It can be seen that augmenting the general model provides a successful approach to the machine learning deghosting.

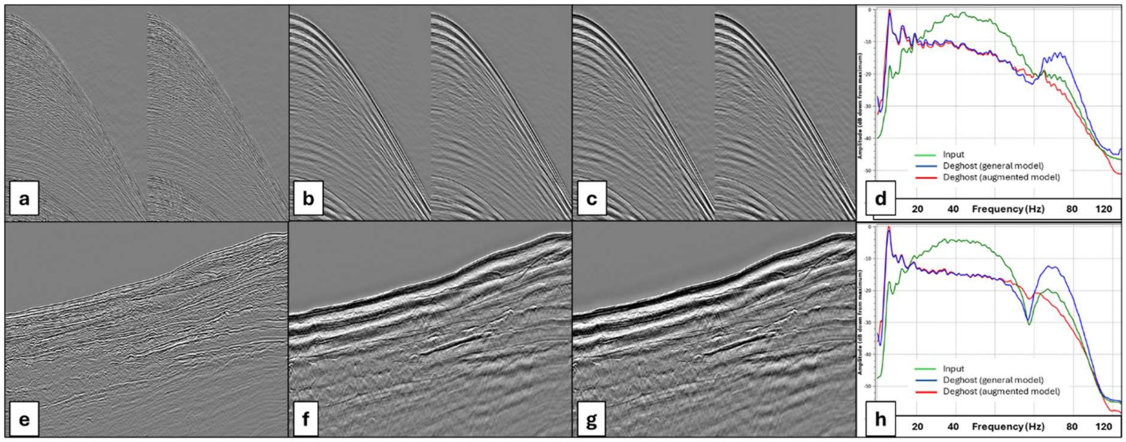

Figure 4 *Deghost results for survey 4: 4a - Input shot gathers. 4b - deghosted shots (general model). 4c - deghosted shots (augmented model). 4d - shot gather spectra. 4e - Input stack. 4f - deghosted stack (general model). 4g - deghosted stack (augmented model). 4h - stack spectra.*

Survey consistency and QC

The consistency of the deghosting result across all the different vintages can be demonstrated by using an RGB quality attribute (Stock et al., 2025). An RGB map is created from a frequency decomposition of stacked data to assess the relative amplitudes over three frequency intervals, usually centred around the notch frequency which are then encoded as an RGB colour value. By mapping this attribute any instabilities or variations in the performance of the deghosting that may relate to changes in the acquisition parameters such as source and receiver depths can be inferred. White or grey colours indicate balanced spectra about the notch while dominant red or blue colours show low or high frequencies being over boosted.

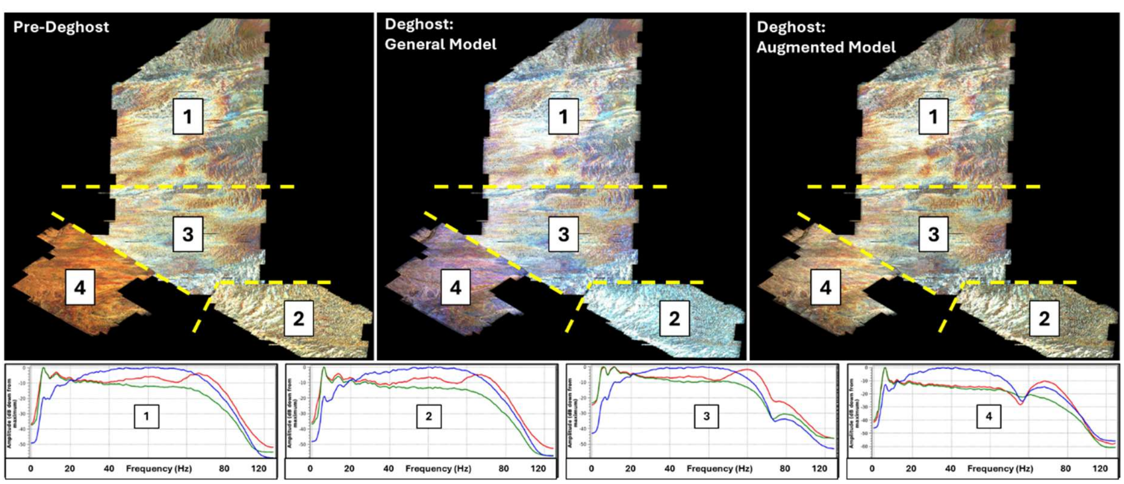

Figure 5 *RGB maps of merged 3D stacks of the four surveys, (pre-deghosting, deghosting with general model and deghosting with augmented model) along with averaged spectra of each survey (blue = pre-deghosting, red = deghosting with general model, green = deghosting with augmented model).*

Figure 5 shows RGB maps created from 3D stacked data after merging the four surveys along with averaged spectra 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. Survey 4 with its deeper receiver depth stands out in the RGB map pre-deghosting with the predominantly red colour indicating a bias towards lower frequencies due to the notch at ~ 75 Hz. The RGB map of the deghosting result with the generalised model shows more consistency across the surveys however the predominance of blue and purple colours indicates an over boosting of high frequencies as can be seen in the spectra. The map of the augmented model result shows good consistency both within each survey and across all four surveys. This is confirmed by the amplitude spectra.

Conclusions

Deghosting is a principal component of any modern processing sequence, but it can be challenging to apply to vintage surveys in reprocessing projects, particularly if traditional deterministic methods are used. A machine learning approach can overcome these challenges by training a suitably general model. Such models can provide good deghosting results but can suffer from inaccuracies and instabilities particularly at higher frequencies. However, it has been shown that these shortcomings can be overcome by augmenting a general model with carefully selected deghosting results produced by a deterministic process and specific to each survey. Such results only need to be produced for small subsets of data. This makes the augmented approach to machine learning deghosting an effective solution to processing multi vintage reprocessing projects. Applied to the KG Basin project it enabled a deghosting result to be produced which was cost and time efficient and consistent in quality throughout all surveys. It should also be noted that by augmenting the training data in the production processing we can also improve the training data for the general model. This then provides an improved baseline model for use in future projects.

Acknowledgements

The authors would like to thank the TGS imaging team in Weybridge, UK, for their work on the processing of the KG basin, India project.

References

Roberts M., Brusova O., Brookes D., Baldock S., Valenciano A. [2025]: Accelerating marine seismic preprocessing with machine learning, *86th EAGE Annual Conference and Exhibition, June 2025, Volume 2025, p1-5.*

Seher, T., Masoomzadeh, H., Yang, C., Dong, S., Drewell, S., Zhang, Z., [2021]: Variable depth streamer deghosting using a linear Radon transform with a time domain sparsity constraint, *82nd EAGE Annual Conference & Exhibition, October 2021, Volume 2021, p. 1-5.*

Stock G., Brookes D., Mitchell C., Yates R., Raffle J. [2025]: Utilising map based process QC attributes to optimise 3D seismic processing flows offshore Nigeria, *86th EAGE Annual Conference and Exhibition, June 2025, Volume 2025, p1-5.*