ML-Based Imputation of Missing DTS (Sonic Shear Slowness) Curves: WAID Case Study

Introduction

Shear sonic slowness (DTS) plays a fundamental role in subsurface characterization, particularly in reservoir evaluation and geomechanical analysis. It is essential for estimating dynamic elastic properties such as shear modulus, Poisson's ratio, and Young's modulus, which are critical inputs for coupled flow–geomechanics simulations, wellbore stability analysis, and hydraulic fracturing design. Despite its importance, DTS data are often incomplete or entirely missing in well logs due to operational constraints, high acquisition costs, or tool malfunctions during logging operations. This lack of reliable shear wave measurements introduces uncertainty in reservoir models and limits the accuracy of geomechanical interpretations.

To address this challenge, recent advances in data-driven techniques have opened new opportunities for reconstructing missing well log data. Machine learning methods, in particular, have demonstrated strong capabilities in capturing complex nonlinear relationships between different petrophysical logs. By leveraging readily available measurements such as compressional sonic slowness (DTC), density (DEN), gamma ray (GR), neutron porosity (NEU), nuclear magnetic resonance (NMR) parameters, photoelectric factor (PE), and resistivity logs, it becomes possible to predict DTS with reasonable accuracy.

In this context, this study proposes a machine learning–based framework to impute missing DTS logs using multi-well petrophysical data from the publicly available WAID dataset. A Random Forest regression model is developed and evaluated using a robust Leave-One-Well-Out cross-validation strategy to ensure generalization across different wells. The proposed approach not only reduces the dependency on costly re-logging operations but also provides a practical and reliable solution for enhancing data completeness in reservoir characterization workflows.

Method

Exploratory data analysis (EDA) was conducted to examine data quality and inter-log relationships, as shown in Figure 1. Input logs were preprocessed, cleaned, and selected based on correlation and physical relevance. A Random Forest model was then trained to predict DTS:

$$DTS_i = f(X_i) + \varepsilon_i$$

Where $f(X_i)$ is the nonlinear regression function approximated by the Random Forest model, and ε_i is the residual error.

Model performance was evaluated using a Leave-One-Well-Out (LOWO) validation strategy.

Results

The RF model demonstrated stable predictive performance across validation wells. The RMSE ranged from 4.66 to 7.78 µs/ft, MAE ranged from 3.71 to 6.18 µs/ft, and R^2 ranged from 0.71 to 0.74. These results indicate robust generalization across wells with varying geological conditions.

After validation, the trained model was applied to reconstruct the missing DTS curve in the fifth well. The predicted DTS profile exhibited smooth and geologically consistent trends aligned with neighboring wells.

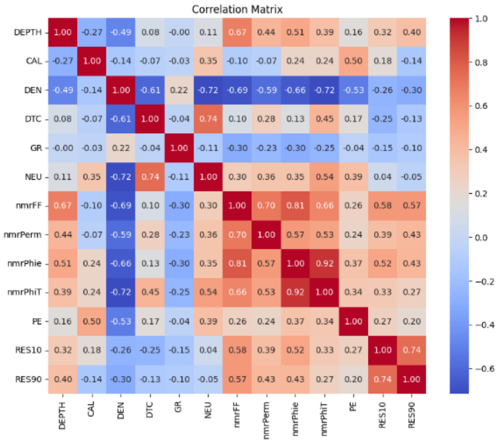

Figure 1 *Correlation Matrix of Petrophysical Logs Used for DTS Prediction.*

The predictive performance of the Random Forest model across all validation wells is summarized in Table 1.

Table 1 *Performance metrics of the Random Forest model under Leave-One-Well-Out validation.*

Well	RMSE (µs/ft)	MAE (µs/ft)	R^2
Well 1	7.784424	6.176733	0.710195
Well 2	6.980135	5.490959	0.742080
Well 3	6.361526	5.288509	0.744982
Well 4	4.659781	3.706762	0.722355
Average	6.4464665	5.1657408	0.729903

Conclusions

This study demonstrates that Random Forest regression can accurately impute missing shear sonic slowness logs using multi-well petrophysical data. The Leave-One-Well-Out validation ensures robustness and cross-well generalization. The proposed methodology provides a practical and cost-effective alternative to re-acquisition of DTS measurements and can be extended to other incomplete well-log datasets used in subsurface characterization.

References

Qu, F., Liao, H., Liu, J., Wu, T., Shi, F., & Xu, Y. (2024). A novel well log data imputation methods with CGAN and swarm intelligence optimization. *Energy*, *293*, 130694. https://doi.org/10.1016/j.energy.2024.130694

Al-Fakih, A., Koeshidayatullah, A., Mukerji, T., Al-Azani, S., & Kaka, S. I. (2025). Well log data generation and imputation using sequence based generative adversarial networks. *Scientific Reports*, *15*(1), 11000. https://doi.org/10.1038/s41598-025-95709-0

Hallam, A., Mukherjee, D., & Chassagne, R. (2022). Multivariate imputation via chained equations for elastic well log imputation and prediction. *Applied Computing and Geosciences*, *14*, 100083. https://doi.org/10.1016/j.acags.2022.100083