http://journal.uir.ac.id/index.php/JGEET

E-ISSN : 2541-5794
P-ISSN : 2503-216X

Journal of Geoscience,
Engineering, Environment, and Technology
Vol 9 No 2 2024

RESEARCH ARTICLE

Multiple Linear Regression Method for Thermal Maturity Prediction Based On Well Logs

Rahmat Catur Wibowo[1,*], Muh Sarkowi[1], Ordas Dewanto[1], Bagus S Mulyatno[1], Ilham Dani[1]
[1] *Geophysics Engineering Universitas Lampung, Sumantri Bojonegoro Street No.1 Bandar Lampung City, Lampung, Indonesia.*

* Corresponding author : rahmat.caturwibowo@eng.unila.ac.id
Tel.:+62-812-7170-2441;
Received: Aug 15, 2022; Accepted: May 22, 2024.
DOI: 10.25299/jgeet.2024.9.2.10270

Abstract

Accurate estimation of thermal maturity is essential in characterizing a source rock, especially using vitrinite reflectance (Ro). The limitations of laboratory data related to the high cost of analysis require a special reliable method to measure the Ro value indirectly in the source rock layer. The proposed method is a continuous prediction of the value of Ro from well logs data using the Multiple Linear Regression (MLR) technique in the Palembang Sub-Basin, South Sumatra Basin. A total of 25 Ro data from 2 wells (RCW-01 and RCW-02) are available from the laboratory's core data analysis results. The Ro data varies from 0.39% to 0.76%, with an average of 0.54%. Prediction of the value of Ro is carried out using the MLR method, which is then carried out training and validation for continuous Ro. The training was carried out using one well (RCW-01) at 2287-3027 m and testing at other intervals (1848-2286 m). The results of the training show an estimation accuracy of R^2 0.99, while the test results produce R^2 0.81. The MLR formula in the RCW-01 well was then applied to the RCW-02 well for the validation test phase. The well RCW-02 produces a good correlation estimate equal to R^2 0.85. Prediction of the value of Ro using the MLR method can be used to evaluate the source rock layer of a sedimentary basin in the form of a continuous interval.

Keywords: Multiple linear regression, well logs, vitrinite reflectance, prediction, thermal maturity

1. Introduction

Thermal maturity is an essential geochemical parameter in evaluating the source rock of a sedimentary basin. Various molecular parameters are widely used in estimating the thermal maturity of sedimentary rock (Abarghani et al., 2019).

Vitrinite reflectance (Ro) has been widely used for determining the thermal maturity of source rock and estimating the amount of an area uplifted or eroded, paleo heat flow, and thermal modeling (Peters and Cassa, 1994; Wibowo et al., 2023). Problems in applying the Ro technique in oil and gas exploration have been detected for a long time, such as identification of differences or variations in vitrinite, especially in dispersed organic material (DOM), suppression of hydrogen-rich kerogen, contamination caused by caving, and reworked organic material. Besides, the Ro technique is unique invaluating thermal maturity in sedimentary basins, especially its simplicity and applicability (Kadkhodaie and Rezaee, 2017). Therefore, to produce an accurate and measurable evaluation, an extensive collection of sample data is needed in the analysis stage (Tariq et al., 2020).

Given the importance of Ro in determining the thermal maturity evaluation stage and the high cost of laboratory sample data (core) analyses, a continuous Ro estimation method is needed based on well-log data (Labani and Rezaee, 2012). Lang (1994) explains a relationship between the value of Ro and depth. The higher the depth value, the higher the temperature or Ro value. Then it is associated with a decrease in the transit time interval (log DT) value in the well-log data. This relationship was then refined by Mallick and Raju (1995) in the Upper Asam basin, India, which resulted in a mathematical relationship between Ro and log DT. However, this study has limitations in the form of only using one data log (log DT) in evaluating thermal maturity, while according to Hussein and Abdula (2018) the resistivity value also can have a relationship with thermal maturity. Especially in the source rock that has reached the oil window, the resistivity value will be higher than the zone still in the diagenesis stage (Afifah and Setiawan, 2019).

This study aims to find an appropriate and simple method for predicting the Ro value in the Palembang Sub-Basin, South Sumatra Basin, using the Multiple Linear Regression (MLR) method based on well logs data (Zhao et al., 2019). The well-logs data used as input are density logs, DT, Gamma-Ray (GR), neutron porosity (NPHI), and deep resistivity (RILD). All these logs are used because they are directly or indirectly related to the evaluation of thermal maturity.

2. Source rock characteristics

Talang Akar shale is in the form of coaly shale, which was deposited in the Late Eocene – Middle Miocene. This formation was deposited in a fluvial-deltaic environment composed of kerogen types I and II (Argakoesoemah and Kamal, 2004). The shale has amorphous and vitrinitic kerogen, which can generate oil and gas. TOC in this formation has good potential, which is around 1.5 – 8% in the Limau area, in the Kuang area, it is potentially less good, with values ranging from 0.33 – 0.9%. This shale is categorized as mature in the Limau area, and late mature in the Kuang and Muaraenim-Lematang areas with T-max values of 436 – 450 ºC and Ro 0.45 – 0.94%. The Lembak and Kuang areas were early-matured with 425 – 433 ºC T-max values and 0.3 – 0.4 % (Table 1) (Wibowo, 2013).

Table 1. Characteristics of the source rock in the South Sumatra Basin, especially in the Talang Akar shale (Wibowo, 2013).

South Sumatra Back Arc basin	Source Rock Analysis							References
	TOC (%)	Ro (%)	Tmax (ºC)	Thickness (ft)	Kerogen Type	Dep. Env.	Maturity Index	
Benakat Shale (South Palembang Sub-Basin)	1.7 – 8.5	-	436 - 441	660 - 2508	II & III	Terrestrial	Peak - Postmature	Sarjono & Sarjito (1989)
Benakat Shale (South Palembang)	0.5 – 16	0.64 – 1.4	435 - 455	-	II & III	Fluvio-deltaic	-	Suseno et al. (1992)
Benakat Shale (Lahat Fm.)	1.0 – 3.0	-	-	-	-	Shallow lacustrine	-	Ginger & Fielding (2005)
Talang Akar Shale (South Palembang Sub-Basin)	0.33 – 8	0.3 – 0.94	436 - 450	990 - 1815	I & II	Fluival, deltaic & shallow marine	Immature – late mature	Sarjono & Sarjito (1989)
Talang Akar Shale	5	-	-	-	-	-	-	Ginger & Fielding (2005)
Talang Akar Shale	5	0.35 – 1.11	-	-	-	Deep marine	-	Argakoesoemah & Kamal (2004)
Gumai Shale (South Palembang Sub-Basin)	0.5 – 11	0.5 – 0.7	400 – 440	495 - 4950	III	Shallow to deep marine	Immature – peak mature	Sarjono & Sarjito (1989)
Gumai Shale	8	-	-	-	III	Marine	Immature	Ginger & Fielding (2005)

3. Theory

3.1 Multiple linear regression (MLR)

Linear regression is the most widely used technique in establishing the relationship between the dependent (output/target) and independent (input) variables. It uses a linear approach to model the correlation function between input and output variables (Emelyanova et al., 2016). In the case of one input variable is called single linear regression, while for multiple input variables, it is called multiple linear regression. The expression between input (xi) and output (yi), assuming a linear relationship, can be written as:

$$y_i = \theta_0 + x_1\theta_1 + x_2\theta_2 + \cdots + x_n\theta_n + \varepsilon_i, i = 1,2, \ldots n \quad (1)$$

In matrix notation, Equation (1) can be expressed as:

$$Y = X\theta^T + \varepsilon \quad (2)$$

where Y is the vector of the observed values y_i of the variable known as the output or dependent variable, X is the vector form of the input or independent variable x_i, θ is the vector of dimension parameters (n + 1)—the elements known as the regression coefficients (for example, θ_0 is the intercept) and is the vector form of the error term ε_i.

The goal is to minimize the error of the cost function (the most widely used function is the mean squared error function) which is defined as:

$$J(\theta_0, \theta_1, \ldots, \theta_n = \frac{1}{2n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \quad (3)$$

where $J(\theta_0, \theta_1, \ldots, \theta_n)$ is the cost-function; $\hat{y}_i$ and y_i represent the predicted and actual output (Mandal et al., 2022).

4. Method

This study uses the MLR method to optimize the parameters related to the Ro value with precision to predict the value using well logs data such as: Density, DT, GR, NPHI RILD. The advantage of this method is the accuracy and simplicity of computation in predicting a value based on many parameters (Jaber and Shuker, 2014).

Twenty-five core data samples from two wells (RCW-01 and RCW-02) were used to calibrate the predicted MLR results. The RCW-01 well is used as a well for the application of MLR because it has complete geochemical data and well logs. The prediction results were then retested at two different intervals to see how accurate the predictions were. After producing a good estimation accuracy value, the MLR equation obtained in RCW-01 was tested for validation on the RCW-02 well. The prediction results' accuracy is determined by the correlation coefficient (CC) value, and the root means square error (RMSE).

5. Result and discussion

To perform the MLR technique, the actual measurement of Ro and the collection of well log data were carried out. Based on Figure 1, the resistivity value has a very important relationship with the Ro value. Based on these results, RILD is used as the MLR input. Ro has a strong relationship with DT, where the resulting CC is -0.69. Density and NPHI have intermediate bonds, which have CC values of -0.38 and -0.22. While GR has a low relationship because of mineral content (Wibowo et al., 2020a), CC of 0.11.

The initial stage in performing the MLR technique is to perform quality control on all data that will be used as input. Figure 2 is all well logs data (RILD, DT, Density, NPHI, and GR) which are used as inputs (independent values) in the MLR technique, and core sample data Ro as the dependent value. Based on the results of MLR processing, the formula is obtained:

$$Ro_{MLR} = 10\text{^}(-1.08823165 - 0.32997306 * Log(NPHI) + 0.23583975 * Log(DT) + 0.23392235 * Log(GR) - 0.15643133 * Log(RILD) - 0.30933119 * Log(DEN)) \quad (4)$$

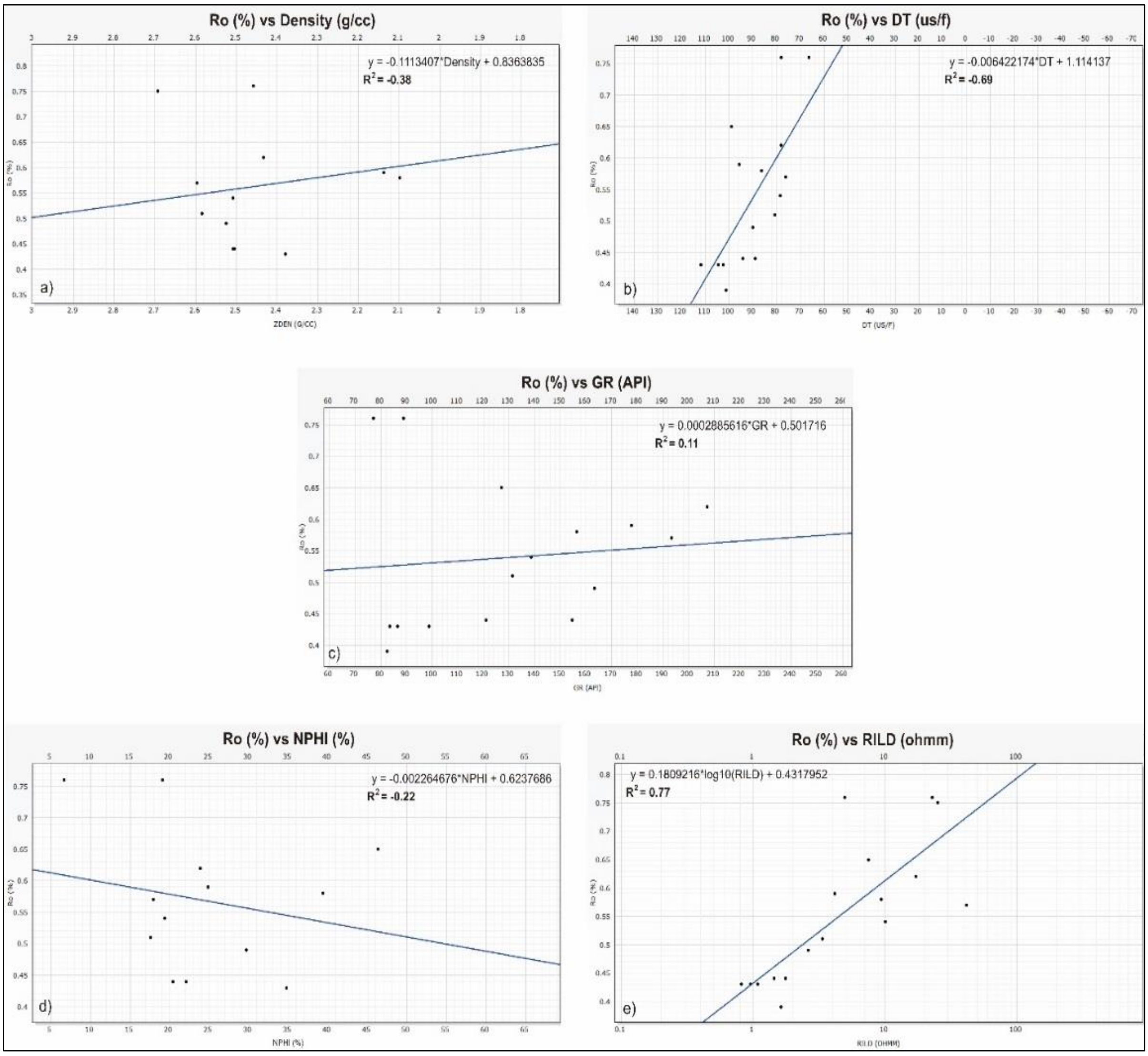

Fig. 1. Correlation coefficient (CC) between Ro from the core and well logs data (Density, DT, GR, NPHI, and RILD).

Figure 3 shows the MLR model's results with a high accuracy level in predicting Ro over the entire RCW-01 well-layer interval, with an R^2 value of 0.95 (Figure 4). To test the MLR method in making predictions, the RCW-01 well was divided into two layers, especially in the Talang Akar Formation (TAF). Layers at the 2287-3027 m interval were used as the MLR input (Figure 5), and the 1848-2286 m interval was used as the blind data (Figure 7). The formula result for this training is as:

$$Ro_{MLR} = 10^\wedge(-1.08823165 - 0.32997306 * Log(CNCF) + 0.23583975 * Log(DT) + 0.23392235 * Log(GR) - 0.15643133 * Log(RILD) - 0.30933119 * Log(ZDEN)) \quad (5)$$

Figure 6 confirms the height of accuracy obtained from the MLR method (eq. 5), and the validation results obtained in Figure 7 produce R^2 0.81 (Figure 8). Based on these results, it can be concluded that the MLR method can be used to predict Ro as a function of several well logs, RILD, DT, Density, NPHI, and GR data.

The MLR formula (eq. 4) was used to validate the TAF well RCW-02. Figure 9 shows the results with high accuracy in predicting Ro. CC value shows R^2 0.85 between prediction Ro and core data Ro. The MLR correlation method produces an accurate model between the predicted and actual Ro values. RMSE value in well RCW-01 is 0.034 with R^2 0.95. Furthermore, when the validation test was carried out on the RCW-02 well, it resulted in an RMSE of 0.044 with an R^2 of 0.85. The decrease in correlation value is due to the distance of the RCW-02 well from the RCW-01 well, which is about 15 km. However, based on these results, the MLR method is confirmed to be used to predict the value of Ro.

In this study, 25 core samples, from two wells were used. Ro has been measured for all samples. The parameters that have a significant influence on Ro were determined, and all samples were used to develop the multiple regression. Variables used for the ultimate equation were RILD, DT, Density, NPHI, and GR (equation 4). The RILD and DT values are important in describing thermal maturity, and both parameters are the main sources of information to estimate rock physics properties (Eskandari et al., 2004). The multiple regression method gave good results during the validation phase in the RCW-1 well, but when it is applied to another well (RCW-2), it usually faces problems. Such problems can be avoided with more parameter input and intelligent solution techniques such as a neural network (Wibowo et al., 2020b; Wibowo et al., 2022). Artificial neural networks are adaptive and parallel information processing systems that have the ability to develop functional relationships between data and provide a powerful toolbox for nonlinear interpolations (Waszkiewicz et al., 2019; Tariq et al., 2020; Adhari and Kardawi, 2022).

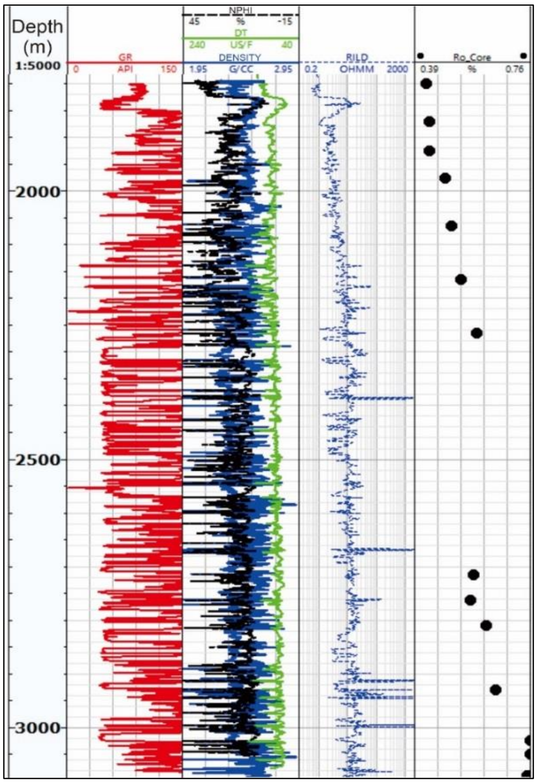

Fig. 2. Well logs data for Talang Akar shale Formation.

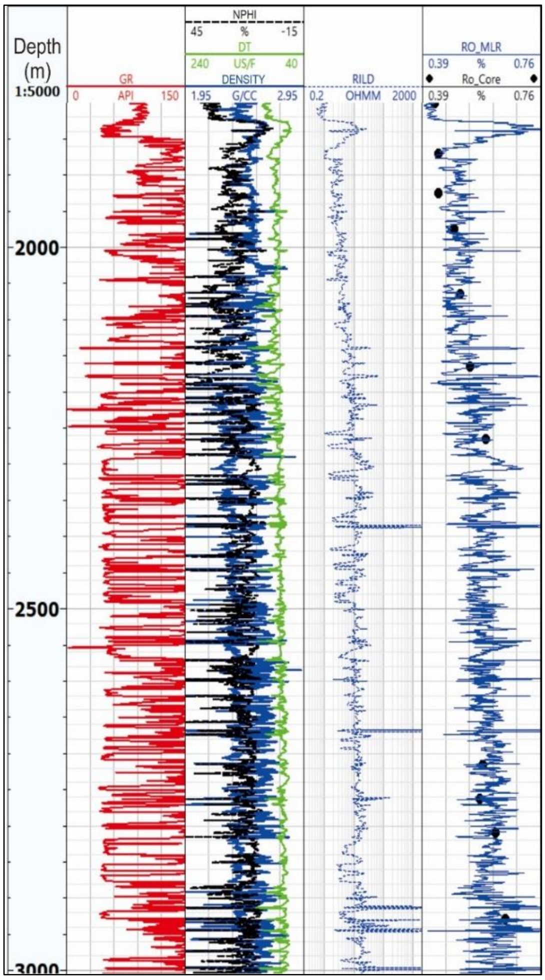

Fig. 3. Well logs data for Talang Akar shale Formation (RCW-01) and prediction result of Ro using MLR method.

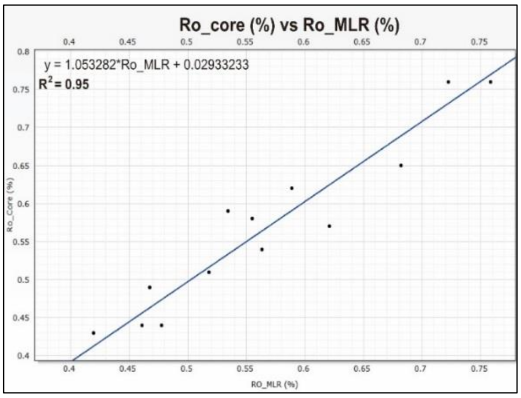

Fig. 4. Coefficient correlation for Ro prediction using MLR model (RCW-01 well).

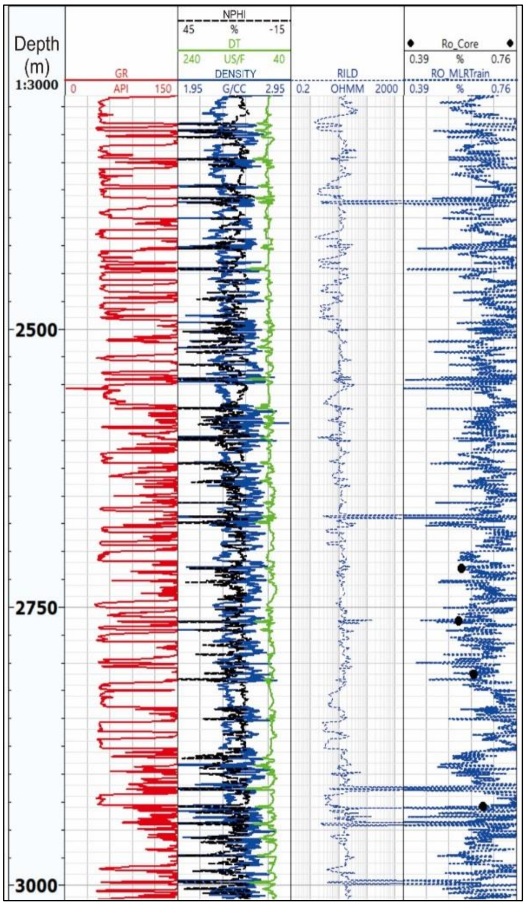

Fig. 5. Well logs data for Talang Akar shale Formation (interval 2287-3027 m) and prediction result of Ro using MLR method.

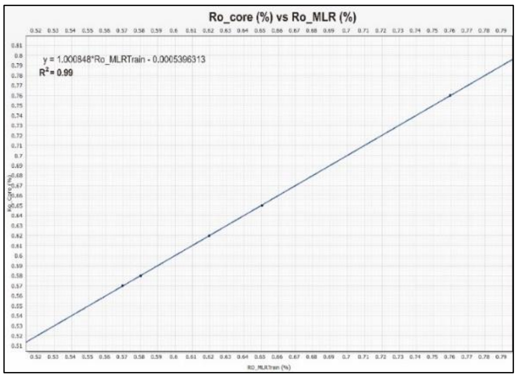

Fig. 6. Coefficient correlation for Ro prediction using MLR model (interval 2287-3027 m).

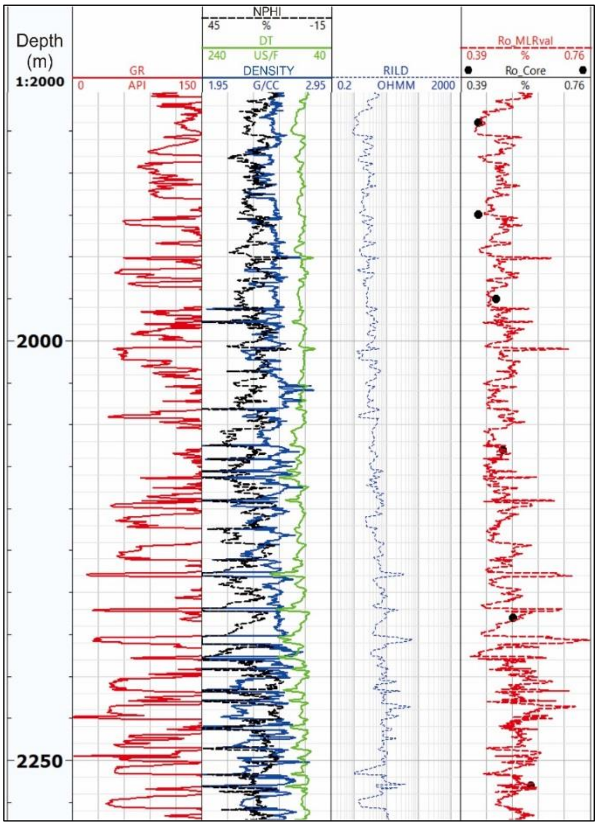

Fig. 7. Well logs data for Talang Akar shale Formation (interval 1848-2286 m) and prediction result of Ro using MLR method.

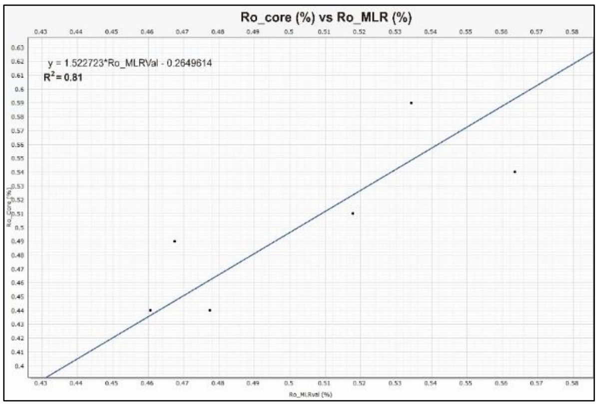

Fig. 8. Coefficient correlation for Ro prediction using MLR model (interval 1848-2286 m).

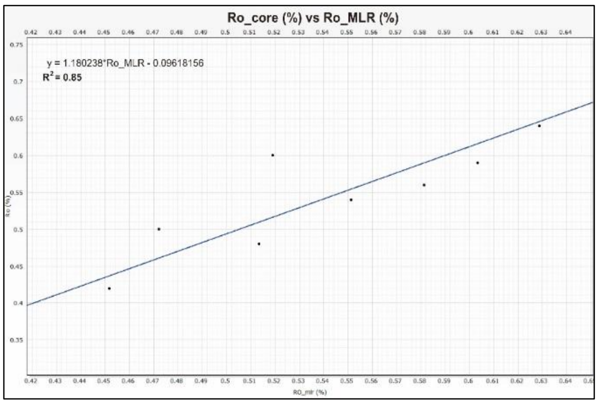

Fig. 9. Actual and predicted Ro for RCW-02 well using MLR method.

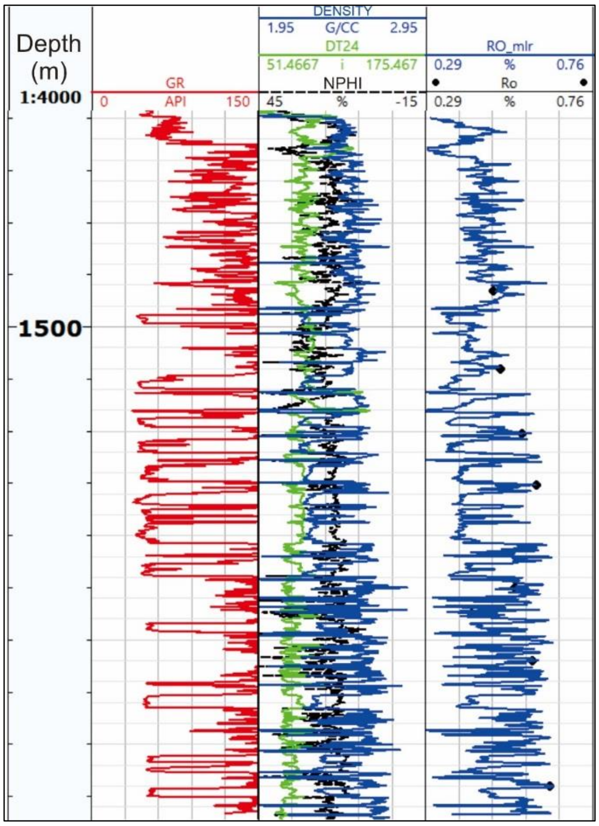

Fig. 10. Well logs data for Talang Akar shale Formation (RCW-02) and prediction result of Ro using MLR method (using eq. 4).

6. Conclusion

The approach using the MLR method has resulted in an accurate value in predicting the value of Ro. The MLR method uses the input of several well logs data such as Density, DT, GR, NPHI, and RILD. The results of modeling using the MLR method yielded CC values of R2 0.95 and RMSE 0.034. While the results of the validation test using other wells (RCW-02) in the formation and lithology which are assumed to be the same, produce a CC value of R2 0.85 with an RMSE of 0.044. It is estimated that the obtained value can still be increased using a non-linear relationship between the input and output parameters, such as the Artificial Neural Network method and other machine learning methods.

Acknowledgments

The authors would like to thank all parties who have helped conduct this research, especially the Faculty of Engineering at the University of Lampung, which has funded it.

References

Abarghani, A., Ostadhassan, M., Bubach, B., Zhao, P., 2019. Estimation of thermal maturity in the Bakken source rock from a combination of well logs, North Dakota, USA. Marine and Petroleum Geology 105, 32–44. https://doi.org/10.1016/j.marpetgeo.2019.04.005

Adhari, M.R., Kardawi, M.Y., 2022. Estimation of density log and sonic log using artifiicial intelligence: an example from the Perth Basin, Australia. Journal of Geoscience, Environment, and Technology 7, 158–166. https://doi.org/ 10.25299/jgeet.2022.7.4.10050

Afifah, P.D., Setiawan, B., 2019. Middle Balck Shale of Airbenakat Formation in Berau Areas, Jambi: are they potential as as source rock. Journal of Geoscience,

Environment, and Technology 4, 128–134. https://doi.org/ 10.25299/jgeet.2019.4.2.1774
Argakoesoemah, R.M.I., Kamal, A., 2004. Ancient Talang Akar deepwater sediments in South Sumatra Basin: a new exploration play, in: Proceedings, Deepwater And Frontier Exploration In Asia & Australasia Symposium. IPA-AAPG, Jakarta, p. 18. https://doi.org/10.29118/ipa.1730.251.267
Emelyanova, I., Pervukhina, M., Clennell, M. ben, Dewhurst, D.N., 2016. Applications of standard and advanced statistical methods to TOC estimation in the McArthur and Georgina basins, Australia. The Leading Edge 35, 51–57. https://doi.org/10.1190/tle35010051.1
Eskandari, H., Rezaee, M.R., Mohammadnia, M., 2004. Applications of Multiple Regression and Artificial Neural Netwrok Techniques to Predict Shear Wave Velocity from Wireline Log Data for a Carbonate Reservoir, South-West Iran. Recorder 29, 19.
Hussein, H.S., Abdula, R.A., 2018. Multiple linear regression approach for the vitrinite reflectance estimation from well logs: A case study in Sargelu and Naokelekan Formations – Shaikhan-2 Well, Shaikhan oil field, Iraq. Egyptian Journal of Petroleum 27, 1095–1102. https://doi.org/10.1016/j.ejpe.2018.03.013
Jaber, A.K., Shuker, M.T., 2014. Multiple Linear Regression Approach for the Permeability Calculation from Well Logs : A Case Study in Nahr Umr Formation - Subba Oil Field, Iraq. International Journal of Science and Research 3, 1408–1415.
Kadkhodaie, A., Rezaee, R., 2017. Estimation of vitrinite reflectance from well log data. Journal of Petroleum Science and Engineering 148, 94–102. https://doi.org/10.1016/j.petrol.2016.10.015
Labani, M.M., Rezaee, R., 2012. Thermal maturity estimation of gas shale layers from conventional well log data: A case study from kockatea shale and carynginia formation of Perth Basin, Australia. Society of Petroleum Engineers - SPE Asia Pacific Oil and Gas Conference and Exhibition 2012, APOGCE 2012 2, 1040–1053. https://doi.org/10.2118/158864-ms
Lang, W.H., 1994. The Determination Of Thermal Maturity In Potential Source Rocks Using Interval Transit Time/interval Velocity. The Log Analyst 35.
Mallick, R.K., Raju, S. v., 1995. Thermal maturity evaluation by sonic log and seismic velocity analysis in parts of Upper Assam Basin, India. Organic Geochemistry 23, 871–879. https://doi.org/10.1016/0146-6380(95)00088-7
Mandal, P.P., Rezaee, R., Emelyanova, I., 2022. Ensemble learning for predicting TOC from well- logs of the unconventional goldwyer shale. Energies (Basel) 15. https://doi.org/10.3390/en15010216
Peters, K.E., Cassa, M.R., 1994. "Applied Source Rock Geochemistry", Chapter 5, in: AAPG Memoir 60. pp. 93–120.
Tariq, Z., Mahmoud, M., Abouelresh, M., Abdulraheem, A., 2020. Data-Driven Approaches to Predict Thermal Maturity Indices of Organic Matter Using Artificial Neural Networks. ACS Omega 5, 26169–26181. https://doi.org/10.1021/acsomega.0c03751
Waszkiewicz, S., Krakowska-Madejska, P., Puskarczyk E., 2019. 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, 67(6):1885–94. Available from: https://doi.org/10.1007/s11600-019-00347-6
Wibowo, R.C., 2013. "Unconventional Reservoir" Shale Gas Potential Based On Source Rock Analysis In Sumatran Back Arc Basin, in: Setiawan, N.I., Budianta, W., Idrus, A. (Eds.), Proceedings of International Conference on Geological Engineering Department, Engineering Faculty, Gadjah Mada University. Geological Engineering Department, Engineering Faculty, Gadjah Mada University, Yogyakarta, pp. 151–163.
Wibowo, R.C., Pertiwi, A.P., Kurniati, S., 2020a. Identification of Clay Mineral Content Using Spectral Gamma Ray on Y1 Well in Karawang Area, West Java, Indonesia. Journal of Geoscience, Environment, and Technology 5, 136–142. https://doi.org/10.25299/jgeet.2020.5.3.4504
Wibowo, R.C., Sarkowi, M., Mulyatno, B.S., Dewanto, O., Zaenudin, A., Amijaya, D.H., 2020b. 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, 1–9
Wibowo, R.C., Dewanto, O., Sarkowi, M., 2022. Total Organic Carbon (TOC) Prediction Using Machine Learning Methods Based on Well Logs Data. In: 2nd International Conference on Science, Technology, and Environment. Bandarlampung: AIP Publishing.
Wibowo, R.C., Putri, A.C.E., and Dewanto, O., 2023. "Analysis of Unconventional Oil and Gas Reservoirs using Well Logging, Geochemical and Seismic Data", JURNAL GEOCELEBES, 7(2), pp. 154–167. doi: 10.20956/geocelebes.v7i2.20603.
Zhao, P., Ostadhassan, M., Shen, B., Liu, W., Abarghani, A., Liu, K., Luo, M., Cai, J., 2019. Estimating thermal maturity of organic-rich shale from well logs: Case studies of two shale plays. Fuel 235, 1195–1206. https://doi.org/10.1016/j.fuel.2018.08.037