Brazilian Journal of Geophysics (2022) 40, 1, 133–143
Brazilian Geophysical Society
ISSN 2764-8044
DOI: 10.22564/brjg.v40i1.2157

SEISMIC WELL TIE USING GEOPHYSICAL LOGS OBTAINED FROM K-NEAREST NEIGHBOR REGRESSION ALGORITHM

Matheus R. S. Barbosa, Vinicius C. Santana, and Alexsandro G. Cerqueira

Universidade Federal da Bahia - UFBA, Geosciences Institute, Geophysics Department, Salvador, BA, Brazil

*Corresponding author email: matheus.radames@ufba.br

ABSTRACT. This paper aims to verify if the seismic slowness log estimated through the supervised machine learning K-nearest neighbor (KNN) algorithm can be a feasible alternative to replace the sonic well log as input for the seismic well tie in a dataset from the Recôncavo Basin. The training and optimization of the regression were performed in a dataset composed of 17 well logs with petrophysical information of gamma-ray, deep and shallow resistivities, and the geological formation, e.g, Pojuca, Marfim, Maracangalha, Candeias, São Sebastião, Água Grande, and Sergi Formations. The metric to evaluate the regressions was the mean absolute error of the measured property and the prediction. The Holdout cross-validation technique was applied to avoid overfitting, and a well log was separated as a blind test to verify the prediction in an unknown dataset. Furthermore, synthetic seismic traces were generated from the slowness log and the prediction using the KNN. The comparison between them shows outstanding results in the visual analysis of the peaks and amplitudes of the main seismic events. In addition, the comparison between the seismic traces close to the synthetic seismic traces reveals a better correlation to the calculated traces using the slowness predicted by the KNN algorithm.

Keywords: K-nearest neighbor; seismic well tie; Recôncavo Basin, sonic log.

INTRODUCTION

The slowness log, also known as compressional acoustic sonic log (t), is an important tool employed to predict petrophysical rock information, such as porosity, pore pressure, and permeability in oil and gas exploration in sedimentary basins with complex geological formations (Cranganu and Breaban, 2013). Furthermore, this information plays a crucial role in the seismic well tie activity, which provides means to identify horizons to correctly pick, estimate wavelets to perform inversions in the seismic data, analyze the amplitude versus offset, or for operations which involve geostatistical modelling in 3D space (White and Simm, 2003; Simm and Bacon, 2014; Bulhões et al., 2015).

Supervised machine learning algorithms can be seen as a feasible solution when the logs are missing. Several works applied regression algorithms to predict petrophysical properties related to well log activities and core samples, such as resistivity, compressional and shear slowness, density, photoelectric, and neutron logs using linear regression, models based on trees, support vector machines, and artificial neural networks (Rolon et al., 2005, 2009; Cranganu and Breaban, 2013; Akinnikawe et al., 2018; Suleymanov et al., 2021). In this sense, the supervised machine learning K-nearest neighbor (KNN) technique is a simple but efficient algorithm that can be used to make classifications and regression tasks. This algorithm predicts a value by analyzing the weighted or uniform distance between the K samples closest to the input in an n-dimensional space in which n is equal to the variables available. According to Faceli et al. (2019), algorithms based on the distance of the closest neighbor are one of the most straightforward machine learning techniques.

In Recôncavo Basin, it is common the absence of the slowness property in the well logs to oil and gas exploration, mainly in the shallow portion of the basin, and old well logs, which turns challenging to develop the well tie seismic task with reliability. However, few works tried to estimate missing sonic logs (da Silva et al., 2014; Barbosa et al., 2021), and the Gardner relation (Gardner et al., 1974) is usually applied to estimate the compressional velocity log and generate a synthetic seismic trace. In this case, the KNN algorithm can be an alternative to estimate the compressional slowness log in a dataset where the sonic is unavailable, and its result can be employed in the synthetic seismic trace generation.

This work aims to verify if the regressions obtained from the K-nearest neighbor algorithm of the slowness can be an appropriate way to replace the sonic log in the well seismic tie in Recôncavo basin using the gamma-ray, deep and shallow resistivity logs as quantitative information, and the geological formation as qualitative information. This paper is organized into four main sections: the first one is called Dataset and the Study Area, which presents some characteristics of the geology of the dataset; the second, named Methodology, contains some aspects related to pre-processing, validation, regression algorithm, optimization, and the well tie performed; the third called Results and Discussions that presents the analysis performed in the blind well log; and finally the section Conclusions, which draws some remarks about this research.

DATASET AND THE STUDY AREA

The well logs used to perform the experiments are located in the Recôncavo basin, in the Northeastern Brazilian region, and have approximately 11,500 km². According to Milani and Davison (1988), this intra-continental rift system was developed during the Cretaceous as a consequence of the breaking of Pangea.

The Rift sequence was filled with fluvial-deltaic and lacustrine sediments represented by the Candeias, Marfim, Maracangalha, and Pojuca Formations, interbedded with a fan-delta system derived from the faulted border from the Salvador Formation. The rift's final stage is marked by the spreading of the fluvial system from the São Sebastião Formation (Silva et al., 2007).

The study area (Figure 1) is in the deepest part of the basin, with the prevalence of lacustrine rocks as pelagic shales, debrite and turbidite sandstones, with occasional deltaic and fluvial rocks in the uppermost well parts.

The regression to obtain the slowness well log (also known as sonic) was performed using information from seventeen well logs in training and one borehole as a blind test (see Figure 1). The following measurements were used as input of the regressions: gamma-ray, deep and shallow induction resistivity, and a numerical label corresponding to the geological formation.

METHODOLOGY

The following workflow (Figure 2) summarizes the methodology applied to verify whether the predicted slowness log can replace the measured data in the seismic well tie procedure. The experiments were conducted in Python using the sci-kit learn library (Pedregosa et al., 2011). Each step mentioned inside the boxes is detailed in the subsections below.

Editing and Preprocessing

The first step of analyzing a dataset consists of editing and preprocessing the data, which are, in this case, the petrophysical information from well logs. This stage comprises the standardization of attributes, concatenation of data from different wells, unreliable data verification, and categorical features creation, such as the "geological formation" to which a given sample belongs. Then, a data preprocessing proposes the analysis of the general behavior of the input variables, aiming to identify and eliminate spurious values without geophysical-geological meaning. The dataset ($\mathbf{X}$) is normalized by the standard deviation (σ) and the mean (μ), as shown in equation (4).

$$\mathbf{X}_{\text{norm}} = \frac{\mathbf{X} - \sigma}{\mu}. \tag{4}$$

Holdout Cross-Validation Method

Cross-validation tests are applied to assess whether the built-in regression model predicts well the value of unknown samples, i.e., to gauge the generalizability of an algorithm, so there is no overfitting or underfitting to the input data.

According to Refaeilzadeh et al. (2009), a common approach is to randomly split the available data set into two non-overlapped parts: one for training and the other for testing. The proportion of data partitioning is arbitrary, and, in general, the largest portion of the samples is assigned to compose the training subset. In this study, 75% of the samples were used to train the model.

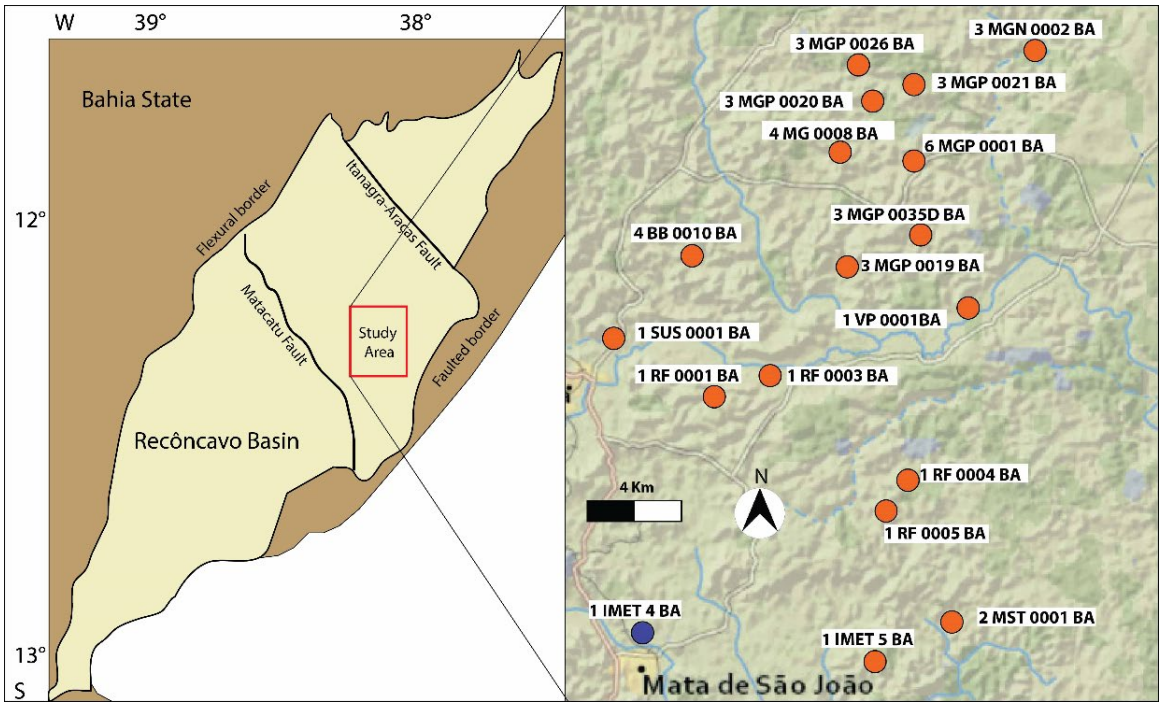

Figure 1: A total of 18 well logs were used to carry out the experiments, in which 17 of them (orange points) were used to train the regression model. The last one (blue point) was used as a blind test to check whether the model was reliable and to perform the seismic well tie.

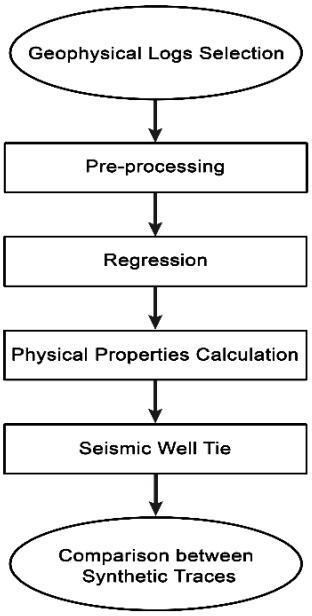

Figure 2: Workflow used to perform the slowness well log regression, generate the synthetic traces, tie the traces to the seismic data and establish a comparison between the results.

The test data (25%) is held out and not used during training. This procedure was performed by the Holdout cross-validation method, which ensures a more accurate estimate for the generalization performance of the algorithm.

K-Nearest Neighbor Algorithm

It is usual to define that the smaller is the distance between two or more samples in the feature space, the more they are similar (Ertel, 2018). The use of the idea behind the statement over an n-dimensional space to build a regression model indicates that the K-nearest neighbor method can predict the value of a sample. The K-nearest neighbor algorithm (KNN) is a non-parametric supervised machine learning technique that uses observed data present in the training subset X to estimate the magnitude of unknown samples based on their similarity or distance d. In other words, to the unknown point is assigned a value based on how much it resembles the points in the training subset.

As said by Ortiz-Bejar et al. (2018), when it is applied to establish a nonlinear regression model, for a given sample x_q, the value of the target attribute $\hat{y}(\mathbf{x_q})$ is calculated as the mean between its most similar (or nearest) k samples $\mathbf{x_i} \in \mathbf{X}$ for each of the input variables weighted by the inverse of the distance w (equation 5).

$$\hat{y}(\mathbf{x_q}) = \frac{1}{k}\frac{\sum_{i=1}^{k} w_i y_i}{\sum_{i=1}^{k} w_i} \tag{5}$$

where

$$w_i = \frac{1}{d(\mathbf{x_i}, \mathbf{x_q})}. \tag{6}$$

A variation of equation (5) can be taken when all the weights are equal to 1. In this case, equation (6) is reduced to the simple mean of the k closest samples (equation 7):

$$\hat{y} = \frac{1}{k}\sum_{i=1}^{k} y_i. \tag{7}$$

Hence, to build the regression model, it is necessary to calculate how far are the training samples regarding the non-estimated point in an n-dimensional space, where n is the number of input variables. This procedure is computationally costly for high dimensionality. It can be performed through distinct distance metrics, e.g., Euclidean, Manhattan, Minkowski, and Hamming. The divergence among them is given by the parameter p, presented in equation (8). The Manhattan and Euclidean distances are defined using p equals to 1 and 2, respectively. These two were used in this work to establish a comparison.

$$d(\mathbf{x_i}, \mathbf{x_q}) = \left[\sum_{j=1}^{D} (\mathbf{x_{i,j}} - \mathbf{x_{q,j}})^p\right]^{1/p}, \tag{8}$$

where D is the number of measurements (logs).

The results obtained using this algorithm are very sensitive to the number of neighbors used to train the model. If this hyperparameter is too small, the final model will overfit, since only a few points would be considered to estimate a new sample. On the other hand, if k is too large, the model will poorly perform in both training and test data. Therefore, optimizing the number of neighbors is an indispensable step to minimize errors in training and validation data.

The Mean Absolute Error (MAE) is used to evaluate regressions since such metric suits well for problems with a large amount of data, besides having a high interpretability. Equation (9) shows how MAE is calculated:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|, \tag{9}$$

in which n is the number of samples and $\hat{y}$ is the calculated data.

Hyperparameter k Optimization

The selection of hyperparameters is part of the model optimization process that aims to minimize the mean absolute error in training and test datasets. In the present work, a variation of the elbow method was used to select the number of neighbors in the K-nearest neighbor technique using uniform weights (equation 9). These curves are constructed iteratively by calculating the mean absolute errors of each subset for each value of K, in a range from one to 200. These graphs highlight the point from which there is a stabilization of the difference between those curves to verify the MAE of this point.

To support the decision-making of the K value, the logarithm of the difference between the mean absolute error of training and test was used in the case where the value of the function was obtained from a mean between the K-nearest neighbor. The resulting graph from this operation (Figure 3) can point out the plateau from which the increase in the number of neighbors does not offer significant interference for the minimization of the metric employed.

It is important to note that this result comes from the adjustment of the model that considers the L2 norm as the distance metric ($p = 2$). Taking into consideration the Manhattan distance ($p = 1$), it was possible to observe that there were no significant changes in errors for the training and test data. Therefore, the later stages were performed using the Euclidean distance to fit the model.

In Figure 4, through the analysis of the mean absolute error on test data, considering the weighted functions mentioned above, it can be seen that there is a brief improvement, for the optimal number of neighbors, in the MAE when the predicted sample's value is weighted by the distance between points (equation 5). It changes from 6.94 $\mu s/ft$ to 6.79 $\mu s/ft$. It is also important to notice that the optimal hyperparameter k is slightly different in both situations: it goes from 7, when the weight function is uniform, to 10.

Table 1 allows us to analyze the model's performance considering the weighted distance for each geological formation present in the dataset. The fact that the predictions made by formations have distinct errors is remarkable. The most reliable estimations are made for the samples of the Sergi and Candeias Formations, while the worst are made in the Pojuca and Marfim Formations.

Finally, to verify the performance of the KNN estimations, a linear regression was performed in the same dataset, and the MAE of the train and test of the linear regression were approximately 25.57 $\mu s/ft$ and 25.24 $\mu s/ft$, respectively.

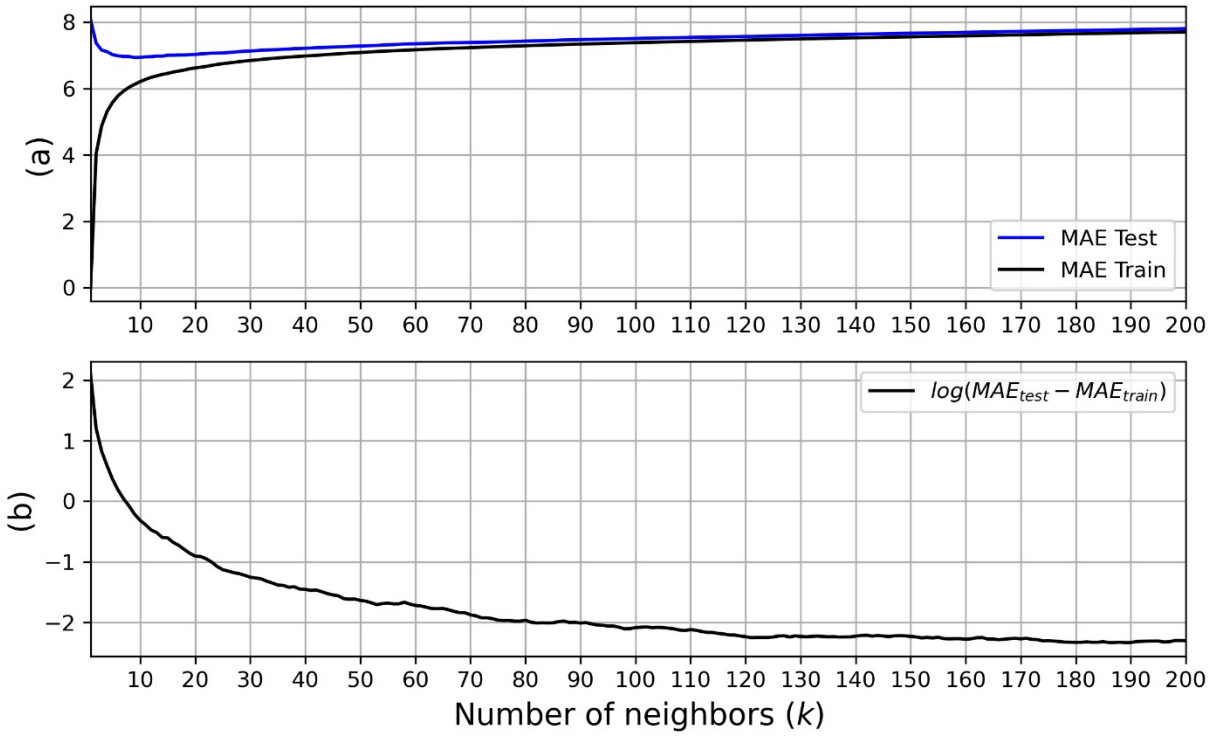

Figure 3: (a) The mean absolute error of the K-nearest neighbor regression with uniform weights applied to train and test datasets using the following attributes: gamma-ray (GR), resistivity (R_{ild}), the difference between resistivity and short normal (SN), and geological formation; (b) the logarithm of the difference between the test and train mean absolute errors.

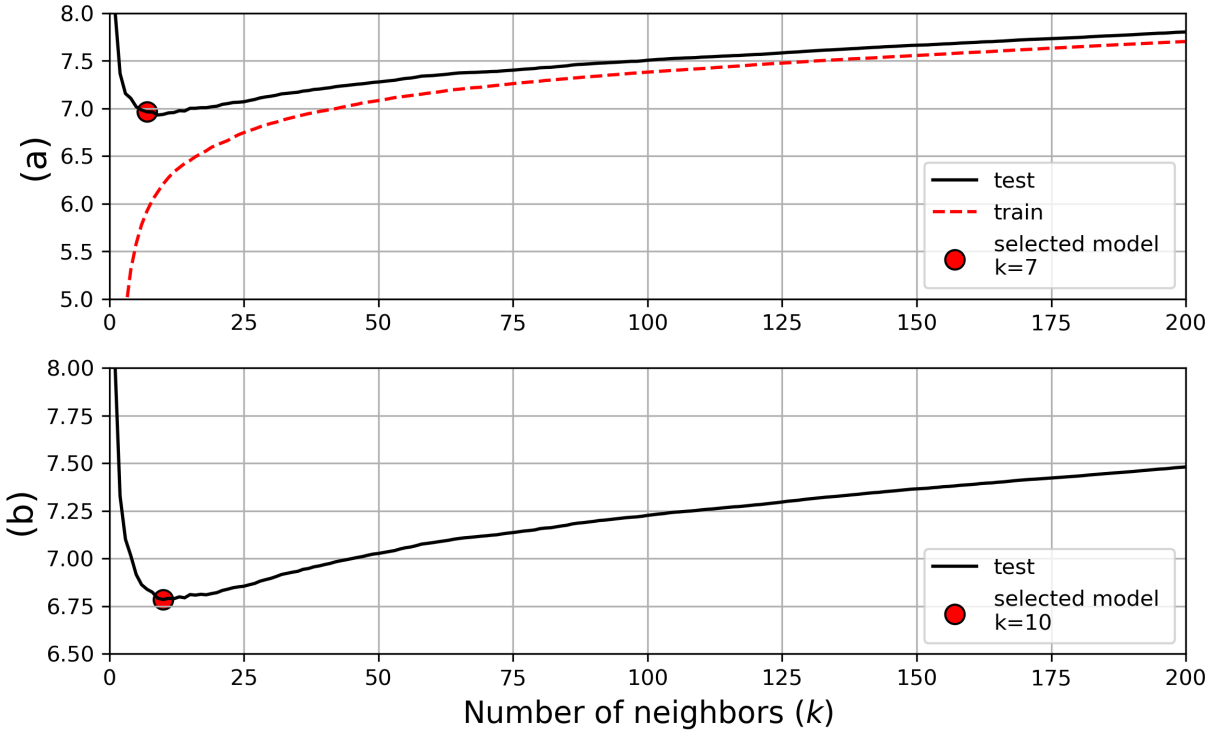

Figure 4: The mean absolute error of the K-nearest neighbor regression applied to train and test datasets when the parameter weighted function is (a) uniform and (b) the distance weighted. Because of the non-uniform weighting, the error for the training data is always zero; hence it does not appear in the graph.

Table 1: Train and test mean absolute error by geological formation.

Geological Formation	Pojuca	Marfim	Maracangalha	Candeias	São Sebastião	Água Grande	Sergi	Total
MAE Test ($\mu s/ft$)	10.09	7.09	6.95	5.21	6.53	5.27	2.08	12.81

Seismic Well Tie

The seismic well tie is an essential stage in seismic interpretation. It relates rock physical properties measured in a borehole with time travel data recorded during seismic reflection activity, providing the correct identification of seismic horizons (White and Simm, 2003). According to Simm and Bacon (2014), seeing the well tie is the interpreters' chance to link the geology and the seismic data.

The classic seismic well tie procedure embodies the synthetic seismic trace alignment with the real seismic trace near the borehole. The generation of the synthetic trace is based on the convolutional model, defined by the convolution of the reflectivity series $\mathbf{r(t)}$ with the seismic wavelet $\mathbf{w(t)}$ added to a noise $\mathbf{n(t)}$, according to equation (10):

$$\mathbf{x(t)} = \mathbf{r(t)} * \mathbf{w(t)} + \mathbf{n(t)}. \tag{10}$$

To achieve this result, it is necessary to perform a series of steps, where the most important are the geophysical log upscale and, consequently, the reflectivity coefficients obtainment; the wavelet extraction; and the stretch and squeeze of the synthetic trace. Firstly, in our experiment using the blind well log, the upscale was done using the median filter with a window equals to 151 samples, as displayed in Figure 5.

After filtering the compressional velocity and density, the acoustic impedance was obtained from equation (2), allowing us to estimate the reflection coefficient (equation 3), a crucial parameter for obtaining the synthetic trace.

Then, to determine the wavelet, the frequency spectrum of the nearest real seismic trace was calculated and four frequencies were selected (Figure 6) to compose the frequency bandwidth of a zero-phase Ormsby wavelet (Figure 7), given by equation (11):

$$w(t) = \frac{\pi f_1^2}{f_2 - f_1} sinc^2(\pi f_1 t) - \frac{\pi f_2^2}{f_2 - f_1} sinc^2(\pi f_2 t) - \frac{\pi f_3^2}{f_4 - f_3} sinc^2(\pi f_3 t) + \frac{\pi f_4^2}{f_4 - f_3} sinc^2(\pi f_4 t) \tag{11}$$

where f_1, f_2, f_3, and f_4 are the main frequencies of the Ormsby wavelet.

In possession of this information, two synthetic traces were generated for comparison: one using the compressional velocity obtained from the sonic well log and the other obtained from the regression using the KNN algorithm.

RESULTS AND DISCUSSION

For a better understanding of this work's results, it is possible to section them in two consecutive phases. The first one covers the pre-processing of geophysical log data up to the regression of the seismic slowness data, while the next step comprises the generation of synthetic traces using the measured sonic log and the log estimated by the KNN method, followed by the comparison between both synthetic traces. The following subsection will present the regression results in the blind dataset, the well tie performed using the slowness curve of the blind data, and the sonic log obtained from KNN. The well tie is supported with the use of the checkshot data

Regression

After the parameter optimization process, the final model was built using the L2 norm as the distance metric; the prediction of a sample's value is weighted by its distance to the K nearest points, and the number of neighbors equals 10. Figure 8 presents the result of the regression applied to the blind well log (1-IMET-4BA), which has the checkshot information available. The mean absolute error between measured and predicted seismic slowness is 12.81 s/ft. The estimated curve suits better to the one measured after 850 meters of depth.

The algorithm had difficulty predicting the slowness in the Marfim Fm., and upper Maracangalha, mainly in regions where the density log presented values lower than 2 g/cm^3, probably due to collapses inside the borehole – information that could not be confirmed because of the absence of the caliper log. Table 2 highlights what can be seen in Figure 8 regarding the prediction error. The algorithm nicely predicted the lower part of the Maracangalha Formation and the Candeias. The predictions in

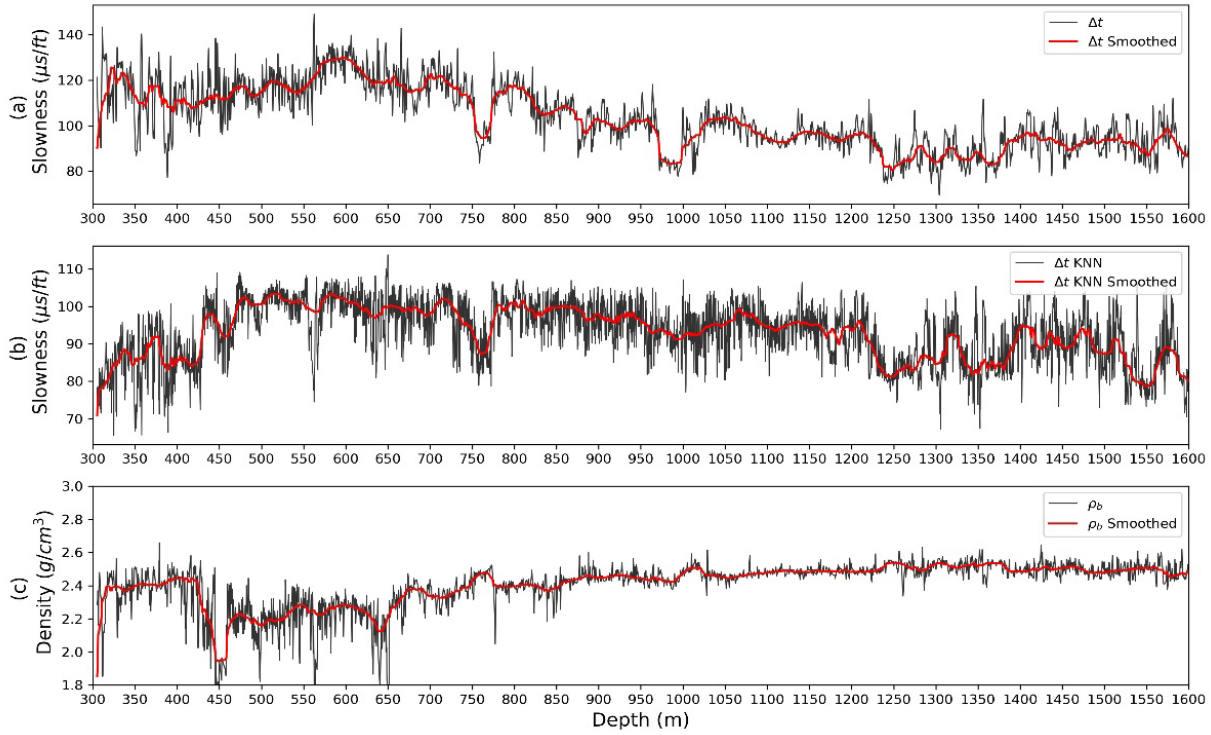

Figure 5: In red, the application of the median filter over the following rock physical properties (in black): (a) measured and (b) predicted seismic slowness and (c) density of the blind well log. The filter window size was chosen through a visual analysis of the curve smoothing.

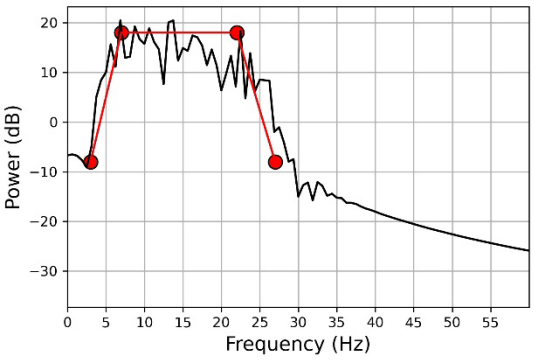

Figure 6: Amplitude spectrum of the nearest real seismic trace. A set of four frequencies, $f_1 = 3$, $f_2 = 7$, $f_3 = 22$, and $f_4 = 27$ Hz, was chosen to construct the zero-phase Ormsby wavelet.

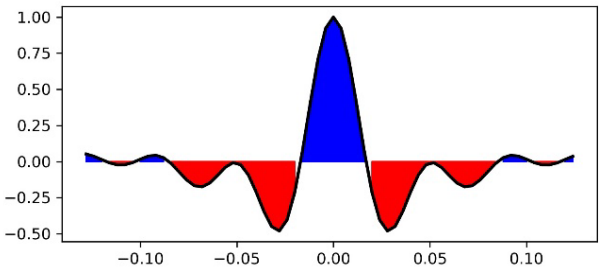

Figure 7: Zero-phase Ormsby wavelet generated using the frequencies $f_1 = 3$, $f_2 = 7$, $f_3 = 22$, and $f_4 = 27$ Hz employed in the seismic well tie of the blind well log.

Table 2**:** Mean Absolute Error by Geological Formation in the blind dataset.

Geological Formation	Pojuca	Marfim	Maracangalha	Candeias	Total
MAE $(\mu s/ft)$	-	21.57	13.51	5.77	12.81

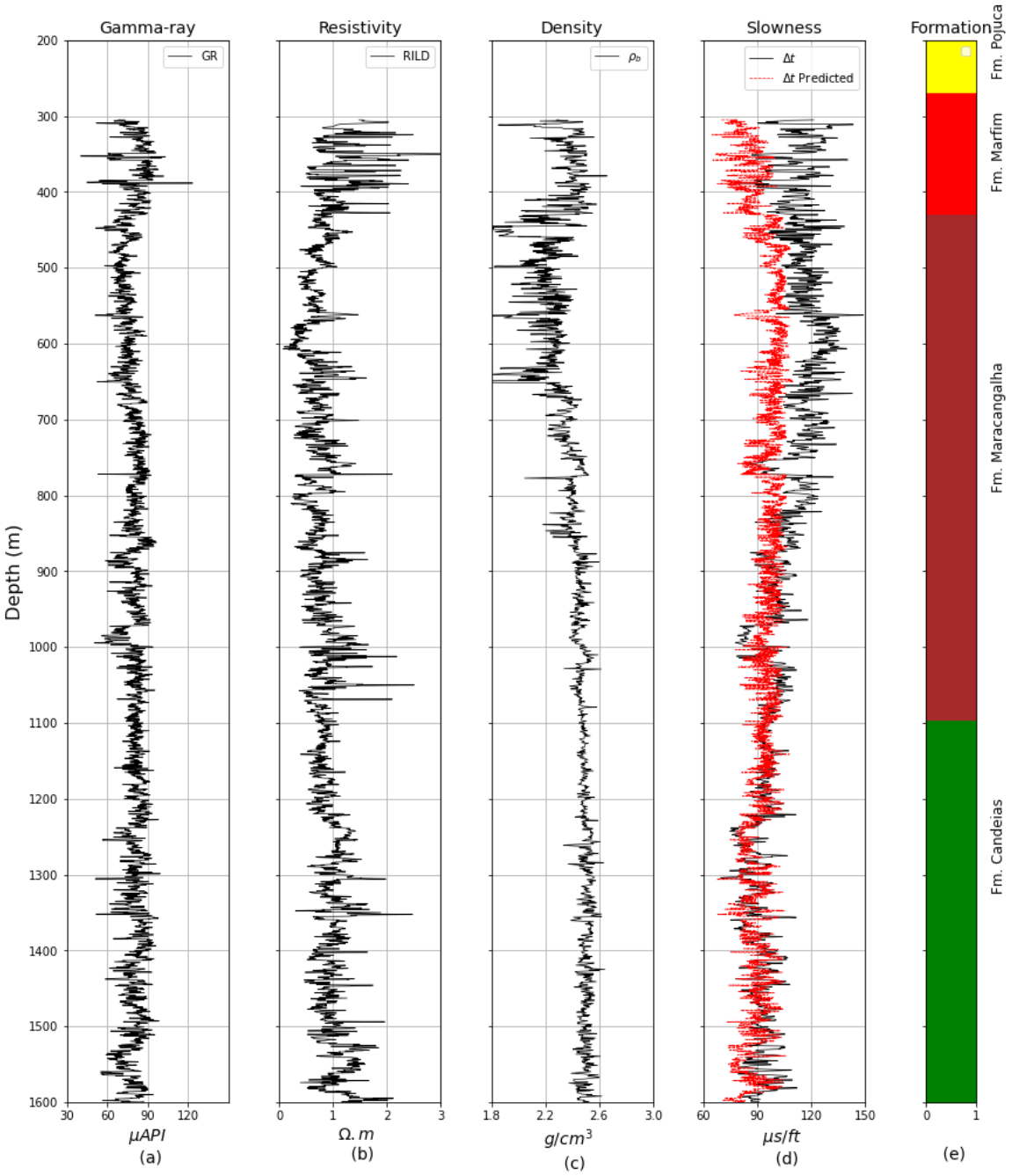

Figure 8: Chart of geophysical logs referring to: (a) gamma-ray, (b) resistivity, (c) density, and (d) measured (continuous black curve) and estimated (dashed red line) seismic slowness. The last track (e) presents the geological formation to which every sample in the blind well log belongs.

Marfim Fm. are inaccurate. Although the Pojuca Formation appears in the geological formation column, it does not have its slowness samples predicted, due to the fact that the resistivity information is missing.

Well to seismic tie

Once the well log data were upscaled, the reflection coefficients were obtained and the wavelet extracted. The whole seismic well tie process took place on the blind test. The synthetic traces were generated using both the measured and estimated seismic slowness to compare each one with the seismic traces close to them.

Figure 9 illustrates them as well as some other rock physical properties. It is remarkable that the labels inside each track indicate from which seismic slowness log that respective property was obtained. As a way to compare the synthetic traces, the Pearson coefficient was calculated between the synthetic traces in the time range from 0.24 to 1 second. This cutout was selected to avoid wavelet edge effects. The Pearson correlation is approximately 0.83

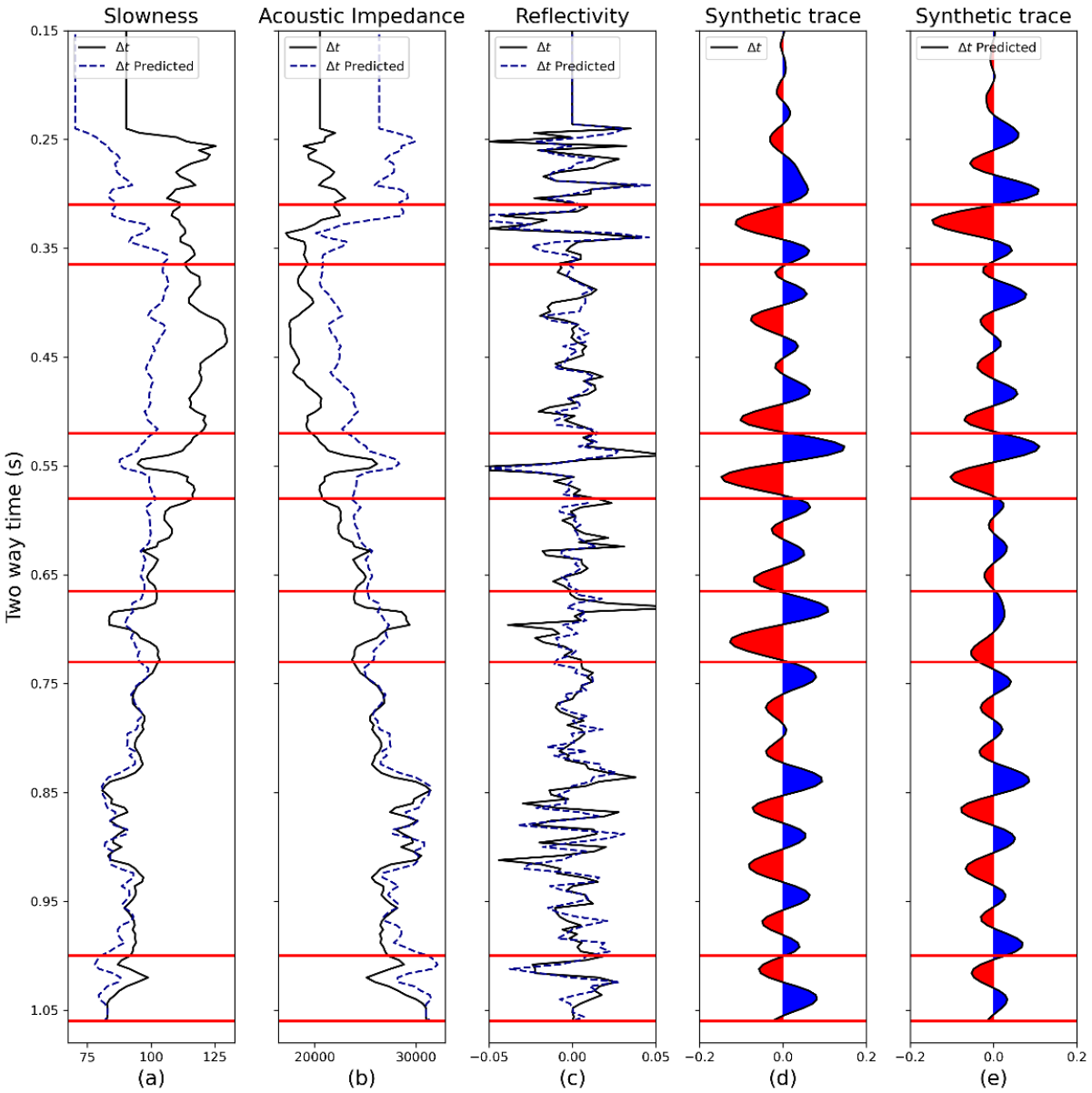

Figure 9: Synthetic traces generated from the sonic log data and the predicted slowness log data. The four time-intervals highlighted by the red lines indicate the main contrasts of acoustic impedance that characterize the leading seismic events.

between the synthetic traces. Furthermore, four intervals were selected for a visual analysis, which allows us to verify that both traces have temporally correlated peaks and troughs, with similar seismic amplitudes, except for the third time interval in Figure 9 (between 0.62 and 0.73 seconds).

To demonstrate that the nonlinear regression using the K-nearest neighbor algorithm is an outstanding alternative to replace the sonic well log when it is not included in the dataset, or its quality is compromised due to borehole problems, we compared the synthetic trace obtained from the slowness and the regression with the closest seismic traces available in the seismic dataset, as can be displayed in Figure 10. The highest Pearson correlation between the seismic trace and the synthetic traces in the time interval of 0.24 s to 1.032 s was approximately 0.36 for the synthetic trace generated using the slowness of the log and 0.42 for the synthetic trace obtained from the regression of the sonic log. Furthermore, it is necessary to enhance that stretch. The squeeze was not performed in any synthetic traces to avoid a non-geological information insertion in the comparisons.

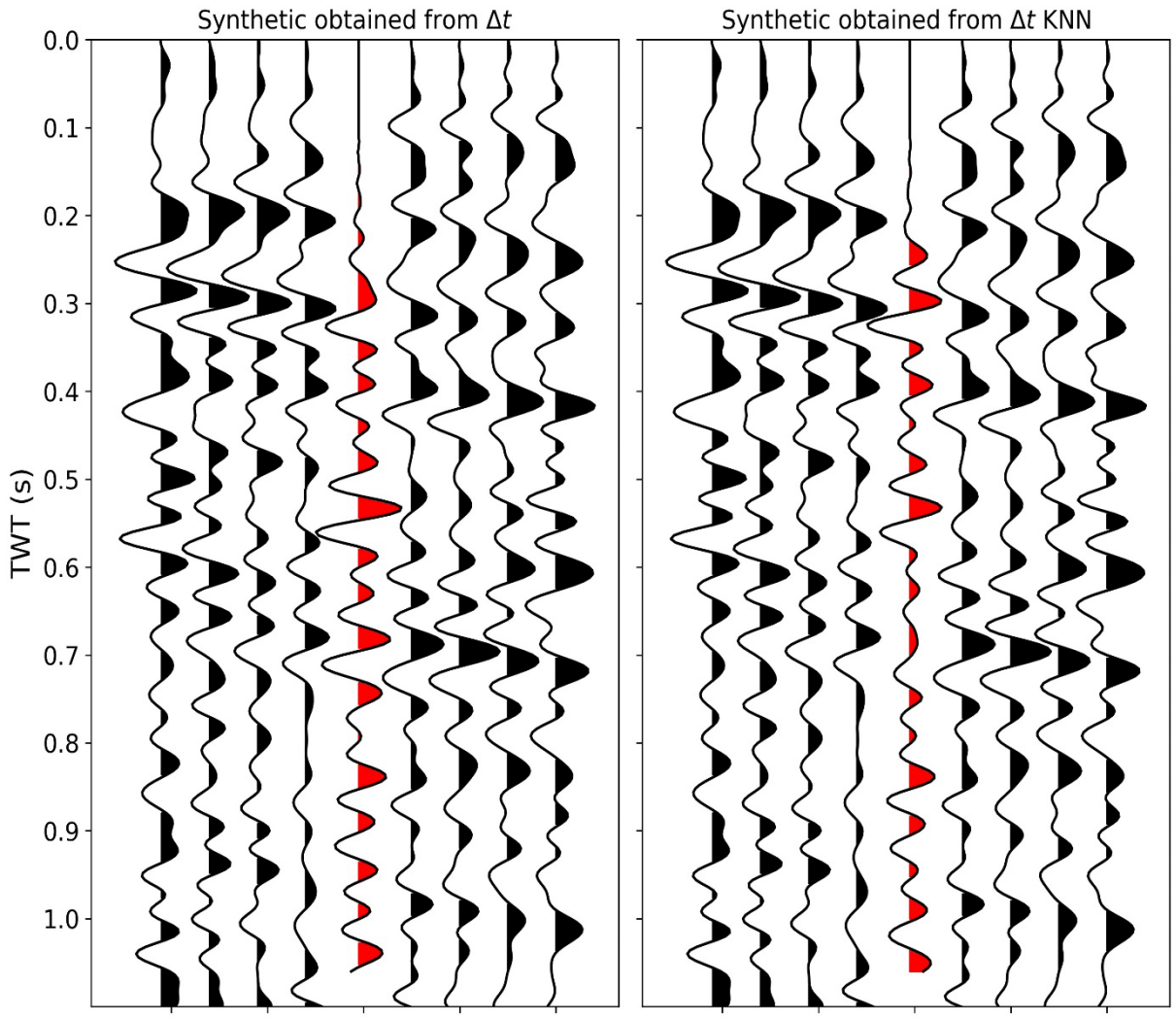

Figure 10: Comparison between (a) the synthetic trace obtained from the measured slowness log (Δt) and (b) the synthetic seismic trace derived from the estimated slowness log ($KNN\ \Delta t$) with the closest traces within the seismic survey.

CONCLUSIONS

The slowness regression estimated from the K-nearest neighbor algorithm presented an outstanding result in comparison to the linear regression performed into the training dataset. To demonstrate the technique effectiveness, a blind well log was used and the Mean Absolute Error was approximately **12.81** $\boldsymbol{\mu s/ft}$. It allowed the authors to obtain more reliable logs to be used in the synthetic seismic trace generation to perform the well tie seismic task. The synthetic trace obtained from the **Δt** log and the regression were very similar, and the comparison between them and the closest seismic trace in the seismic survey using the Pearson correlation pointed out a better result to the synthetic trace calculated using slowness estimated from KNN.

ACKNOWLEDGMENTS

The authors would like to thank Instituto Nacional de Ciência e Tecnologia de Geofísica do Petróleo (INCT-GP) for the scholarship of Author 1, and Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB) for the scholarship of Author 2. Authors 1, 2 and 3 acknowledge Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Author 3 also thanks Projeto FINEP – Apoio à Rede de P&D em Gás Não Convencional do Brasil (GASBRAS) for supporting this research and for his scholarship. This study was also financed by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. Finally, the authors also thank Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP) for providing the dataset and dGB Earth Science for the OpendTect software license.

REFERENCES

Akinnikawe, O., S. Lyne, J. Roberts, and Devon Energy Corp E&P, 2018, Synthetic Well Log Generation Using Machine Learning Techniques: SPE Unconventional Resources Technology Conference, doi: 10.15530/urtec-2018-2877021.

Barbosa, M.R.S., A.G. Cerqueira, and V.C. Santana, 2021, Non-linear regression model as a replacement for seismic slowness log data in the construction of synthetic traces: 17th International Congress of the Brazilian Geophysical Society, Rio de Janeiro, RJ, Brazil, Sociedade Brasileira de Geofísica, SBGf.

Bulhões, F.C, C.D.M.R. Formento, J.C.S. de O. Lyrio, G.A.S. De Amorim, G.D. Ferreira, E.S. Pereira, and R.F. Castro, 2015, Geostatistical 3D Density Modeling: Integrating Seismic Velocity and Well logs: 14th International Congress of the Brazilian Geophysical Society, Rio de Janeiro, RJ, Brazil, Sociedade Brasileira de Geofísica, SBGf, doi: 10.1200/sbgf2015-054.

Cranganu, C., and M. Breaban, 2013, Using support vector regression to estimate sonic log distributions: A case study from the Anadarko Basin, Oklahoma: Journal of Petroleum Science and Engineering, **103**, 1–13, doi: 10.1016/j.petrol.2013.02.011.

Da Silva, A.A.N., B.F. Bahia, T.C.S, Sant'ana, V.C. Santana, and M. Holz, 2014, Modelagem de perfis geofísicos sintéticos em poços da bacia do Recôncavo: VI Simpósio Brasileiro de Geofísica, Porto Alegre, RS, Brazil, Sociedade Brasileira de Geofísica, SBGf, doi: 10.22564/6simbgf2014.035.

Ertel, W., 2018, Introduction to artificial intelligence: 2nd ed., Springer, Switzerland, 356 pp, ISBN: 978-3319584867.

Faceli, K., A.C. Lorena, J. Gama, and A.C.P.L.F. de Carvalho, 2019, Inteligência artificial: uma abordagem de aprendizado de máquina: Rio de Janeiro, Brazil, LTC, 378 pp, ISBN: 978-8512618805.

Gardner, G.H.F., L.W. Gardner, and A.R. Gregory, 1974, Formation velocity and density – the diagnostic basics for stratigraphic traps: Geophysics, **39**, 770–780, doi: 10.1190/1.1440465.

Milani, E.J., and I. Davison, 1988, Basement control and transfer tectonics in the Recôncavo-Tucano-Jatobá rift, Northeast Brazil: Tectonophysics, **154**, 41–70, doi: 10.1016/0040-1951(88)90227-2.

Ortiz-Bejar, J., M. Graff, E.S. Tellez, J. Ortiz-Bejar, and J.C. Jacobo, 2018, K-Nearest Neighbor Regressors Optimized by using Random Search: 2018 IEEE International Autumn Meeting on Power, Electronic and Computing (ROPEC), doi: 10.109/ROPEC.2018.8661399.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, 2011, Scikit-learn: Machine learning in Python: Journal of Machine Learning Research, **12**, 2825–2830.

Refaeilzadeh, P., L. Tang, and H. Liu, 2009, Cross validation, *in* Liu, L., and M.T. Özsu, Eds., Encyclopedia of database systems: Springer, Boston, MA, p. 532–538, doi: 10.1007/978-0-387-39940-9_565.

Rolon, L., S.D. Mohaghegh, S. Ameri, R. Gaskari, and B. McDaniel, 2005, Developing Synthetic Well Logs for the Upper Devonian Units in Pennsylvania: SPE Eastern Regional Meeting, OnePetro, Morgantown, West Virginia, SPE-98013-MS, doi: 10.2118/98013-MS.

Rolon, L., S.D. Mohaghegh, S. Ameri, R. Gaskari, and B. McDaniel, 2009, Using artificial neural networks to generate synthetic well logs: Journal of Natural Gas Science and Engineering, **1**, 118–133, doi: 10.1016/j.jngse.2009.08.003.

Silva, O.B., J.M. Caixeta, P.S. Milhomem and M.D. Kosin, 2007, Roteiros geológicos – guia de campo da Bacia do Recôncavo, NE do Brasil: Boletim de Geociências da Petrobras, **15**, 423–431.

Simm, R., and M. Bacon, 2014, Seismic Amplitude: An interpreter's Handbook: Cambridge University Press, 279 pp, ISBN: 978-1107011502, doi: 10.1017/CBO9780511984501.

Suleymanov, V., H. Gamal, G. Glatz, S. Elkatatny, and A. Abdulraheem, 2021, Real-Time Prediction for Sonic Slowness Logs from Surface Drilling Data Using Machine Learning Techniques: SPE Annual Caspian Technical Conference. OnePetro, doi: 10.2118/207000-MS.

White R.E., and R. Simm, 2003, Tutorial: good practice in well ties: First Break, **21**, doi: 10.3997/1365-2397.21.10.25640.

Barbosa, M.R.S., Santana, V.C. and Cerqueira, A.G.: study conception and design, algorithm implementation, result analysis and interpretation, draft writing, manuscript review, suggestions for improving the intellectual content.

Received on December 30, 2021 / Accepted on May 11, 2022