Submission code: ZLX6AYJML7

See this and other abstracts on our website: https://home.sbgf.org.br/Pages/resumos.php

Density model building: A comparison between geostatistical and machine learning methods for a regional case study

Frank Cenci Bulhoes (Petrobras), Mario Florencio Paiva (SLB), Gleidson Ferreira (Petrobras), Alex Passamani (Petrobras)

Density model building: A comparison between geostatistical and machine learning methods for a regional case study

Please, do not insert author names in your submission PDF file

Abstract

In a regional context, detailed velocity models for domain conversion can be constructed using velocity-density transformation alongside inverted gravimetric data, even when information is limited. Such studies often rely on 2D seismic data, while 3D surveys typically cover only small areas, leaving gaps in data coverage. Gravimetric surveys become crucial as they can cover regions lacking seismic data, providing valuable density information for regional studies. Although density models can be created through a 3D grid using geostatistical methods, they require precise variogram parametrization, necessitating an experienced modeler. Machine learning methods offer an alternative, making the process accessible to geoscientists unfamiliar with geostatistics. This study compares workflows based on both methodologies, discussing their advantages, disadvantages, and potential impacts on project timelines.

Introduction

Density models typically involve creating a 3D grid divided into geological zones, populated with spatial cells filled with modeled data (Bulhoes et al., 2015). Geostatistical methods like ordinary kriging enable the interpolation of nearby information to fill these cells where data is lacking, honoring spatial correlations. This allows for the quantification and prediction of unsampled locations through linear unbiased estimation, minimizing error variance and enhancing predictive confidence. However, the assumptions of stationarity and isotropy must be satisfied (Oliver *et al*., 2015).

While kriging is effective for datasets with strong spatial autocorrelation, it requires careful variogram modeling and is sensitive to data quality. Creating a 3D grid can be time-consuming, and optimal kriging parameters depend on data characteristics, posing challenges for non-geomodellers (Pyrcz *et al*., 2014).

Conversely, machine learning techniques, increasingly accessible through commercial platforms, utilize various algorithms, such as convolutional neural networks and tree-based models, to identify patterns in data. These methods offer flexibility in modeling complex relationships and are generally more user-friendly, although they may require more data and computational resources, with uncertainty quantification not always directly provided.

Methodology

This study was conducted in an offshore area of the Campos Basin, northeastern Rio de Janeiro, Brazil, where water depths range from 50 m to approximately 2200 m. The project includes 220 wells across various geological settings, featuring Turonian turbidites (Carapebus Formation) and Albian carbonates (Quissamã Formation) in the post-salt section, as well as Aptian carbonates (Lagoa Feia Formation) in the pre-salt section (Winter *et al*. 2007).

For geostatistical modeling, the spatial covariance structure of density from well logs was analyzed and fitted using a variogram. This allowed for the interpolation of values at unsampled points with weights derived from the covariance structure. Density logs from all 220 wells were upscaled into a 3D grid skeleton divided into nine zones, as outlined in Table 01. Given the established relationship between velocity and density (Gardner *et al*., 1974), seismic interval velocity volumes served as a secondary variable in a cokriging method to enhance density property interpolation, even in data-scarce areas (Azevedo & Soares, 2017).

For the machine learning density model, we employed the Fast Tree method, an efficient decision tree model based on gradient boosting (Ke *et al*., 2017). This model utilized seismic interval velocity from select wells to train the density log, estimating it for all available wells. Once the model reached the desired accuracy, it was used to estimate volumetric density properties directly from the interval velocity volume, avoiding the need for a 3D grid. Additionally, measured density logs were conditioned to estimate the shallow non-logged sections, following the approach described by Brocher (2005).

Results

Model results for kriging interpolation and machine learning methods can be inspected in Figure 01.A and Figure 01.B, respectively, display a cross-section of the studied area. According to Figure 02, extracted logs from each model are displayed jointly with measured density and corresponding filtered version for a few wells.

ZONE	SURFACES	CORRELATION	ZONE	SURFACES	CORRELATION
01	SEABED - MAZUL	Kriging (r & r^2) = 0.96/0.93 ML (r & r^2) = 0.91/0.82	06	TOP SALT – BASE SALT	
02	MAZUL - CRETACEOUS	Kriging (r & r^2) = 0.94/0.90 ML (r & r^2) = 0.75/0.56	07	BASE SALT - DPA	Kriging (r & r^2) = 0.94/0.88 ML (r & r^2) = 0.81/0.66
03	CRETACEOUS - TURONIAN	Kriging (r & r^2) = 0.92/0.84 ML (r & r^2) = 0.86/0.75	08	DPA - DPJ	Kriging (r & r^2) = 0.96/0.92 ML (r & r^2) = 0.91/0.84
04	TURONIAN - ALBIAN	Kriging (r & r^2) = 0.95/0.91 ML (r & r^2) = 0.83/0.69	09	DPJ - BASEMENT	Kriging (r & r^2) = 0.97/0.95 ML (r & r^2) = 0.94/0.88
05	ALBIAN – TOP SALT	Kriging (r & r^2) = 0.96/0.93 ML (r & r^2) = 0.85/0.72			

Table 01: Structure of 3D grid presented the delimitation of each zone and the correlation between measured density log *versus* kriging density log and measured density log *versus* ML density log for each zone.

As a specific quality control factor for kriging interpolation, the histogram of data distribution displays that density property follows the general trend of the input data (well logs), indicating a reasonably controlled interpolation process, although some artifacts are visible. Also, kriging propagated the property across zones without overburden control, which was highly structurally controlled instead of by depth. In turn, the ML model presents a precise overburden control displaying smoother results where the presence of artifacts was minimized (note that the model is strongly influenced by interval velocity, which is controlled by overburden). Also, metrics from the trained model can be analyzed in terms of how good the model predicts the data. R squared, Mean Absolute error (MAE), and Root-mean squared error (RMSE) should be enough to evaluate the efficiency of a model driven by data (Boutayeb et al., 2025). Thus, the final trained model achieved $R^2 = 0.92$, MAE = 0.04g/cm³, and RMSE = 0.07g/cm³, which indicates the model's high confidence in describing the relationship between its variables with minimal error. Lastly, the correlation coefficient (R) and the determination coefficient (R^2) were calculated between the measured density versus the kriging density model and ML density model extracted in each well position. These parameters were calculated for each zone created in the 3D grid for all 220 wells available, and it can be inspected following the scheme of Table 01 in the correlation column.

Analyzing the well response of models, kriging density logs present higher values for R & R^2 than ML density logs for each analyzed zone - salt layer assumed halite a constant value of 2.1 g/cm³ for both models, according to Yamamoto *et al*. (2019). As in the kriging model, all 220 wells were used to propagate density property, explaining the high correlation because geostatistical interpolation methods depend on a high amount of data to increase accuracy (Azevedo & Soares,

2017) while ML methods do not require it. Ingesting a high quantity of data will overfit the model, generating a biased result. About 5% of wells were used here, and representative ones were chosen regarding water column depth and main target. ML results are generally smaller than kriging; however, they present high accuracy and can reasonably represent the relationship between velocity and density.

Conclusions

Both models were developed from a regional perspective to define the main density trend distribution across a large area, anticipating a wide range of values. Kriging interpolation is well-suited for datasets with clear spatial correlations but requires expertise to develop a well-defined variogram, which can be complex. In contrast, machine learning methods are flexible and user-friendly, effectively managing large amounts of complex data, making them accessible to geoscientists. However, they can be computationally demanding, especially with extensive datasets. Ultimately, the choice of method depends on the specific problem and data characteristics. In this case, the ML model effectively captured depositional aspects, with distribution influenced by the overburden, yielding reliable results with less effort.

Acknowledgments

The authors are Grateful to Petrobras – Petróleo Brasileiro for the provision of all data used in this research as well as to SLB for further resources and software access for the development of this study.

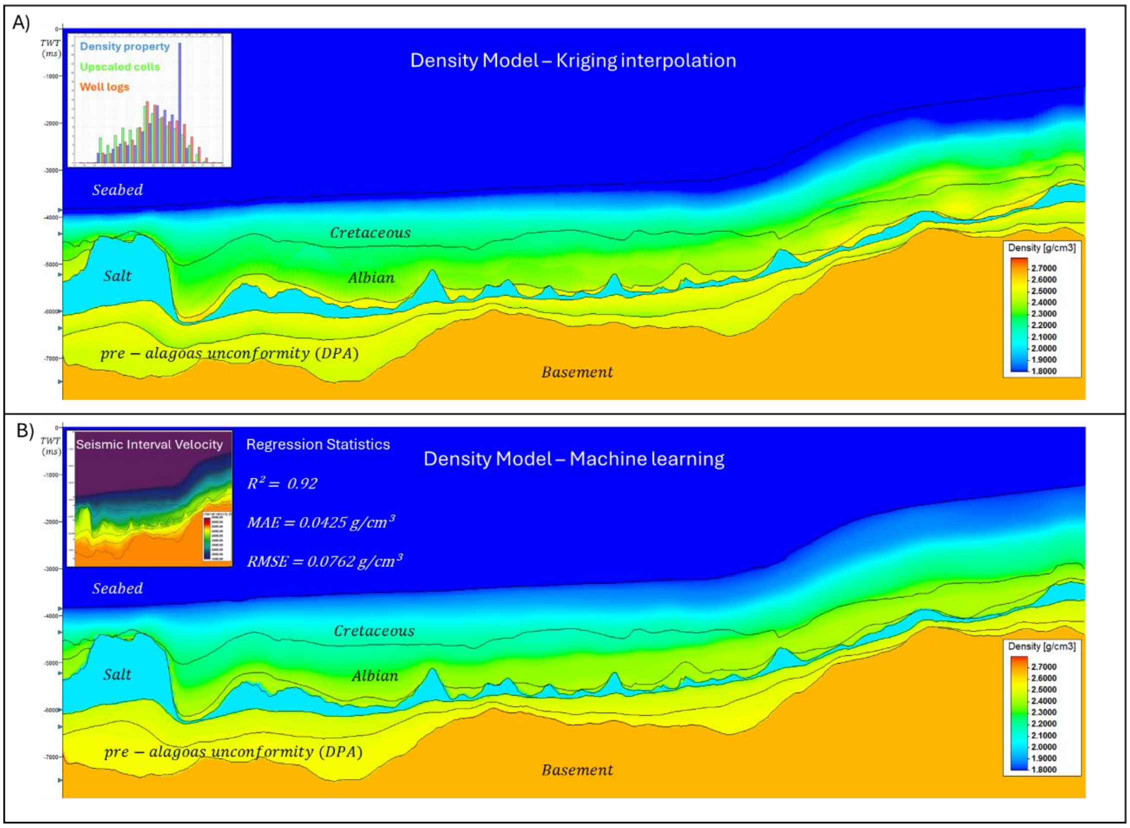

Figure 01: Density model calculated through kriging interpolation (A) and ML method (B). Figure 01.A, distribution histogram of input and modeled data displays the behavior of data. In Figure 01.B seismic interval velocity jointly with regression statistics can be inspected respectively for QC purpose.

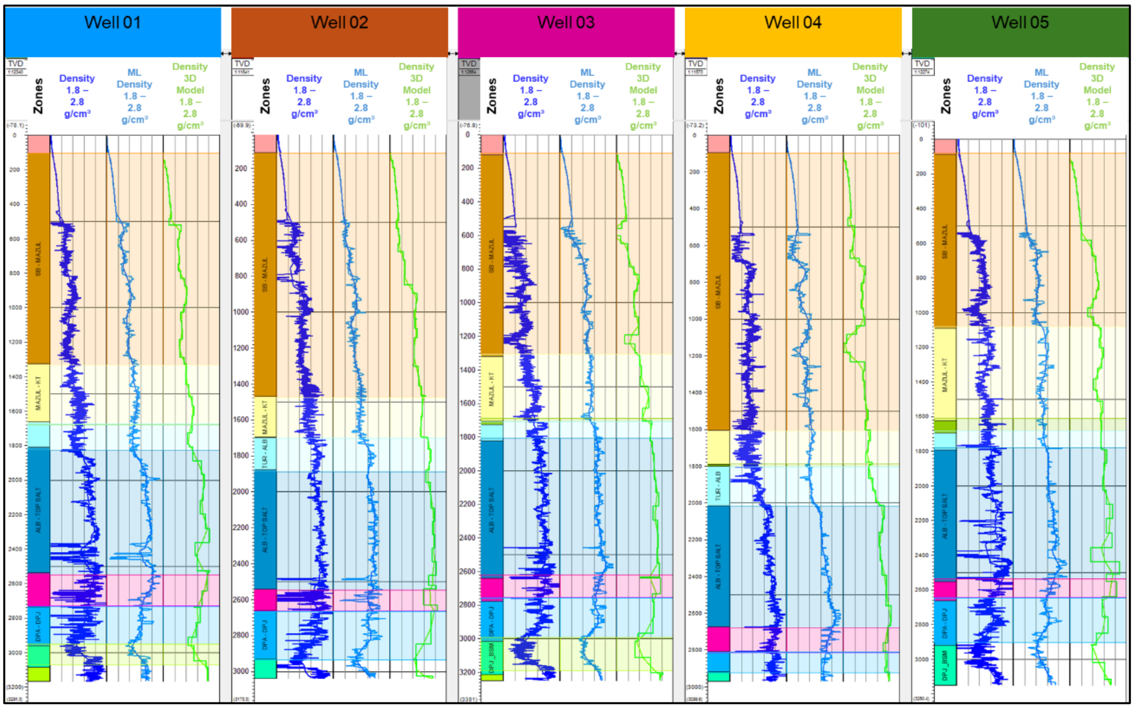

Figure 02: Overview of a set of wells displaying respectively measured density, ML estimated density and kriging density logs jointly with its correspondent filtered log.

References

Azevedo, L., and A. Soares, 2017, Geostatistical Methods for Reservoir Geophysics: Cham, Springer International Publishing, Advances in Oil and Gas Exploration & Production, 159 p., doi:10.1007/978-3-319-53201-1

Boutayeb, A., Lahsen-Cherif, I., and Khadimi, A. E., 2025, When Machine Learning Meets Geospatial Data: A Comprehensive GeoAI Review, in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 18, pp. 13135-13191, 2025, doi: 10.1109/JSTARS.2025.3568715.

Brocher, T. 2005. Empirical Relations between Elastic Wavespeeds and Density in the Earth's Crust. Bulletin of the Seismological Society of America. 95. 2081 - 2092.

Bulhões, F. C. et al. - Geostatistical 3D Density Modeling: Integrating Seismic Velocity and Well Logs. 14th International Congress of the Brazilian Geophysical Society held in Rio de Janeiro, Brazil, August 3-6, 2015

Pyrcz, M. J., and C. V. Deutsch (eds.), 2014, Geostatistical Reservoir Modeling: New York, Oxford University Press, 448 p

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Neural Information Processing Systems.

Winter, W.R.; Jahnert, R.J. & França, A.B. 2007. Bacia de Campos. Boletim de Geociências da Petrobrás, 15 (2): 511-529.

Yamamoto, T. M., Maul, A. & Lupinacci, W. 2019. Analysis of Velocities, Density and Seismic Facies of Ariri Formation – Santos Basin. In: 16th International Congress of the Brazilian Geophysical Society held in Rio de Janeiro, Brazil, 19-22 August 2019.