Generation of Synthetic Density Log Response using Multivariate Adaptive Regression Splines

S. Asante-Okyere[1] and H. Osei[1]

[1]Department of Petroleum and Natural Gas Engineering, School of Petroleum Studies, University of Mines and Technology, Tarkwa, Ghana.

Asante-Okyere S. and Osei H. (2022), "Generation of Synthetic Density Log Response using Multivariate Adaptive Regression Splines", *Ghana Mining Journal*, Vol. 22, No. 2, pp. 49 - 57.

Abstract

There are instances in well logging operations where log response can be missing or inaccurate for a specific depth of interest due to wellbore conditions such as wellbore size, wellbore rugosity and mud cake effects. The conventional approach is to rerun the logs at definite depths, however, this remedial technique is costly, time-consuming and prone to errors due to the presence of a casing. Machine learning methods are currently implemented as an innovative way of predicting missing log responses. The present study seeks to investigate the potential of multivariate adaptive regression splines (MARS) as a density log predictive model. The performance of the developed MARS model was judged with the widely used artificial neural networks (ANN). The results reveal that MARS generalise better when predicting the density log response of the testing data. The MARS density log model achieved the highest correlation of 0.869, an error rate of 0.01196 and 0.1094 for *MSE* and *RMSE* respectively on the withheld dataset. While back propagation neural network (BPNN) and radial basis function neural network (RBFNN) had 0.855 and 0.802 as *R*, 0.0128 and 0.0147 as *MSE*, 0.1131 and 0.1212 as *RMSE* respectively. Therefore, a cost-effective MARS model can accurately generate synthetic density well log response.

Keywords: Multivariate Adaptive Regression Spline, Density Log, Radial Basis Function Neural Network, Back Propagation Neural Network

1 Introduction

Well logs have played a vital role in the global petroleum and gas exploration and reservoir evaluation since their first successful implementation at the beginning of the 20th century. They are essential for the oil and gas industry to understand the in situ subsurface petrophysical properties (Alger *et al*., 1963). Widespread field application has demonstrated that the measurement of formation density is a valuable and revealing technique for well logging, and as such the formation density tool was devised. The density well log tool measures the bulk densities of formations and corresponds to the bulk density variations of the formation. Therefore, density log affords valuable data on porosity, lithology and fluid content of formations (Misra and Wu, 2020).

If there exist wellbore conditions that prevent a direct contact to the borehole walls, the reading of the tool can be affected leading to unreliable results. The most critical borehole environments that impact the quality of the density log and the validity of the measured data are wellbore size, wellbore rugosity and mud cake effects. Also, a well log suite may miss some definitive logs for the entire well depth or specific depths of interest. These missing log responses may be required in the estimation of particular reservoir properties (Rezaee and Applegate, 1997; Rajabi *et al*., 2010). Such drawbacks in the well logging operation may need new drilling or a rerunning of the logging tool. On the other hand, drilling a new well or rerunning the logging tool generates additional cost and some log types cannot be measured due to the well casing.

To combat this challenge, recent studies have proposed the application of machine learning techniques in generating synthetic well logs (Chen *et al*., 2005; Rolon *et al*., 2009; Korjani *et al*., 2016; Akkurt *et al*., 2016; Blanes de Oliveira and Carneiro, 2021). The widely utilized neural network algorithm in generating density log data has been the Back Propagation Neural Network (BPNN). Long *et al*. (2016) incorporated clustering method and BPNN in predicting density well log. Similarly, BPNN was adopted in estimating density and resistivity logs by Salehi *et al*. (2017). Other well log types such as Photo-Electric (PE) and Unconfined Compressive Strength (UCS) logs which are difficult to obtain were compared by BPNN, decision tree, gradient boosted and random forest in Akinnikawe *et al*. (2018). Deep learning algorithms have also been employed for generating synthetic well log data in studies such as, Kim *et al*. (2020), Zhang *et al*. (2018) and Tatsipie and Sheng (2021).

*Manuscript received April 27, 2022
Revised version accepted September 22, 2022
https://dx.doi.org/10.4314/gm.v22i2.8

It is important to mention that Multivariate Adaptive Regression Splines (MARS) can accurately model non-linearities and extract patterns among different variables (Friedman, 1991; Hastie *et al*., 2008). Compared to other machine learning models, MARS is more precise and easier to implement (Fard *et al*., 2019). Considering the scarcity of studies on the potential of MARS in predicting well log response, the present research investigated the capability of MARS in generating accurate density well log response from other well logs. The performance of the MARS model was quantitatively compared to the widely used machine learning models of back propagation and radial basis function neural networks.

2 Resources and Methods Used

2.1 Data

The well log data used for the purpose of this study was from two wells located in the Ordos Basin in the northwest of China. The well logs are composed of Gamma ray (GR), caliper (CAL), sonic travel time (AC), bulk density (DEN) and compensated neutron (CNL). The input variables were GR, CAL, AC and CNL while DEN became the output. The visual description of the well log data used for the research is shown in Figures 1 and 2. At a sampling depth of 0.25 m, the total number of well log points of Well A for training the predictive models was 4475 and 1118 data points of Well B were considered the testing data.

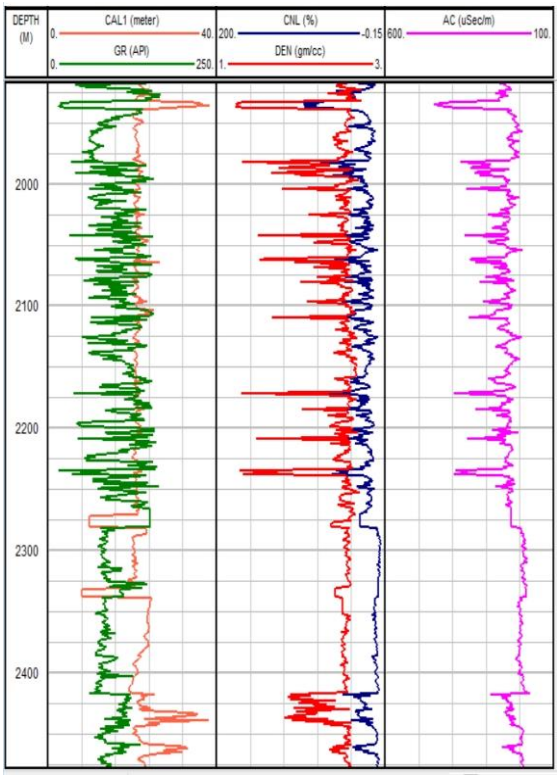

Fig. 1 Well Log Data for Well A used for Training the Models

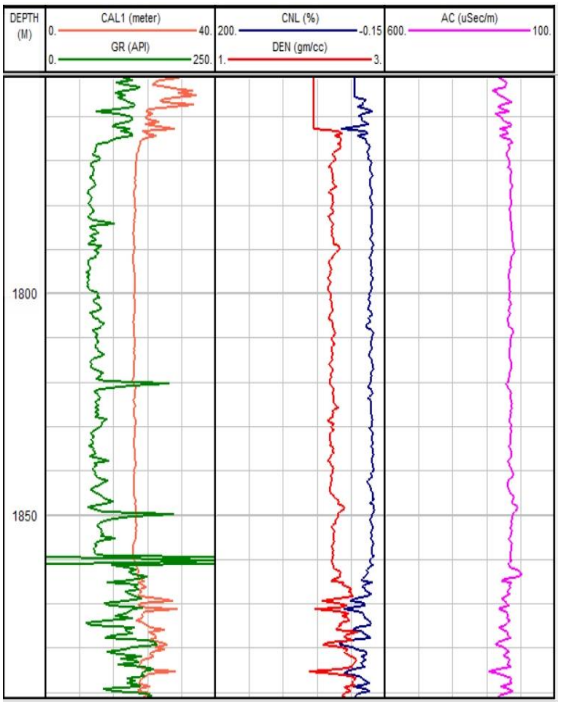

Fig. 2 Well Log Data for Well B used for Testing the Models

2.2 Multivariate adaptive regression splines

MARS is a non-parametric technique for generating non-linear patterns within a set of dependent and independent variables. The MARS method is

composed of a continuous piecewise polynomial function (splines) and can be expressed as (Friedman, 1991);

$$f(x) = \sum_{i=1}^{n} k_i BF_i(x) \quad (1)$$

where *ki* represents the least squares generated coefficient and *BFi* is the basis function. For the purpose of this study, *x* represents the input well logs. The MARS algorithm uses the optimal basis function and the knot, which is the position of where two individual spline polynomial functions coincide, to develop the model.

The MARS model is developed by initially stacking up knots and basis functions in order to minimize the prediction error. After the initial stacking up process is completed, the remaining basis function that did not contribute to the prediction outcome will be eliminated. The basis function elimination is performed using an optimal number of terms and they are commonly determined using the Generalized Cross-Validation (GCV) index. The GCV is best adopted based on its less computational requirements (Friedman, 1991; Hastie *et al*., 2008; Qi *et al*., 2021). In this study, MARS was implemented using the Salford Predictive Modeler (SPM) version 8.

2.3 Back propagation neural network

Due to its simplicity and efficient computation of the gradient descent, BPNN is the widely used supervised neural network. The general operation of the BPNN algorithm requires a forward propagation of input data and reverse transmission of output error. The main objective of BPNN is to obtain the optimal weights that can generate the least error margin. The mathematical expression of the optimum weight is provided in Eq. (2) (Konate´ *et al*., 2015).

$$W^* = \arg\min E_p(w) \quad (2)$$

where *w* is weight matrix and *Ep(w)* is an objective function on *w*. *E(w)* is the error that is to be reduced at any point of *w* as seen in Eq. (3) (Konate´ *et al*., 2015; Asante-Okyere *et al*., 2018).

$$E(w) = \sum_{p} E_p(w) \quad (3)$$

where *p* is the number of training samples and the error for each sample well log point is given as *Ep(w)* as represented in Eq. (4) (Konate´ *et al*., 2015; Asante-Okyere *et al*., 2018).

$$E_p = \frac{1}{2}\sum_{p}\left(d_{pj} - y_{pj}(w)\right)^2 \quad (4)$$

BPNN is influenced by the type of training algorithm, activation function, and more importantly, the number of hidden neurons (Ziggah *et al*., 2016). When training the BPNN, the optimal number of hidden neurons was obtained through a sequential trial and error approach. The Levenberg-Marquardt algorithm was adopted as the training algorithm, while the hyperbolic tangent function was used as the activation function as provided in Eq. (5) (Ziggah *et al*., 2016). BPNN model was coded and implemented in MATLAB R2019b.

$$f(x) = \tanh(x) = \frac{2}{1+e^{-2x}} - 1 \quad (5)$$

2.3 Radial Basis Function Neural Network

RBFNN was initially used by Broomhead and Love (1988), and Moody and Darken (1989) to mitigate the overfitting and underfitting problem posed by the BPNN. The structure of RBFNN is similar to BPNN, with input, hidden and output layers. The hidden layer accommodates the radial basis function (Bullinaria, 2004). The radial basis function used in the study was the Gaussian function expressed in Eq. (6) (Bullinaria, 2004).

$$\phi_j(x) = \exp\left(-\frac{\|x - c_i\|^2}{2\delta_i^2}\right) \quad (6)$$

In the Gaussian function, the Euclidian distance between the centre of the basis function and the input vector *x* is determined as the radial distance d_i (Bullinaria, 2004).

$$d_i = \|x - c_i\|^2 \quad (7)$$

where, $\phi_j(x)$ denotes the output of the basis function and δ is the spread of the interpolated function. The output layer linearly sums up the weighted value from the hidden neurons in determining the overall output for the neural network using Eq. (8) (Bullinaria, 2004).

$$y_k(x) = \sum_{i=1}^{M} W_{k,i}\phi_j(x) \quad (8)$$

where $y_k(x)$ is the output from the kth output neuron. *M* is the total number of basis function and *W(k,i)* is the weight between kth and ith output neuron. The gradient descent rule was adopted to train the RBFNN. MATLAB R2019b was the software used to code and implement RBFNN model.

The workflow for developing the density log prediction models is represented in Fig. 3.

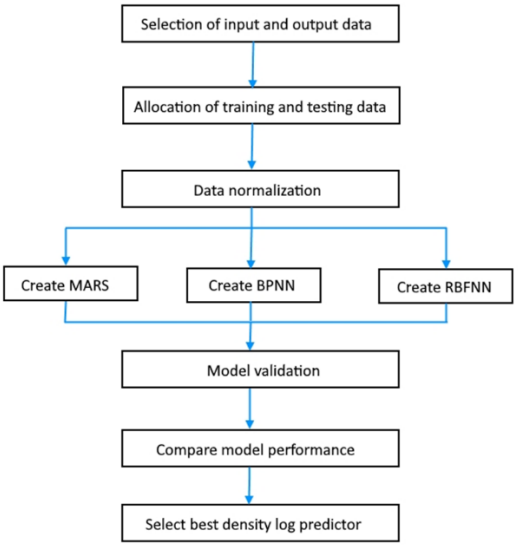

Fig. 3 Workflow of Developing MARS, BPNN and RBFNN for Density Log Prediction

2.4 Statistical Model

The performance of the proposed density log response prediction model from BPNN and RBFNN was determined using statistical parameters namely correlation coefficient (*R*), mean square error (*MSE*), and root mean square error (*RMSE*). The correlation between the prediction from the models and measured density log response was explained by the *R* value (Eq. 9). The *R* value ranges from 0 to 1, with a strong correlation existing if the *R* value approaches 1. The average square and the root average square deviation between the observed and predicted density values were also expressed using the *MSE* and *RMSE* scores shown in Eqs. (9) and (11) respectively. The MSE and RMSE value provides an indication of how much the estimates deviate from the measured density log data and reflects the quality of the predictive model.

$$R = \frac{\sum_{i=1}^{n}\left(m_i - \overline{m}\right)\left(g_i - \overline{g}\right)}{\sqrt{\sum_{i=1}^{n}\left(g_i - \overline{g}\right)^2}\sqrt{\sum_{i=1}^{n}\left(m_i - \overline{m}\right)^2}} \quad (9)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(m_i - g_i\right)^2 \quad (10)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(m_i - g_i\right)^2} \quad (11)$$

where *g* is the generated density log response, *m* is the measured density log response, and *n* is the total number of observations

3 Results and Discussion

3.1 MARS model

During the calibration process of the MARS model, the best outcome was achieved at a maximum basis function of 15, with the maximum interactions of 1. Minimum observations between knots were specified as 0 and the degree of freedom for knot optimization was observed to be 3.

The training performance of the optimal MARS model for generating density well log response is summarized in Table 1. From Table 1, it was identified that the MARS model during training attained an *R* value of 0.903 which represents a strong positive correlation between the estimated density log response and the measured density log data. The error statistics of *MSE* and *RMSE* for the training stage of the model are 0.0104 and 0.102 respectively.

The trained MARS model was tested on the withheld data. From the statistical indicators, an *R* value of 0.869 was obtained. The error metrics of *MSE* and *RMSE* of 0.01196 and 0.1094, respectively.

Table 1 Training and testing performance of MARS model

Statistical Indicator	Train	Test
R	0.903	0.869
MSE	0.0104	0.01196
RMSE	0.1020	0.1094

3.2 BPNN Model

The optimal BPNN structure was identified at a structure of 4 inputs, a single hidden layer with 6 nodes and an output (*i.e.*, 4-6-1). The optimal BPNN had *R* value of 0.912 during training. When the trained BPNN model was tried on the withheld testing data of Well B, 0.855 was recorded as the *R* value.

The error metrics of *RMSE* and *MSE* recorded values of 0.09747 and 0.0095 for training, and 0.1131 and

0.0128 for validation, respectively. The performance of the BPNN is summarized in Table 2.

Table 2 Training and Testing performance for BPNN model

Statistical Indicator	Train	Test
R	0.9122	0.855
MSE	0.0095	0.0128
RMSE	0.0975	0.1131

3.2 RBFNN Model

The optimum architecture for the RBFNN was obtained by adjusting the spread parameter based on a sequential trial and error approach. Therefore, the best performing RBFNN structure was identified at a spread parameter of 0.9. This represents a RBFNN structure of 4 inputs, a hidden layer with a maximum of 50 hidden neurons, a width parameter of 0.9, and 1 output. The correlation coefficient obtained was 0.9161 for training. During testing on Well B, the estimation from the trained RBFNN had *R* value of 0.8019. The *RMSE* value during training and testing was 0.0954 and 0.1212, respectively. In contrast, the *MSE* score for the optimal RBFNN was 0.0091 and 0.0147 during training and validation, respectively. The performance of the RBFNN as the model parameter is summarised in Table 3.

Table 3 Training and Testing performance for RBFNN model

Statistical Indicator	Train	Test
R	0.9161	0.8019
MSE	0.0091	0.0147
RMSE	0.0954	0.1212

Table 4 Predictive Performance of MARS, BPNN and RBFNN

Testing	*R*	*RMSE*	*MSE*
MARS	0.869	0.1094	0.01196
BPNN	0.855	0.1131	0.0128
RBFNN	0.802	0.1212	0.0147

3.3 Predictive Performance

The practicality of the models was compared using the testing performance as the developed models were tried on the independent, withheld dataset. The testing results reveal the generalization capability of the models since the testing data did not contribute to the learning process of the models. The results as the models were tried on the independent testing data gives a clear indication of the performance of the developed models when implemented in real life situations.

From Table 4, the MARS model generated a more excellent density log prediction than BPNN and RBFNN. MARS had the highest correlation of 0.869, which reflects the density log response generated 86.9 % accuracy. In comparison, BPNN and RBFNN produced 0.855 and 0.802, respectively (Table 4). It was also identified that the prediction outcome from MARS had the least error margin when compared with BPNN and RBFNN. The predicted density log response from MARS had the least *MSE* and *RMSE* value as shown in Table 4.

Therefore, it can be inferred from the improved generalization ability that MARS is an excellent predictor for accurate density well log response from other well logs and can easily be implemented without the need to rerun the logging tool. A visual illustration of the performance of MARS, BPNN and RBFNN in generating density log response and the measured density log is shown presented in Figure 4.

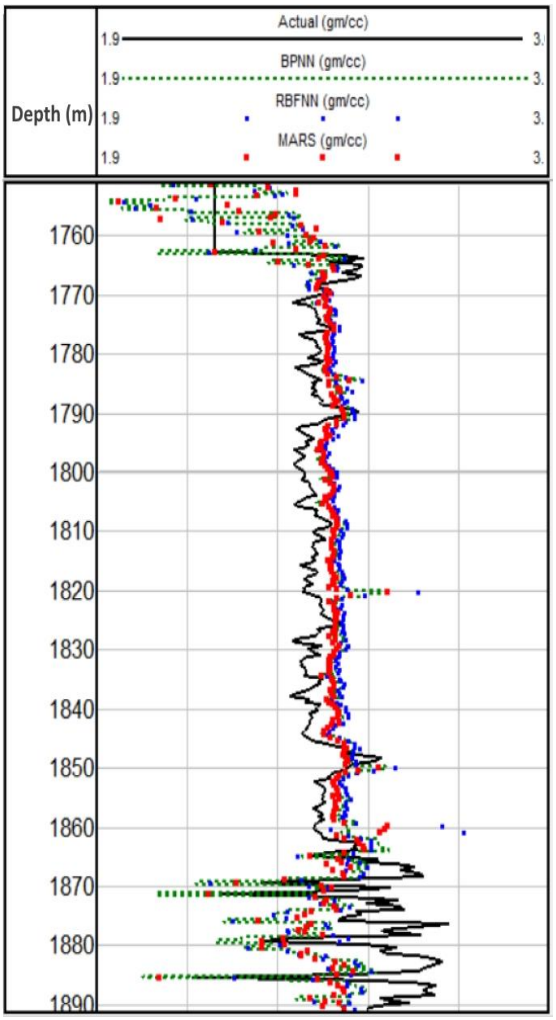

Fig. 4 Prediction Performance of Density Log Response for MARS, BPNN and RBFNN

4 Conclusions

In this research, an examination of the effectiveness of MARS in producing synthetic density log response was conducted using data from wells situated in the Ordos Basin in China. A comparative analysis was further performed to confirm the prediction output of the proposed MARS model with the widely adopted BPNN and RBFNN.

During training, the BPNN and RBFNN produced a correlation score (*R*) of 0.912 and 0.916, respectively. The error statistics of *MSE* and *RMSE* for BPNN and RBFNN were 0.0095 and 0.0975; 0.0091 and 0.0954, respectively for training. The MARS model obtained an *R* value of 0.903, *MSE* and *RMSE* score of 0.0104 and 0.102 respectively, during the training process.

However, it was identified during the testing stage that the MARS model outperformed the BPNN and RBFNN models when the predictive capability was examined. The MARS model produced density log estimates having an *R* value of 0.869 compared to the 0.855 and 0.802 generated by BPNN and RBFNN, respectively. Similarly, a lower error margin was recorded for MARS as it obtained 0.01196 and 0.1094 for *MSE* and *RMSE* respectively, compared to 0.0128 and 0.1131 for BPNN, 0.0147 and 0.1212 for RBFNN.

A deduction can therefore be drawn from the improved predictability of the MARS model that it is a reliable alternative in providing accurate density log predictions without the need to rerun the logging tool.

References

Akinnikawe, O., Lyne, S., and Roberts, J. (2018), "Synthetic Well Log Generation using Machine Learning Techniques", *SPE/AAPG/SEG Unconventional Resources Technology Conference*, Houston, Texas, USA, pp. 1-16.

Akkurt, R., Conroy, T. T., Psaila, D., Paxton, A., Low, J. and Spaans, P. (2018), "Accelerating and Enhancing Petrophysical Analysis with Machine Learning: A Case Study of an Automated System for Well Log Outlier Detection and Reconstruction", *SPWLA 59th Annual Logging Symposium*, London, UK, pp. 25.

Alger, R. P., Raymer, L. L., Hoyle, W. R. and Tixier, M.P. (1963), "Formation Density Log Applications in Liquid-Filled Hole", *Journal of Petroleum Technology*, Vol. 15, No. 3, pp. 321-332.

Asante-Okyere, S., Xu, Q., Mensah, R. A., Jin, C. and Ziggah, Y. Y., (2018), "Generalized Regression and Feed Forward Back Propagation Neural Networks in Modelling Flammability Characteristics of Polymethyl Methacrylate (PMMA)", *Thermochimica Acta*, Vol. 667, pp. 79–92.

Blanes de Oliveira, L.A. and Carneiro, C. (2021), "Synthetic Geochemical Well Logs Generation using Ensemble Machine Learning Techniques for the Brazilian Pre-Salt Reservoirs", *Journal of Petroleum Science and Engineering*, Vol. 196, pp. 108080.

Broomhead, D. S. and Lowe, D. (1988), "Multivariable Functional Interpolation and Adaptive Networks", *Complex Systems*, Vol. 2, pp. 321-335.

Bullinaria, J. A. (2004) "Introduction to Neural Networks", *Unpublished MSc Lecture Notes*, University of Birmingham, UK, pp. 1-16.

Chen, D., Quirein, J., Simth Jr, H., Hamid, S. and Grable, J. (2005), "Neural Network Ensemble

Selection using a Multi-Objective Genetic Algorithm in Processing Pulsed Neutron Data", *Petrophysics*, Vol. 46, No. 5, pp. 323–334.

Fard, M.B., Mirbagheri, S.A., Pendashteh, A. and Alavi, J., (2019), "Biological Treatment of Slaughterhouse Wastewater: Kinetic Modeling and Prediction of Effluent" *Journal of Environmental Health Science and Engineering*, Vol. 17, No. 2, pp.731-741.

Friedman, J. H. (1991), "Multivariate Adaptive Regression Splines", *Annals of Statistics*, Vol. 19, No. 1, pp. 1–67.

Hastie, T., Tibshirani, R., & Friedman, J. (2008), *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2nd edition, Vol. 1, Springer Publications, New York, pp. 1-350.

Kim, S., Kim, K. H., Min, B., Lim, J. and Lee, K. (2020), "Generation of Synthetic Density Log Data Using Deep Learning Algorithm at the Golden Field in Alberta, Canada", *Geofluids*, Vol. 2020, pp. 26.

Konate´, A. A., Pan, H., Khan, N. and Yang, J. H. (2015), "Generalized Regression and Feed-Forward Back Propagation Neural Networks in Modelling Porosity from Geophysical Well Logs", *J. Petrol Explor. Prod. Technol.*, Vol. 5, No. 2, pp. 157–166.

Korjani, M., Popa, A. S., Grijalva, E., Cassidy, S. and Ershaghi, I. (2016), "A New Approach to Reservoir Characterization using Deep Learning Neural Networks", *SPE Western Regional Meeting*, Anchorage, AK, USA, pp. 1-15.

Long, W. Chai, D. and Aminzadeh, F. (2016), "Pseudo Density Log Generation using Artificial Neural Network", *SPE Western Regional Meeting*, Anchorage, Alaska, pp. 1-21.

Misra, S. and Wu, Y. (2020), "Chapter 10 - Machine Learning Assisted Segmentation of Scanning Electron Microscopy Images of Organic-Rich Shales with Feature Extraction and Feature Ranking", in: Misra, S. and Wu, Y. (Eds.), Machine Learning for Subsurface Characterization, Gulf Professional Publishing, pp. 289-314.

Moody, J. and Darken, C. (1989), "Fast Learning in Networks of Locally-Tuned Processing Units", *Neural Comp*., Vol. 1, No. 2, pp. 281-294.

Rajabi, M., Bohloli, B. and Ahangar, E., (2010), "Intelligent Approaches for Prediction of Compressional, Shear and Stoneley Wave Velocities from Conventional Well Log Data: A Case Study from the Sarvak Carbonate Reservoir in the Abadan Plain (South-western Iran)", *J. Comput. Geosci.*, Vol. 36, pp. 647–664.

Rezaee, M. R. and Applegate, J. K., (1997), "Shear Velocity Prediction from Wire Line Logs, an Example from Carnarvon Basin, NW Shelf, Australia", *SEG (Society of Exploration Geophysicists) Expanded Abstracts*, Vol. 16, pp. 945–947.

Rolon, L., Mohaghegh, S. D., Ameri, S., Gaskari, R. and McDaniel, B. (2009), "Using artificial neural networks to generate synthetic well logs", *Journal of Natural Gas Science and Engineering*, Vol. 1, No. 5, pp. 118–133.

Salehi, M. M., Rahmati, M., Karimnezhad, M. and Omidvar, P. (2017), "Estimation of the Non Records Logs from Existing Logs using Artificial Neural Networks", *Egyptian Journal of Petroleum*, Vol. 26, No. 4, pp. 957–968.

Tatsipie, N. R. K. and Sheng, J. J. (2021), "Generating Pseudo Well Logs for a Part of the Upper Bakken using Recurrent Neural Networks", *Journal of Petroleum Science and Engineering*, Vol. 200, pp. 108253

Qi, X., Wang, H., Pan, X., Chu, J. and Chiam, K. (2021), "Prediction of Interfaces of Geological Formations using the Multivariate Adaptive Regression Spline Method", *Underground Space*, Vol. 6, No. 3, pp. 252-266.

Zhang, D., Chen, Y. and Meng, J., (2018), "Synthetic Well Logs Generation via Recurrent Neural Networks", *Petroleum Exploration and Development*, Vol. 4, pp. 629-639.

Ziggah, Y.Y., Youjian, H., Yu, X., Yu, X., Basommi, L.P., (2016), "Capability of Artificial Neural Network for Forward Conversion of Geodetic Coordinates (ϕ, λ, h) to Cartesian Coordinates (X, Y, Z)", *Math. Geosci.*, Vol. 48, pp. 687.

Authors

S. Asante-Okyere is currently a Lecturer in the Department of Petroleum and Natural Gas Engineering, Univerisity of Mines and Technology. He holds a PhD and MEng in Oil and Natural Gas Engineering from the China University of Geosciences, China and BSc from the Kwame Nkrumah University of Science and Technology, Ghana. He is a member of the Society of Petroleum Engineers (SPE), American Association of Petroleum Geologists (AAPG), International Association of Mathematical Geosciences (IAMG) and other professional organizations. His research interest is in the application of artificial intelligence for petrophysical analysis, reservoir characterization and source rock geochemistry.

H. Osei is currently a senior lecturer at the Department of Petroleum and Natural Gas Enginering of the University of Mines and Technology, Tarkwa, Ghana. He holds MSc degree in Petroleum Engineering from African University of Science and Technology, Abuja, Nigeria. He obtained his PhD degree in Mechanical Engineering from the Universiti Teknologi PETRONAS (UTP), Perak, Malaysia. He is a member of Society of Petroleum Engineers, International Association of Engineers, Universal Association of Mechanical and Aeronautical Engineers, Teaching and Education Research Association and Ghana Institution of Engineering and Technology. His research interests include reservoir and production engineering, multiphase flow, fluid dynamics and separations technology.