EAGE

D1-6

What is our a Priori Information? An Initial Attempt at Generating Representative Earth Models

M. Roberts[1], O. Brusova[1], K. Gonzalez[1], A. Valenciano[1]

TGS[1]

Summary

Summary is not available

Introduction

Elastic FWI (Tarantola, 1986) is currently the state-of-the-art method for recovering detailed elastic models of the sub-surface from seismic data. Applications include Timelapse monitoring (Zheng et al., 2013) and pore-pressure prediction (Roberts et al., 2007). A key issue is that Elastic FWI is an under-constrained inverse problem. Tarantola (1986) framed FWI as a Bayesian inverse problem enabling prior information to be captured in a model covariance matrix. The big drawback of such approaches is the assumption of Gaussian distribution. While a reasonable approximation for a single lithology, typically, this is unknown a priori. Recently, deep learning has presented several techniques for capturing prior information in deep neural networks and has been used to help constrain the inverse problem. One promising approach is using diffusion networks to capture structural and petrophysical priors (Taufik *et al*, 2024). OpenFWI (Deng *et al*, 2022) provides a useful synthetic test dataset for ML velocity estimation. However, OpenFWI does not provide a realistic *a priori* distribution of the earth model. This is the problem we solve in this paper by working with a global dataset of 60,090 offshore wells. In this dataset, there were a mere 3,101 shear wave sonic logs. Firstly, we train a model to estimate the shear wave sonic log from other logs and run inference on the complete set of offshore wells to provide a more representative and global dataset. Secondly, we extract structural information from 3D seismic volumes. Thirdly, the Vp, Vs, and density logs from a random well are combined with the structural information from the seismic volume to generate a representative 3D earth model. This can be repeated for many well and seismic combinations to generate a rich dataset of representative models.

Well Data

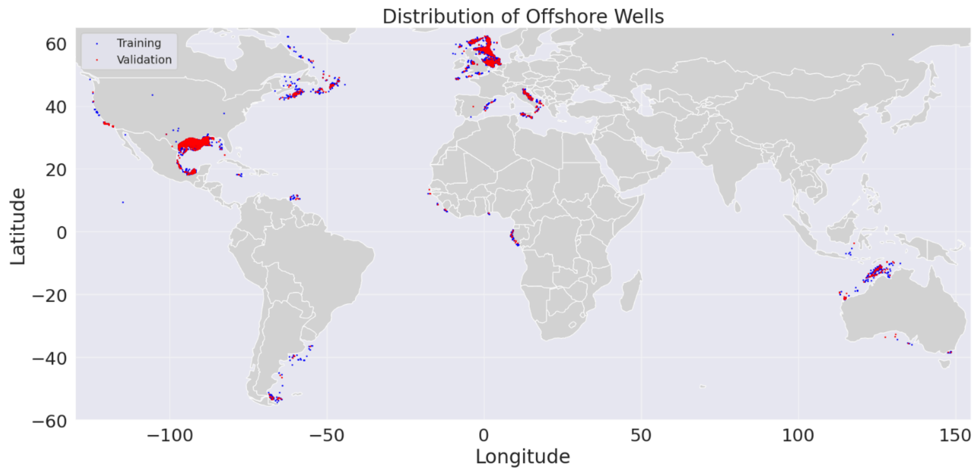

Figure 1 *Global distribution of training (blue) and validation (red) wells.*

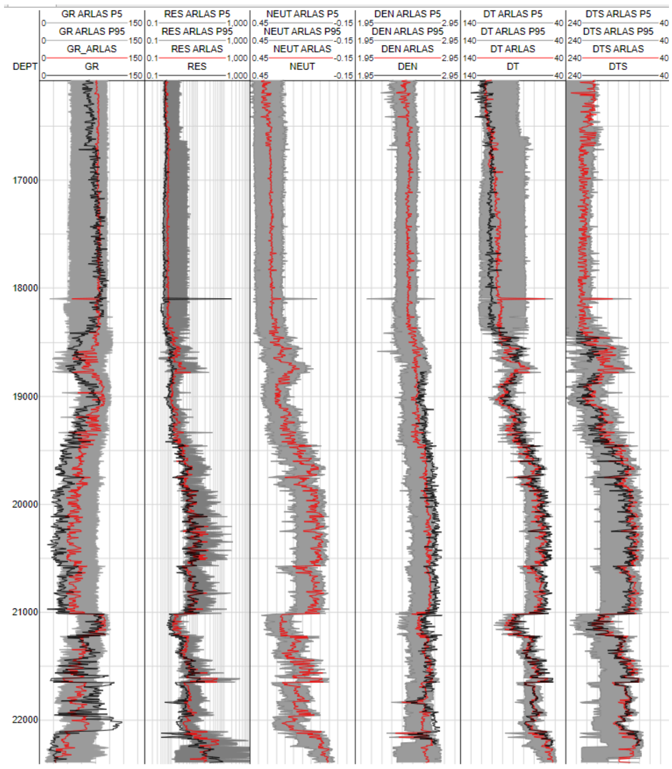

Figure 2 *SEQ Figure * ARABIC 2 Example of ARLAS predictions (red curves) as compared to measured data (black curved). The plot shows: (1) gamma ray, (2)deep resistivity, (3) neutron porosity, (4) bulk density, (5) compressional slowness, (6) shear slowness. P5-P95 interval in grey.*

Our dataset consists of 60,090 offshore wells from different basins around the world. The location of the wells used is shown in Figure 1. The dataset exhibits an uneven distribution of wells across global regions, with the Gulf of Mexico accounting for the majority (88%). While almost half of the wells contain acoustic measurements, only 5% include shear wave information. The scarcity of shear wave data underscores the need for reliable prediction methods. We use 75 % of all data for model training and the remaining 15% for model validation. Training and validation data is randomly sampled from different global basins. Figure 1 shows the distribution of training and validation wells globally.

Before machine learning, the raw well-log data underwent rigorous cleaning and quality control processes. These steps were essential for ensuring data accuracy and consistency and for reliable model training and inference. The clean-up workflow included automated steps like log categorization, depth alignment, normalization, and removal of low-quality data. A key innovation in the clean-up process involves the introduction of synthetic curve comparisons for sonic log classification. This method uses equations to calculate synthetic compressional and shear slowness logs from resistivity data. Comparing actual sonic logs to these synthetic models provides a robust and automated method for classifying sonic curves, even when log names and descriptions are unreliable.

The core of the ARLAS methodology (Gonzalez *et al.* 2023) is the Gradient Boosting Tree (GBT) machine learning model. GBT is an ensemble learning approach that builds a predictive model by sequentially combining multiple decision trees, iteratively minimizing the error between predictions and actual values. This technique excels at handling complex data patterns and achieving high prediction accuracy.

The ARLAS model leverages various features, including compressional sonic logs, gamma-ray logs, resistivity logs, neutron porosity, bulk density, true vertical depth, and well-location data. This multi-dimensional approach allows the model to capture complex relationships between log properties and their spatial variation. Here, we extend the method to include shear logs and offshore wells.

The trained ARLAS model was tested on a held-out dataset, demonstrating good agreement with measured data in areas with ample training data. However, challenges arise in shallow sections and regions with limited data availability, highlighting the need for further model refinement. P5 and P95 curves (representing the 5th and 95th percentiles of the model predictions) were generated to assess model uncertainty and reliability. These bounds provide a range within which the true value is expected to lie with 90% confidence. Large discrepancies between P5 and P95 indicate potential overfitting and lower confidence in the predictions.

3D model generation

The approach taken here for extracting structural information from seismic data follows the work of Fomel (2010). "Plane-wave destruction" leverages the concept of local plane waves, modeled as simple linear equations representing seismic events with varying slopes. The technique involves predicting each trace in a seismic section from its neighbor based on estimated slopes and subtracting the prediction from the original trace, resulting in a residual representing non-planar wave components. This process uses a local operator to propagate each trace along estimated dominant slopes, determined by minimizing the prediction residual through regularized least-squares optimization. The technique can be extended to 3D by applying it to both inline and crossline directions and is used here. "Predictive painting" then takes this a step further, allowing a reference trace to be recursively propagated to distant neighbors, following the local structure of seismic events, essentially "painting" the information from the reference trace onto neighboring traces. This process allows for a more global understanding of the seismic data structure.

By applying predictive painting to a reference trace containing only time values, we can derive a "relative geologic age" attribute. This attribute, as defined by Stark (2004), represents the time shift between a given trace and the reference trace, effectively indicating how much older or younger the geology at that trace is compared to the reference point.

To generate a random yet representative 3D elastic model based off the relative geological time, a well is chosen at random from the global dataset. Several QCs are made to ensure the log has a long enough section of usable data to generate a meaningful representative model (length, absence of salt). If selected, the logs are filtered with a smoothing operator (the smoothing length is also randomly chosen) to generate models suitable for FWI models generated with different frequency sources. It is assumed that FWI. Figure 3 outlines the complete flow for one seismic image. The top left shows a patch of the image taken from the data library, followed by the structural dip and resulting relative age volume. The bottom row contains the models taken after a lookup from the well data.

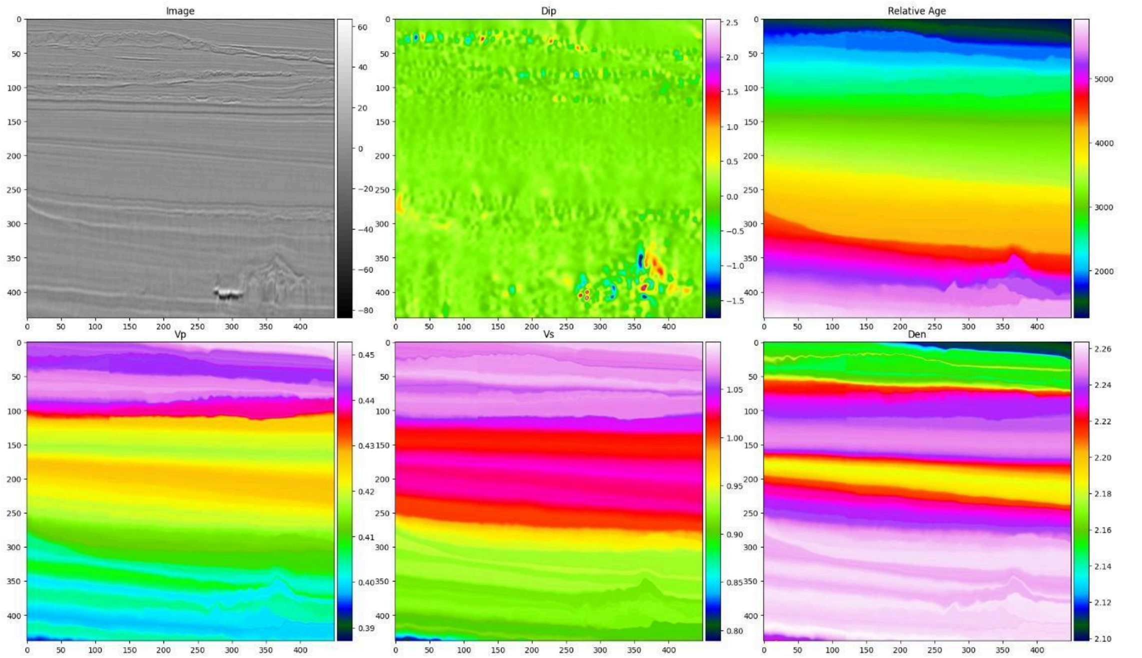

Figure 3 *This shows the progression from a stacked seismic image to a representative Vp, Vs, and density volume. Top row: seismic image, estimated dip, relative geological time. Bottom row: Vp, Vs, and density volumes.*

Future Work

We are currently developing a large number of representative geological models for a specific structural area. This effort will be expanded to encompass other areas in the future. The seismic data showcased in this work includes an igneous sill, a geological feature not currently captured in our modeling process. While Roberts et al. (2024) have demonstrated the ability to generate robust 3D models for salt formations, further research is needed to develop models for other geobodies, such as volcanics. This advancement would allow us to create even more realistic geological models.

Conclusions

This work presents a novel strategy for predicting continuous Vp, Vs, and density logs for offshore wells, significantly expanding the availability of crucial well log data. We demonstrated the feasibility of generating realistic elastic models directly from seismic images and a limited set of well logs. These models hold significant potential to enhance downstream applications, such as regularizing Full Waveform Inversion (FWI) using diffusion models trained on this synthetic data. This approach offers a promising path toward more accurate and reliable subsurface characterization, paving the way for improved exploration, reservoir management, and production strategies in offshore environments.

Acknowledgments

We are thankful to TGS management for permission to show the data. The authors thank Burak Bitlis and Seet Li Yong for their help in preparing the seismic data.

References

Deng, C., Feng, S., Wang, H., Zhang, X., Jin, P., Feng, Y., Zeng, Q., Chen, Y., & Lin, Y. (2022). OPENFWI: Large-scale Multi-structural Benchmark Datasets for Full Waveform Inversion. *Advances in Neural Information Processing Systems*, 6007–6020.

Fomel, S., (2010) Predictive painting of 3-D seismic volumes. Geophysics, 75, no. 4, A25-A30

Gonzalez, K., Brusova, O., & Valenciano, A. (2023). A machine learning workflow for log data prediction at the basin scale. *First Break*, *41*(2), 73–80. https://doi.org/10.3997/1365-2397.fb2023015

A. Roberts, M., & E. Hornby, B. (2007). Application of 2D Full Waveform Inversion to Walkaway VSP Data for the Estimation of Sub-Salt Elastic Parameters. *69th EAGE Conference and Exhibition Incorporating SPE EUROPEC 2007*, cp-27-00113. https://doi.org/10.3997/2214-4609.201401531

Roberts, M., Warren, C., Lasscock, B., & Valenciano, A. (2024). A Comparative Study of the Application of 2D and 3D CNNs for Salt Segmentation. *85th EAGE Annual Conference & Exhibition (Including the Workshop Programme)*, *2024*(1), 1–5. https://doi.org/10.3997/2214-4609.202410573

Stark, T. J. (2004). Relative geologic time (age) volumes—Relating every seismic sample to a geologically reasonable horizon. *The Leading Edge*, *23*(9), 928–932. https://doi.org/10.1190/1.1803505

Tarantola, A. (1986). A strategy for nonlinear elastic inversion of seismic reflection data. *Geophysics*, *51*(10), 1893–1903.

Taufik, M. H., Wang, F., & Alkhalifah, T. (2024). Learned Regularizations for Multi‐Parameter Elastic Full Waveform Inversion Using Diffusion Models. *Journal of Geophysical Research: Machine Learning and Computation*, *1*(1). https://doi.org/10.1029/2024JH000125

Zheng, Y.Y., Barton, P.J., Singh, S.C. (2013) Time-lapse Waveform Inversion from a Compacting Reservoir 75th EAGE Conference and Exhibition