Total Organic Carbon (TOC) Prediction Using Machine Learning Methods Based On Well Logs Data

Rahmat Catur Wibowo[1, a)], Ordas Dewanto[1, b)], and Muh. Sarkowi[1, c)]

[1]*Geophysical Engineering, Engineering Faculty, Universitas Lampung*
Sumantri Brojonegoro Street, No. 1, Rajabasa, Bandar Lampung City, Lampung, 35145, Indonesia

[a)] Corresponding author: rahmat.caturwibowo@eng.unila.ac.id
[b)] ordasdewanto@gmail.com
[c)] muh.sarkowi@eng.unila.ac.id

Abstract. TOC (Total Organic Carbon) Determination is very important for the evaluation of each source rock unit. Methods that rely on extensive laboratory testing are limited by the availability and integrity of rock samples. Prediction of TOC from wells log data is available for most wells drilled, providing a rapid evaluation of organic content, providing a continuous record while eliminating sampling problems. Therefore, the ideal method for determining the TOC fraction in source rock units would be to use general well log data. The purpose of this study is to apply a machine learning method with the Multi-Layer Perceptron Artificial Neural Network (ANN) technique to predict the best correlation from a new empirical correlation that can be used to estimate TOC using well logs. Eighteen data points from the Talang Akar Formation were used for training and testing the MLP-ANN model. The results obtained show that the MLP-ANN model predicts TOC using only well logs: bulk density (RHOB), compressional time (DT), deep resistivity (ILD), gamma-ray (GR), and neutron porosity (NPHI) with accuracy. high (CC 0.98 and average absolute percentage error (AAPE) 5%). The TOC correlation that has been developed is simple and can be applied using any computer without the need for an ANN model or special software. The developed technique will help reservoir geophysicists and geologists to estimate TOC values using only well logs with high accuracy.

INTRODUCTION

Shale is an alternative to conventional natural gas and is currently a popular material in petroleum exploration and production. With this type of fuel, the evaluation of the hydrocarbon potential is very important. Total Organic Carbon (TOC) content is one of the most important parameters to evaluate shale quality and hydrocarbon potential. In previous years, TOC was directly measured from core rock samples in the laboratory; However, this method is highly dependent on variations in the core rock sample. The associated TOC measurements are also limited and cannot satisfy the evaluation of the entire rock formation. The host rock has a specific response to gamma-ray (GR), density (RHOB), transit time velocity (Dt), and resistivity (ILD/LLD), which makes it distinguishable from surrounding rocks [1-4]. Therefore, wireline logs can be used as source rock potential indicators. Wireline log-based TOC predictions can provide continuous and in-situ potential evaluation results of target formations through boreholes.

Previous researchers have studied the TOC evaluation method based on wireline logs. Fertl and Rieke (1980) [5] and Fertl and Chilingar (1988) [6] used GR spectral logs to identify organic-rich rocks and analyze their relationship to total GR, uranium, and thorium-potassium ratios. Schmoker and Hester (1983) [7] proposed the log RHOB technology for TOC estimation, which is well established and frequently used for various shale formations. Mendelzon and Toksoz (1985) [8] used a multivariate method to establish a quantitative relationship between wireline logs and TOC determined using core rock samples; the regression equation also has a high coefficient of determination. Carpentier et al. (1991) [9] proposed the CARBOLOG method to obtain an in-situ estimate of TOC content for the long sonic transit time and high resistivity in organic matter-rich shales. However, this requires calibration based on available TOC data measured in the laboratory [10]. In addition, Passey et al. (1990) [11] proposed the DlogR technique, which performs an overlay between porosity logs (e.g., sonic, density, and neutron) and resistivity logs to calculate TOC content. Kamali and Mirshady (2004) [12] combined the DlogR method and the Neuro-Fuzzy approach to determine the existing TOC content; it can also be used on gas carrier shale [13]. Passey et al. (2010) [14] revised the calibration of the DlogR method to include organic-rich rocks of high maturity and identified mature hydrocarbon source rocks from gas-bearing shale. Bakhtiar et al. (2011) [4] applied DlogR and neural network methods to estimate TOC content for source rock evaluation. Therefore, the DlogR technique is the most popular TOC evaluation method using cable logs. However, the DlogR technique must manually select a baseline, and the background TOC level varies regionally and is difficult to determine. In recent years, intelligent systems and neural networks have been applied to TOC prediction. Kadkhodaie-Ilkhchi et al. (2009) [15] applied a committee engine to estimate TOC content from petrophysical data. Khoshnoodkia et al. (2011) [16] also used an intelligent method to investigate the TOC determination of the Gadvan formation using conventional wireline logs. However, TOC calculations using neural networks are complex and involve many parameters; selecting the relevant parameters is usually tricky.

In most cases, TOC estimation is achieved by constructing multivariate or straightforward regression methods. In addition, in recent years, several direct TOC estimation techniques have been introduced, including geochemical wireline logs. The technology can provide direct TOC measurements without complex adjustment algorithms [17-18]; however, this technology has not been widely applied because only a few wells were measured.

Machine learning (ML) and artificial intelligence (AI) are both captivating fields that integrate computing power with human intelligence to produce intelligent and reliable solutions to highly nonlinear and highly complex problems [19]. Engineering journals have reported many articles using AI and ML for regression, function approximation, and classification problems in the last two decades. With the advent of soft computing techniques, several correlations utilizing AI field techniques have emerged, especially in reservoir characterization [20-21], reservoir engineering, and reservoir geomechanics. In petroleum geochemistry, this correlation can be seen in the work of Rahaman and Vasant (2020) [22]. In recent years, the multi-layer perceptron (MLP) artificial neural networks (ANN) have been actively used by researchers to predict porosity [23]; researchers have also used MLP-ANN to predict permeability [24].

Based on the literature study, it is known that much attention is paid to predicting the TOC content of shale rich in organic matter. However, the TOC prediction model cannot predict TOC accurately or requires a lot of laboratory data (core) to determine the suitable parameters (fit). Therefore, this study intends to investigate the log-based TOC prediction for Talang Akar organic material-rich shale using the MLP-ANN technique. This study will utilize ML and evolutionary algorithms to obtain an optimal model.

METHODS

In this study, conventional well logs of RHOB, DT, ILD, GR, and NPHI, collected from the Talang Akar shale, were used to train the MLP-ANN model to predict the TOC measured in the laboratory. This ML model is used in this study to estimate TOC because of its proven high accuracy in evaluating petroleum and related geological parameters. A total of 18 core data points and log data were collected from the Talang Akar shale. Figure 1 (columns 2-5) shows the log data collected from the Talang Akar shale used to develop the model. The ML model optimization process is continued until the minimum absolute percentage error (AAPE) average value and the highest coefficient of determination (R^2) are obtained between the predicted TOC and the measured core. The trained and optimized ML model was then tested using another data set from the same well and validated using data points collected from Gumai.

The relative importance of selected training well log data on the predictability of TOC values was then studied. I was comparing the relative importance between the different conventional well logs used to train the ML model and the TOC values measured in the laboratory. The data used to train the ML model, TOC, highly correlate on RHOB, while moderately related with DT, ILD, GR, and NPHI.

Five conventional well logs used to train the ML model were selected based on their relative importance to the core measured TOC. However, the selection was consistent with their reported association with TOC. For example, ILD is believed to be influenced by the presence of kerogen in the source rock [25]; DT decreased with increasing TOC [26]; some studies have confirmed that GR can significantly improve TOC prediction [27], but the relationship is controversial for others; RHOB decreased with increasing kerogen content, and hence, organic matter in the formation increased; and NPHI is sensitive to the presence of organic matter even in small amounts and must go through an inorganic material correction to find out [11]. Due to the reasons mentioned above, the five conventional well logs RHOB, DT, ILD, GR, and NPHI, were considered to develop the TOC model in this study.

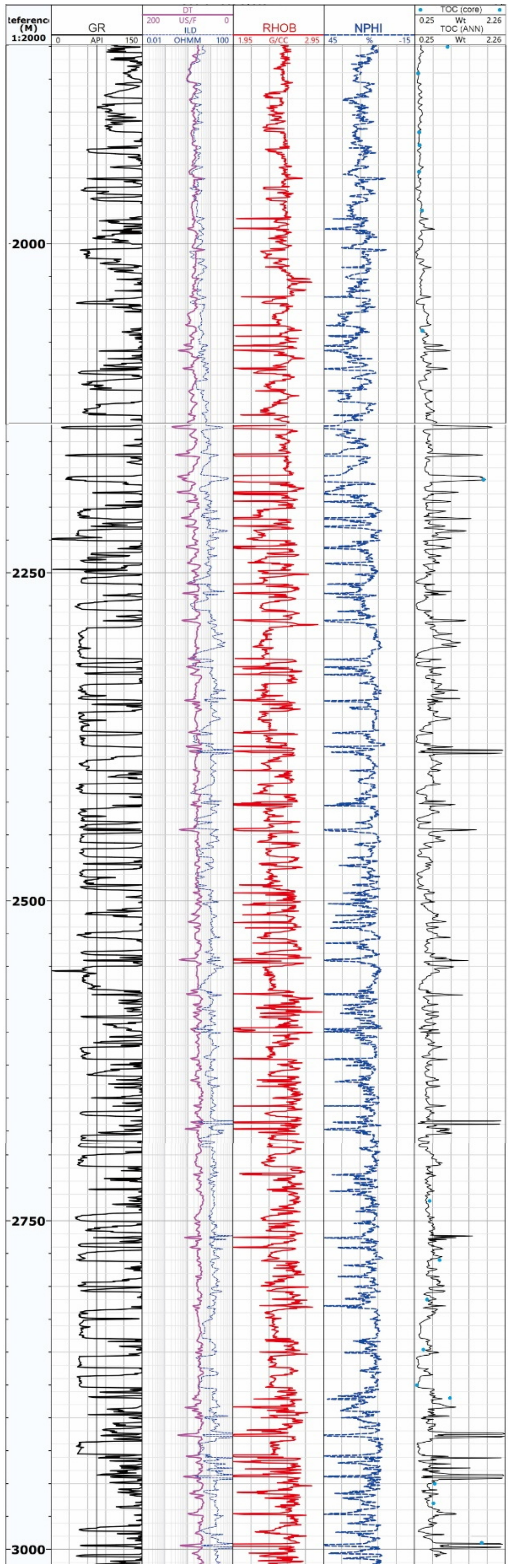

FIGURE 1. Well log data was collected from Talang Akar shale formation to develop the ML models (column 2-5, start from left) and TOC MLP-ANN prediction compare versus TOC core (column 6).

Principles of MLP-ANN

Multilayer Perceptron (MLP) is a popular supervised learning technique in ANN whose architecture has been used for several forecasting problems in the literature [24]. It is a distributed mathematical model inspired by the behavior of the human brain and nervous system. MLP consists of three layers; input layer, hidden layer, and output layer. The hidden layer may have one or more activation functions [28]. The input for this research is five conventional well logs, while the output is Total Organic Carbon. Figure 2 shows a schematic diagram of the MLP-ANN and the respective neural network architecture.

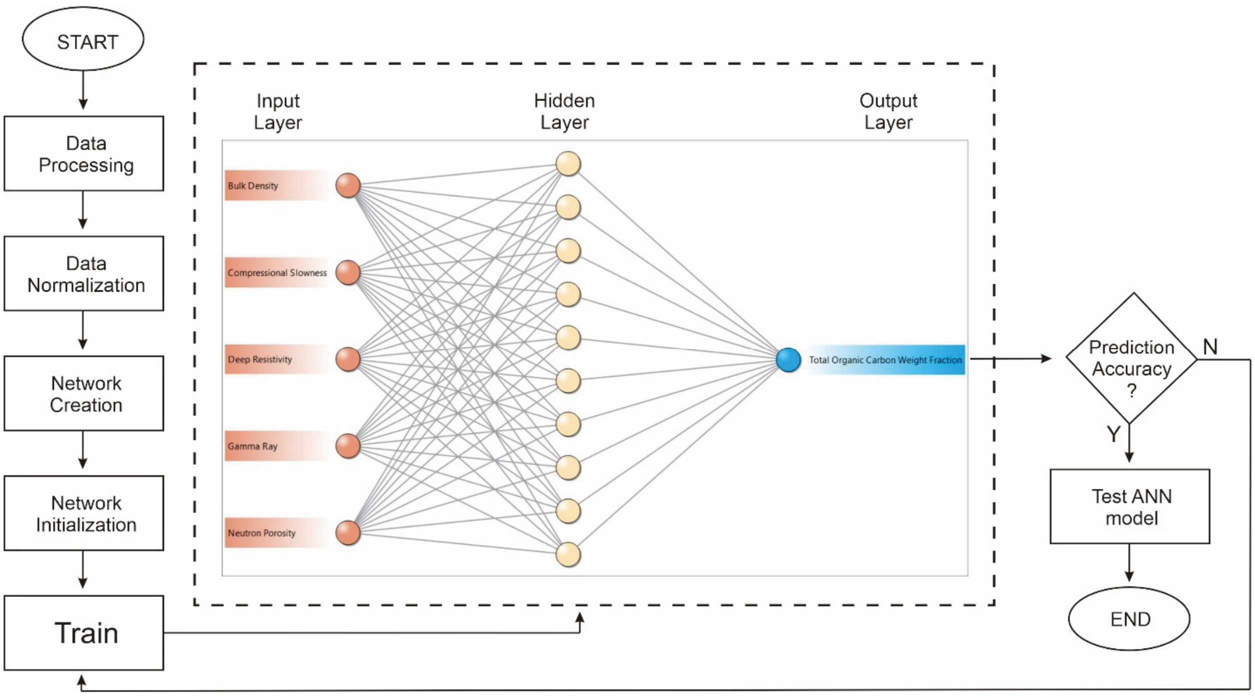

FIGURE 2. The schematic diagram for MLP-ANN.

RESULTS AND DISCUSSION

Figure 1 shows that the MLP-ANN model predicts TOC content with high accuracy with CC 0.87 and AAPE 10.0% (after validation). Another data set from the same well was used to test the MLP-ANN model using invisible data. TOC ranges from 0.3 to 1.8% by weight; RHOB ranges from 2.0 to 2.8 g/cc; DT ranges from 52 to 124 us/ft; ILD ranges from 4 to 163-ohm m; GR ranges from 27 to 166 degrees API, and NPHI ranged from 2 to 45%. The test data range lies within the same range as the training data. Figure 3 confirms the high accuracy of the MLP-ANN model in predicting TOC, where R^2 is 0.98. Based on these results, it can be concluded that the MLP-ANN model can be used to predict TOC as a function of the well logs of ILD, GR, DT, RHOB, and NPHI. To avoid overfitting, 15% of the data points were used randomly to assess the generalizability of the developed network. Figure 4 shows that the optimum number of iterations, resulting in the lowest mean square error (MSE), is 49. Network performance on the test dataset is also considered (where MSE is calculated for each evaluation). Network training and testing performance should be good. If the training alone is good, it will be considered overfitting, and the network is not selected.

CONCLUSION

In this study, an ML model based on the MLP-ANN technique was developed to estimate total organic carbon (TOC) using conventional deep well logs of resistivity, gamma rays, sonic transit time, neutron porosity, and bulk density. The model was developed and tested using data collected from the Talang Akar shale and then validated using invisible data from the Gumai shale. The ML model showed high TOC predictability for both formations evaluated in this study. The model is adequate for estimating TOC using the well logs used in this study (CC 0.87 and AAPE 10%).

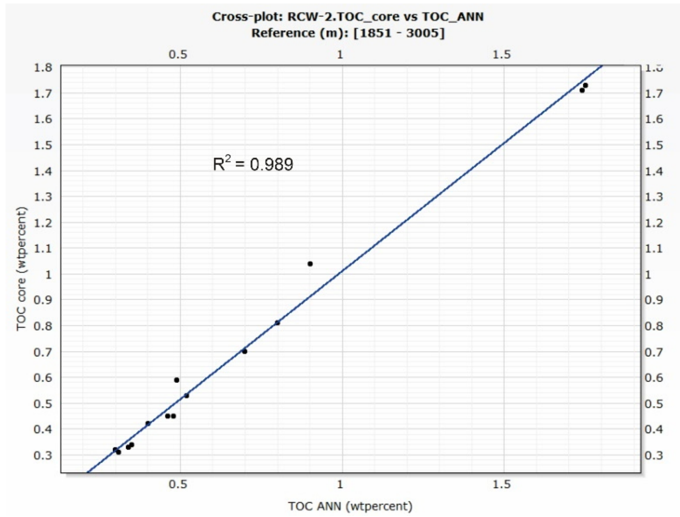

FIGURE 3. Coefficient of determination for TOC prediction for training data (18 data points) using the MLP-ANN model.

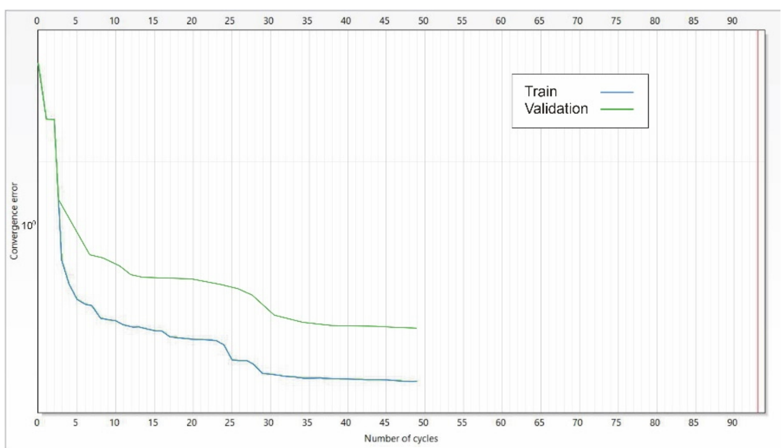

FIGURE 4. Training and validation errors.

ACKNOWLEDGMENTS

We would like to thank the LPPM, the University of Lampung, which has funded this research, and all those who have supported the completion of this article.

REFERENCES

1. Sachsenhofer RF, Leitner B, Linzer HG, Bechtel A, Ćorić S, Gratzer R, et al. Deposition, erosion and hydrocarbon source potential of the Oligocene Eggerding Formation (Molasse Basin, Austria). Austrian J Earth Sci. 2010;103(1):76–99.
2. Bodin S, Fröhlich S, Boutib L, Lahsini S, Redfern J. Early toarcian source-rock potential in the central high atlas basin (central morocco): Regional distribution and depositional model. J Pet Geol. 2011;34(4):345–63.
3. El Sharawy MS, Gaafar GR. Application of well log analysis for source rock evaluation in the Duwi

Formation, Southern Gulf of Suez, Egypt. J Appl Geophys [Internet]. 2012;80:129–43. Available from: http://dx.doi.org/10.1016/j.jappgeo.2011.12.005
4. Amiri Bakhtiar H, Telmadarreie A, Shayesteh M, Heidari Fard MH, Talebi H, Shirband Z. Estimating total organic carbon content and source rock evaluation, applying logr and neural network methods: Ahwaz and marun oilfields, SW of Iran. Pet Sci Technol. 2011;29(16):1691–704.
5. Fertl WH, Rieke HH. Gamma Ray Spectral Evaluation Techniques Identify Fractured Shale Reservoirs and Source-Rock Characteristics. JPT, J Pet Technol. 1980;32(11):2053–62.
6. Fertl WH, Chilingar G V. Total Organic Carbon Content Determined From Well Logs. SPE Form Eval. 1988;3(2):407–19.
7. Schmoker JW, Hester TC. Organic Carbon in Bakken Formation, United States Portion of Williston Basin. Am Assoc Pet Geol Bull. 1983;67(12):2165–74.
8. Mendelzon JD, Toksoz MN. Source Rock Characterization Using Multivariate Analysis Of Log Data. Source Rock Charact Using Multivar Anal Log Data. 1985 Jun 17;
9. Carpentier B, Huc AY, Bessereau G. Wireline logging and source rocks estimation of organic carbon by the Carbolog method. Log A. 1991;32(3):279–97.
10. Yun H, Xiang J, Liu Z. Estimation method of organic carbon log and its application in ShengLi Oilfield. Cejing Jishu/Well Logging Technol. 2000 Jan 1;24(5):372-376+381.
11. Passey QR, Creaney S, Kulla JB, Moretti FJ, Stroud JD. A Practical Model for Organic Richness from Porosity and Resistivity Logs. Am Assoc Pet Geol Bull. 1990;74(12):1777–94.
12. Kamali MR, Mirshady AA. Total organic carbon content determined from well logs using ΔLogR and Neuro Fuzzy techniques. J Pet Sci Eng. 2004;45(3–4):141–8.
13. Pan Renfang WY and SZ. Geochemical Parameters for Shale Gas Exploration and Basic Methods for Well Logging Analysis [Internet]. Vol. 14, China Petroleum Exploration. 2009. p. 6–9. Available from: http://www.cped.cn
14. Passey QR, Bohacs KM, Esch WL, Klimentidis R, Sinha S. From Oil-Prone Source Rock to Gas-Producing Shale Reservoir – Geologic and Petrophysical Characterization of Unconventional Shale-Gas Reservoirs. In: The CPS/SPE International Oil & Gas Conference and Exhibition. 2010. p. 29.
15. Kadkhodaie-Ilkhchi A, Rahimpour-Bonab H, Rezaee M. A committee machine with intelligent systems for estimation of total organic carbon content from petrophysical data: An example from Kangan and Dalan reservoirs in South Pars Gas Field, Iran. Comput Geosci. 2009;35(3):459–74.
16. Khoshnoodkia M, Mohseni H, Rahmani O, Mohammadi A. TOC determination of Gadvan Formation in South Pars Gas field, using artificial intelligent systems and geochemical data. J Pet Sci Eng [Internet]. 2011;78(1):119–30. Available from: http://dx.doi.org/10.1016/j.petrol.2011.05.010
17. Herron MM, Grau JA, Herron SL, Kleinberg RL, Machlus M, Reeder SL, et al. Total organic carbon and formation evaluation with wireline logs in the green river oil shale. Proc - SPE Annu Tech Conf Exhib. 2011;5(October):3709–27.
18. Radtke RJ, Lorente M, Adolph B, Berheide M, Fricke S, Grau J, et al. A New Capture And Inelastic Spectroscopy Tool Takes Geochemical Logging To The Next Level. Petrophysics - SPWLA J Form Eval Reserv Descr. 2012 Jun 16;
19. Tariq Z, Mahmoud M, Abdulraheem A. Core log integration: a hybrid intelligent data-driven solution to improve elastic parameter prediction. Neural Comput Appl [Internet]. 2019;31(12):8561–81. Available from: https://doi.org/10.1007/s00521-019-04101-3
20. Wibowo RC, Mulyatno BS. Karakterisasi Reservoar Menggunakan Metode Inversi Impedansi Akustik dan Neural Network Pada Lapangan “ ICL ” Cekungan Sumatera Selatan. In: PROCEEDINGS PIT HAGI 2012. Palembang: HAGI; 2012. p. 2–4.
21. Wibowo RC, Sarkowi M, Mulyatno BS, Dewanto O, Zaenudin A, Amijaya DH, et al. Thinned coal distribution modeling based on integrated geological and geophysical data : Case study CBM resources in Central Palembang Sub-Basin. In: 2nd International Conference on Earth Science, Mineral, and Energy. Yogyakarta: AIP Publishing; 2020. p. 1–9.
22. Rahaman MS., Vasant P. Artificial Intelligence Approach for Predicting TOC From Well Logs in Shale Reservoirs: A Review. In: IGI Global. 2020. p. 46–77.
23. Hamada GM, Ahmed E, Chaw Y N. Artificial neural network (ANN) prediction of porosity and water saturation of shaly sandstone reservoirs. Adv Appl Sci Res [Internet]. 2018;9(2):26–31. Available from: www.pelagiaresearchlibrary.com
24. Waszkiewicz S, Krakowska-Madejska P, Puskarczyk E. Estimation of absolute permeability using artificial

neural networks (multilayer perceptrons) based on well logs and laboratory data from Silurian and Ordovician deposits in SE Poland. Acta Geophys [Internet]. 2019;67(6):1885–94. Available from: https://doi.org/10.1007/s11600-019-00347-6
25. Heslop KA. Generalized Method for the Estimation of TOC from GR and Rt. In: AAPG Annual Convention and Exhibition. New Orleans; 2010.
26. Liu Y, Chen Z, Hu K, Liu C. Quantifying Total Organic Carbon (TOC) from Well Logs Using Support Vector Regression. In: GeoConvention 2013: Integration [Internet]. 2013. p. 6. Available from: http://www.geoconvention.com/archives/2013/281_GC2013_Quantifying_Total_Organic_Carbon.pdf
27. Zhao T, Verma S, Devegowda D, Jayaram V. TOC estimation in the Barnett Shale from triple combo logs using support vector machine TOC estimation from triple combo logs using support vector machine. In: SEG Annual Meeting. New Orleans; 2015. p. 791–5.
28. Kişi Ö. Multi-layer perceptrons with Levenberg-Marquardt training algorithm for suspended sediment concentration prediction and estimation. Hydrol Sci J. 2004;49(6):1025–40.