Density log construction utilizing machine learning

Caique P. de Carvalho* (Universidade Federal do Pará, Brasil), José J. S. de Figueiredo (Universidade Federal do Pará, Brasil & INCT-GP, Brasil) e Carolina B. da Silva (Universidade Federal do Pará, Brasil)

Abstract

Determination of the reflection coefficients is a key element to a well-to-seismic tie, and the density log has a major petrophysical importance as it is used to calculate the acoustic impedance and the reflectivity log. Many authors have developed empirical relations to determine the bulk density log from other logs information such as compressional velocity and shale volume fraction. Machine learning techniques have been applied in image and voice recognition, medical diagnosis, statistics and many other problems involving regression and classification including some in geophysics. The idea in this work is to compare the accuracy of an Artificial Neural Network model, which calculates the density log by having other logs as input, with the existing empirical models, and determine which one presents the best adjustment. An ANN model was created and the comparison with the empirical models was made by statistical analysis such as calculating the mean squared error, the relative error and correlation factor. The ANN model presented smaller errors and higher precision on the adjustment compared to the empirical models.

Introduction

The density log plays a major role both in well logging and in well to seismic tie, its importance is due to the petrophysical information that can be determined by the rocks' densities and applied to the oil and gas industry. A formation evaluation can be constructed utilizing the density log associated with other logs information. By calculating the acoustic impedance of the medium which result from the product of the density and the P-wave velocity, it is possible to calculate the seismic reflection coefficients which are fundamental in the well to seismic tie.

From laboratory studies containing sedimentary samples of different basins, geologic ages, and depths, Gardner et al. (1974) developed an empirical equation that calculates the density from the P-wave velocity information, also allowing the calculation of acoustic impedance and seismic reflection coefficients from P-wave velocity information alone.

Besides Gardner, other authors like Brocher (2005), Lindseth (1979), Christensen and Mooney (1995) and Birch (1960) also developed empirical equations to calculate the density log. Two important empirical models were proposed by Oloruntobi and Butt (2019) which consider for the calculation of the density log, not only the P-wave velocity information (V_p) but also the volume fraction of shale (V_{sh}). The first model is a linear equation obtained by modifying the equation of Han et al. (1986), the second model was obtained from Gardner's equation in which the term for the volume fraction of shale was added. The equations' parameters were calculated in the article utilizing core samples from USA and Gulf of Mexico wells, but they can also be calculated for other formations utilizing a linear regression method for the first model and a non-linear regression method for the second model. A special focus will be given to the empirical equation of Gardner et al. (1974) and to the second empirical model of Oloruntobi and Butt (2019) because these equations were developed for sedimentary formations and present small errors on the density log calculation.

Artificial Neural Networks (ANN) were first developed by McCulloch and Pitts (1943) as a mathematical model and then evolved with the works of authors like Rosenblatt (1958), Hebb (1949), Rochester et al. (1956), Widrow and Hoff (1960). Nowadays, the applications of Artificial Neural Networks are very broad, for instance, it might be found in medical diagnostics, voice recognition, fraud detection, and many others. Applications in geophysics may be found in the works of Van der Baan and Jutten (2000), Kohli and Arora (2014) and Long et al. (2016), these last two are examples of well logging applications.

A problem that arises in the construction of the density log utilizing empirical equations, either with the original parameters calculated by the authors or by re-calibrating them with linear or nonlinear regressions for specific formations, is that the error between real and calculated density logs might be significant due to geological factors as geochronological differences in the formations and fractured rocks. As an attempt to overcome this issue, an Artificial Neural Network algorithm will be used to create a model that constructs the density log by having other geophysical logs as input data. The results of the ANN model will be compared with the results from the empirical equations to determine which one presents the best adjustment.

Methodology

The objective of this work is to build an Artificial Neural Network model and compare it with the empirical equations. It is possible to observe in Figure 1 all the steps of the work that lead to the best fit selection. The empirical equations utilized in this work are listed in the following section, the point is to test each equation for well logs of different sedimentary basins and analyze their limitations by calculating the relative error (R.E), the mean

squared error (M.S.E), and adjustment factor (R) between the real density log and the density log calculated by the empirical equation and compare with the results from the ANN model.

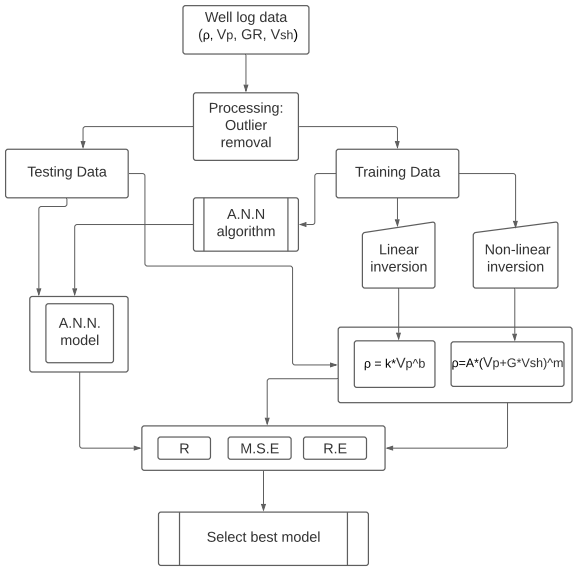

Figure 1: Illustration of the flowchart detailing the steps of the work's methodology.

The function NNSTART (Neural Network Start) from the MATLAB programming language is utilized to obtain the ANN models. The data set utilized to perform the training, validation, and testing steps of the ANN contains well log information from the Campos Basin in Brazil, the Norne field in Norway, Taranaki Basin in New Zealand, and some well logs from Alaska, USA. A second data set containing well logs from the Penobscot field in Nova Scotia, Canada was utilized exclusively to perform additional tests. All the well log data was first submitted to an outlier removal phase, as shown in Figure 1, utilizing a despike function and then submitted to the steps of the empirical equations and machine learning algorithm.

The re-calibration of the empirical equations' parameters by inversion was made utilizing a reference well from Norne field.

Empirical models

The parameters k and B for Gardner's equation on its original article, for a P-wave velocity measured in Km/s, and bulk density in g/cm^3, are equal to 1.74 and 0.25, respectively. These parameters can also be obtained by linear regression for other sedimentary basins utilizing the Least Squares method; nevertheless, the values are generally different. The parameters A, G, and m for Oloruntobi and Butt (2019) model II also known as modified Gardner's model are equal to 1.350, 1.651, 0.390, respectively, they can also be obtained for other formations by non-linear regression utilizing the Levenberg-Marquardt method (Levenberg, 1944; Marquardt, 1963) but the values will also be different from the original article ones.

Gardner et al. (1974) equation is given by:

$$\rho_{\mathrm{b}} = k\left[V_p\right]^B, \tag{1}$$

where ρ_{b} is the bulk density, V_p is P-wave velocity and k and b are constants.

By applying the natural logarithm on both sides of equation 1, a linear equation is obtained. It is now possible to find a solution for the constants k, b utilizing the least squares method. Modifying equation 1 with the natural logarithm and applying on the Least Squares equation:

$$\begin{bmatrix} 1 & ln(V_{p1}) \\ 1 & ln(V_{p2}) \\ \vdots & \vdots \\ \vdots & \vdots \\ 1 & ln(V_{pn}) \end{bmatrix} \begin{bmatrix} ln(k) \\ B \end{bmatrix} = \begin{bmatrix} ln(\rho_{b1}) \\ ln(\rho_{b2}) \\ \vdots \\ \vdots \\ ln(\rho_{bn}) \end{bmatrix} \tag{2}$$

Now it is possible to obtain the values of the vector c, which is $\mathrm{c} = [ln(k), b]$ for equation 1 by replacing each element of equation 2 on the Least Squares equation. Considering that: $ln(k) = x_r$, then it is possible to find the value of k by applying the exponential function on both sides: $k = exp(x_r)$. After these steps, it was possible to obtain the new values for parameters k and B utilizing the reference well from the Norne field.

Oloruntobi and Butt (2019) model II (Modified Gardner):

$$\rho_{\mathrm{b}} = \mathrm{A}\left[V_p + \mathrm{GV_{sh}}\right]^m. \tag{3}$$

The $\mathrm{V_{sh}}$ log was calculated from the Gamma-ray (GR) log utilizing Clavier et al. (1971) equation:

$$\mathrm{V_{sh}} = 1.7\sqrt{3.38 - (I_{GR} + 0.7)^2}. \tag{4}$$

The Levenberg-Marquardt algorithm was utilized both on the regularization of the weights during the training stage of the neural network and on the determination of the modified Gardner's model parameters A, G, and m for different geologic formations, nevertheless, it is important to point out that while the Jacobian matrix of the neural network regularization is composed of the partial derivatives of the function with respect to the weights as shown on equation 9, the Jacobian matrix of the modified Gardner's model is composed of the function's derivatives with respect to the parameters A, G, and m:

$$\mathbf{J} = \begin{bmatrix} \frac{\partial \rho_1}{\partial A} & \frac{\partial \rho_1}{\partial G} & \frac{\partial \rho_1}{\partial m} \\ \vdots & \vdots & \vdots \\ \frac{\partial \rho_n}{\partial A} & \frac{\partial \rho_n}{\partial G} & \frac{\partial \rho_n}{\partial m} \end{bmatrix}_{nx3}, \tag{5}$$

where $\frac{\partial \rho_n}{\partial A}$, $\frac{\partial \rho_n}{\partial G}$ and $\frac{\partial \rho_n}{\partial m}$ are the partial derivatives of the function with respect to the parameters. When the parameters k and B from the traditional Gardner's equation and also the parameters A, G, and m from the modified Gardner's equation were calculated by inversion utilizing the first data set, consisting of well logs from Campos, Norne, Alaska and Taranaki basins; their values were very close to the ones of their respective original articles, therefore, a single well from the Norne field was chosen to be the reference well.

ANN model

Lv et al. (2017) shows that for a Neural Network with two layers, the ANN architecture will be represented by:

$$a^2 = f^2\left(\sum_{i=1}^{S} w_{1,i}^2 f^1\left(\sum_{j=1}^{R} w_{i,j}^1 p_j + b_i^1\right) + b^2\right), \quad (6)$$

which was also the one utilized in this work, $w_{i,j}^1$, p_j, b_i^1 and f^1 represent the weights, input vector, bias and activation function for the hidden layer, R is size of the input vector p_j, while $w_{1,i}^2$, b^2, and f^2 represent the weights, bias and activation function for the second layer (output layer), and a^2 represents the network's output. The input data might also be a matrix, however, to establish the matrix multiplication, the number of columns of the matrix containing the weights needs to be equal to the number of lines of the input matrix, this also guarantees that each weight value is associated to each input value.

It is important to notice in equation 6 that there is an inner equation which represents the outputs of the first layer, each of these elements is multiplied with the weights of the second layer and then summed, showing that the outputs of the first layer are inputs of the second layer.

The next steps after the training of the ANN will be validating and testing. Table 1 shows the total number of samples from the data utilized for the ANN model and how they were split, 80% for training, 10% for validation, and 10% for testing. Additional tests were made with the second data set containing a well log from the Penobscot field in Canada which was used to perform tests on rocks with different lithology, also called: "blind tests".

Table 1: Number of samples for the ANN model and percentage used for each step.

	Samples	%
Total	69669	100%
Training	55735	80%
Validation	6967	10%
Testing	6967	10%

The logs used as input data for the ANN model were the P-wave velocity (V_p) calculated from the sonic log (DT), and the volume fraction of shale (V_{sh}) calculated from the Gamma-ray log (GR).

The Equation 7 describes the ANN model, which has as input, a matrix containing two lines and n rows, each line representing a log (V_p and V_{sh}). The hidden layer contains 25 neurons, each neuron containing two weights. On the second layer there is only one neuron containing 25 weights.

$$a^2 = f^2\left(\sum_{i=1}^{25} w_{1,i}^2 f^1\left(\sum_{j=1}^{2} w_{i,j}^1 p_{jk} + b_i^1\right) + b^2\right), k = 1,2...n. \quad (7)$$

The ideal number of neurons on the hidden layer was reached after a number of tests which confirmed that a number of neurons larger than 25 would generate an overfitting of data.

The Levenberg-Marquardt method was also utilized on the ANN algorithm to optimize the solution for the weights values. The algorithm utilized for neural networks is called Levenberg-Marquardt backpropagation, and modifying it for the weights update gives:

$$w^{l+1} = w^l - \left[\mathbf{J}^T\mathbf{J} + \mu\mathbf{I}\right]^{-1}\mathbf{J}^T\mathbf{e}. \quad (8)$$

The Jacobian matrix will be:

$$J = \begin{bmatrix} \frac{\partial e_1(w)}{\partial w_1} & \frac{\partial e_1(w)}{\partial w_2} & \cdots & \frac{\partial e_1(w)}{\partial w_n} \\ \frac{\partial e_2(w)}{\partial w_1} & \frac{\partial e_2(w)}{\partial w_2} & \cdots & \frac{\partial e_2(w)}{\partial w_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial e_N(w)}{\partial w_1} & \frac{\partial e_N(w)}{\partial w_2} & \cdots & \frac{\partial e_N(w)}{\partial w_n} \end{bmatrix} \quad (9)$$

For the comparison of the empirical equations of traditional and modified Gardner models (Gardner et al., 1974; Oloruntobi and Butt, 2019) with the ANN model that generate the density log, two types of tests were made. The first test consisted of comparing the mean squared error of the ANN model with the mean squared errors for the traditional and modified Gardner equation utilizing the original parameter values from the articles, the second test consisted of comparing the ANN model with traditional and modified Gardner models in which parameters were calculated in a well log from a different field utilizing linear and non-linear regression.

RESULTS

Figure 2 shows the adjustments of the density calculated by the empirical equations and the density obtained by the ANN. The blue curve is the real density, the red curve is the density calculated by the empirical equation and the black curve is the density calculated by the ANN model. Table 2 shows the values of mean squared error (MSE) and correlation factor (R) obtained during the training, validation, and testing steps of the ANN algorithm while creating the ANN model.

Table 2: Number of samples, Mean Squared Error and correlation factor R for the ANN model.

	Samples	MSE	R
Training	55735	0.00421	0.86092
Validation	6967	0.00413	0.86487
Testing	6967	0.00438	0.85900

The Table 3 shows the mean squared errors for the ANN, empirical model of Gardner et al. (1974) and Oloruntobi and Butt (2019).

For the data set composed by a Campos basin and a Penobscot field well log, tests were performed using the first ANN model, and the adjustment is shown in the Figures below. The mean squared errors are displayed in Table 3, the mean squared error between the real density and the density calculated by the ANN model for a Campos basin well log was 0.0097 while the mean squared error for the same field but with the traditional Gardner equation utilizing the parameters K and B from Gardner et al. (1974) was 0.0147 as shown in Figure 2. When the parameters

Table 3: Mean Squared Errors for a well data from the Campos basin and for a well from the Penobscot basin. A.V. refers to the parameters' values from their respective original article while N refers to the parameters calculated in a Norne field data.

Model	MSE (Campos)	MSE (Penobscot)
Traditional Gardner (A.V.)	0.0147	0.0105
Modified Gardner (A.V.)	0.0209	0.0105
Traditional Gardner (N)	0.0260	0.0254
Modified Gardner (N)	0.0355	0.0099
ANN	0.0097	0.0074

k and B were calculated on a well log from the Norne field and applied on the Campos basin, the mean squared error was 0.0260.

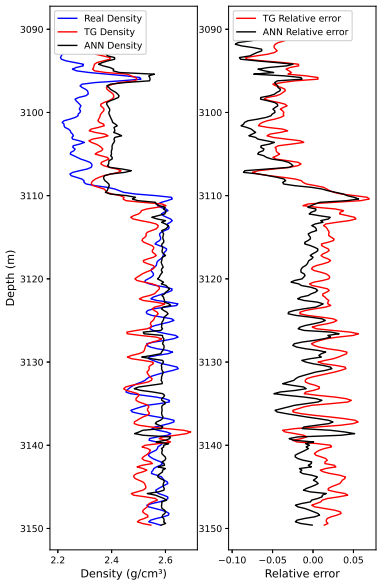

Figure 2: Comparison between the fit of the density log constructed with the traditional Gardner model with its original article parameter values (red curve) and the density log constructed by the ANN model (black curve). Both applied in a Campos basin data.

The Figure 3 shows the adjustment for the Penobscot field data comparing the ANN model with traditional Gardner's equation, the ANN model had a mean squared error of 0.0074 while Traditional Gardner's equation had a mean squared error of 0.0105. When the parameters k and B were calculated on the Norne field, Gardner's equation performed worse, showing a mean squared error of 0.0254 which is also greater than the error presented by the ANN, the adjustment is shown in Figure 4.

Similar tests were conducted comparing the modified Gardner's equation of Oloruntobi and Butt (2019). The mean squared error was 0.0209 for the modified Gardner in a Campos basin well log utilizing the parameters A, G, and m from Oloruntobi and Butt (2019). When the parameters were calculated by non-linear regression from Norne field data, the mean squared error was 0.0355, both results showed a larger error compared to the ANN model which

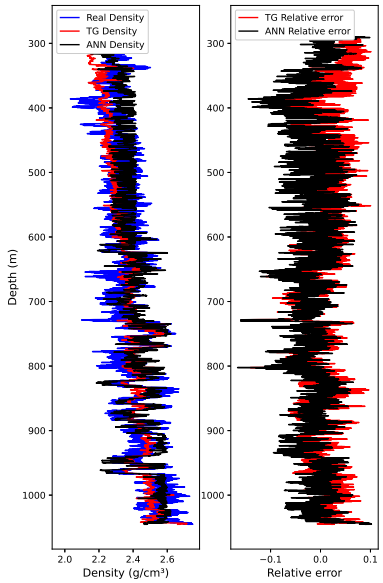

Figure 3: Comparison between the fit of the density log constructed with the traditional Gardner model with its original parameter values (red curve) and the density log constructed by the ANN model (black curve). Both applied in the Penobscot data.

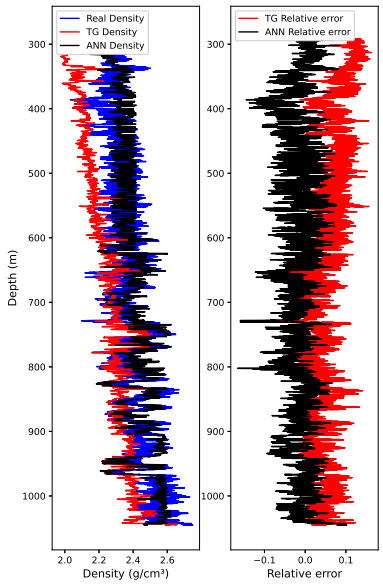

Figure 4: Comparison between the fit of the density log constructed with the traditional Gardner model with its parameter values calculated in a Norne field well log by linear regression (red curve) and the density log constructed by the ANN model (black curve). Both applied in the Penobscot data.

was 0.0097, these results are shown in Table 3. Figure 5 show the adjusted curves and relative errors.

When the modified Gardner equation was applied to the Penobscot data set utilizing the original parameters from the article, the mean squared error was 0.0105, while the

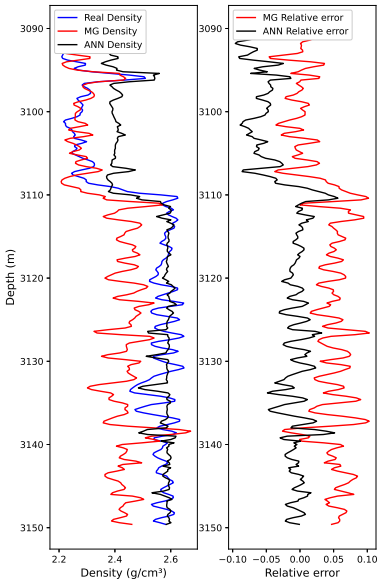

Figure 5: Comparison between the fit of the density log constructed with the modified Gardner model with its article values (red curve) and the density log constructed by the ANN model (black curve). Both applied in a Campos basin data.

parameters calculated by non-linear regression on a well log from Norne showed a mean squared error of 0.0099, both errors are larger than the ANN model error which was 0.0074. Figure 6 show the ANN and empirical equations fits and their respective relative errors.

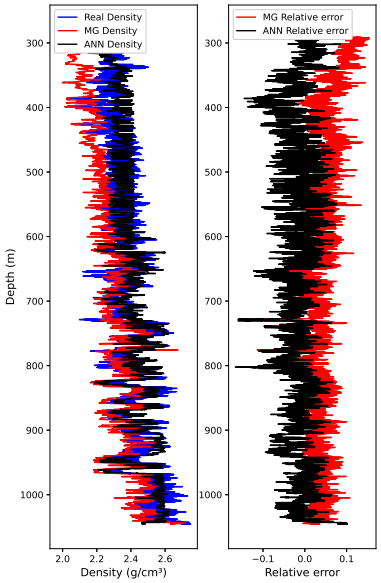

Figure 6: Comparison between the fit of the density log constructed with the modified Gardner model with its original parameter values (red curve) and the density log constructed by the ANN model (black curve). Both applied in the Penobscot data.

DISCUSSION

The original parameters values utilized in the empirical equations of Gardner et al. (1974) and Oloruntobi and Butt (2019) were very close to the values obtained by linear and non-linear regression for the first data set in this work which contains geologic information from very distinct sedimentary basins, however, the ANN model created in this work, tested with the same geologic data, showed smaller mean squared error values and the correlation factor values were closer to one.

The empirical equations of Gardner et al. (1974) and Oloruntobi and Butt (2019) had a good performance on constructing the density log by presenting a low mean squared error and high correlation factor, however, when the equations' parameters were obtained utilizing linear and non-linear inversion for a reference well from the Norne field data, the errors increased significantly, showing that the article parameters are a better option for the calculation.

It was possible to observe that utilizing the equation of Oloruntobi and Butt (2019) with the parameters obtained by inversion in a Norne field well log, it presented a good adjustment when applied in the Penobscot field data, which might indicate that both fields have similar lithology composed of shaly rocks. For the comparison of the ANN model with the empirical models on the Penobscot data, it is possible to observe in Figures 4 and 6, that in depths ranging from 300 m to 650 m, the ANN model makes a much better adjustment with the real density curve than the empirical models, from 650 m to 1000 m, the ANN density and modified Gardner density are close, however, it is possible to observe in the Figure that the relative error of the ANN model is close to zero while the relative error of the modified Gardner (MG) model is closer to 10%.

For the Campos basin well log utilized on the tests for the comparison of the empirical models with the ANN model shown in Figures 2 and 5, it is possible to observe that from depths between 3095 m and 3110 m, the traditional Gardner and the ANN model do not make a good adjustment compared to the modified Gardner model, which might indicate a very shaly formation or a wall's caving, however, for the rest of the well, for depths from 3110 m to 3150 m for example, the ANN model makes a better adjustment than both empirical models.

References

Beale, M. H., M. T. Hagan, and H. B. Demuth, 2010, Neural network toolbox: User's Guide, MathWorks, **2**, 77–81.

Birch, F., 1960, The velocity of compressional waves in rocks to 10 kilobars: 1.: Journal of Geophysical Research, **65**, 1083–1102.

Brocher, T. M., 2005, Empirical relations between elastic wavespeeds and density in the earth's crust: Bulletin of the seismological Society of America, **95**, 2081–2092.

Christensen, N. I., and W. D. Mooney, 1995, Seismic velocity structure and composition of the continental crust: A global view: Journal of Geophysical Research: Solid Earth, **100**, 9761–9788.

Clavier, C., W. Hoyle, D. Meunier, et al., 1971, Quantitative interpretation of thermal neutron decay time logs: part i. fundamentals and techniques: Journal of Petroleum Technology, **23**, 743–755.

Gardner, G., L. Gardner, and A. Gregory, 1974, Formation velocity and density—the diagnostic basics for stratigraphic traps: Geophysics, **39**, 770–780.

Han, D.-h., A. Nur, and D. Morgan, 1986, Effects of porosity and clay content on wave velocities in sandstones: Geophysics, **51**, 2093–2107.

Hebb, D. O., 1949, The organization of behavior: a neuropsychological theory: J. Wiley; Chapman & Hall.

Kailath, T., 1980, Linear systems: Prentice-Hall Englewood Cliffs, NJ.

Kayri, M., 2016, Predictive abilities of bayesian regularization and levenberg–marquardt algorithms in artificial neural networks: a comparative empirical study on social data: Mathematical and Computational Applications, **21**, 20.

Kohli, A., and P. Arora, 2014, Application of artificial neural networks for well logs: IPTC 2014: International Petroleum Technology Conference, European Association of Geoscientists & Engineers, 1–8.

Leon, S. J., I. Bica, and T. Hohn, 1998, Linear algebra with applications: Prentice-Hall Upper Saddle River, NJ.

Levenberg, K., 1944, A method for the solution of certain non-linear problems in least squares: Quarterly of applied mathematics, **2**, 164–168.

Lindseth, R. O., 1979, Synthetic sonic logs—a process for stratigraphic interpretation: Geophysics, **44**, 3–26.

Long, W., D. Chai, F. Aminzadeh, et al., 2016, Pseudo density log generation using artificial neural network: Presented at the SPE Western Regional Meeting, Society of Petroleum Engineers.

Lv, C., Y. Xing, J. Zhang, X. Na, Y. Li, T. Liu, D. Cao, and F.-Y. Wang, 2017, Levenberg–marquardt backpropagation training of multilayer neural networks for state estimation of a safety-critical cyber-physical system: IEEE Transactions on Industrial Informatics, **14**, 3436–3446.

Marquardt, D. W., 1963, An algorithm for least-squares estimation of nonlinear parameters: Journal of the society for Industrial and Applied Mathematics, **11**, 431–441.

McCulloch, W. S., and W. Pitts, 1943, A logical calculus of the ideas immanent in nervous activity: The bulletin of mathematical biophysics, **5**, 115–133.

Oloruntobi, O., and S. Butt, 2019, The new formation bulk density predictions for siliciclastic rocks: Journal of Petroleum Science and Engineering, **180**, 526–537.

Rochester, N., J. Holland, L. Haibt, and W. Duda, 1956, Tests on a cell assembly theory of the action of the brain, using a large digital computer: IRE Transactions on information Theory, **2**, 80–93.

Rosenblatt, F., 1958, The perceptron: a probabilistic model for information storage and organization in the brain.: Psychological review, **65**, 386.

Stigler, S. M., 1981, Gauss and the invention of least squares: the Annals of Statistics, 465–474.

Van der Baan, M., and C. Jutten, 2000, Neural networks in geophysical applications: Geophysics, **65**, 1032–1047.

Widrow, B., and M. E. Hoff, 1960, Adaptive switching circuits: Technical report, Stanford Univ Ca Stanford Electronics Labs.