

Prediction of the Rate of Penetration (ROP): Comparison Between Nineteen Machine Learning Algorithms

Husam H. Alkinani, Abo Taleb T. Al-Hameedi, and Shari Dunn-Norman, Missouri University of Science and Technology

Copyright 2020, AADE

This paper was prepared for presentation at the 2020 AADE Fluids Technical Conference and Exhibition held at the Marriott Marquis, Houston, Texas, April 14-15, 2020. This conference is sponsored by the American Association of Drilling Engineers. The information presented in this paper does not reflect any position, claim or endorsement made or implied by the American Association of Drilling Engineers, their officers or members. Questions concerning the content of this paper should be directed to the individual(s) listed as author(s) of this work.

Abstract

The rate of penetration (ROP) is one of the most important factors in drilling operations. Because of the complicated relationship between parameters affecting ROP, it is not easy to be estimated using analytical methods. In this work, nineteen machine learning algorithms were used to create models that can predict ROP based on weight on bit (WOB), flow rate (Q), revolutions per minute (RPM), and the unconfined compressive strength (UCS), utilizing data from over 3000 wells drilled worldwide. 5-fold cross-validation was implemented to ensure sufficient representation for all data points in the training process. The findings showed that the exponential Gaussian process regression (GPR) model had the highest R^2 (0.91) and the lowest root mean square error (RMSE) (0.96). Thus, the exponential GPR model was selected as the best algorithm to train the model. The exponential GPR model was also compared with previously developed models and the pros and cons of all previous models and the model developed in this study were discussed. The exponential GPR model can be used to optimize ROP in an effective and easy way by altering WOB, ROP, and/or Q. Due to the large data availability in the oil and gas industry, machine learning and other artificial intelligence methods can play a vital role in revolutionizing the industry by cutting costs and decreasing the non-productive time.

Introduction

The rate of penetration (ROP) can be described as “the volume of rock removed per unit area (feet) per unit time (hour). It can be referred to as the speed of breaking the rock under the bit. In general, it measures the speed or the progress of the bit when it drills the formation” (Bourgoyne and Millheim, 1986). The cost for penetrating the formation is 20-30% of the total cost required to drill a well, according to recent wells drilled in Iraq (Basra Oil Company, 2017). Therefore, many service companies and operators are focused on understanding and optimizing ROP to drill wells faster. However, there is a tradeoff between having too high or low ROP. High ROP can cause many problems related to wellbore instability, poor hole cleaning, lost circulation, etc.; while low ROP increases the delivery time of the well (Akgun, 2002). Thus, there should be a balance to maximize ROP and minimize the associated problems in order to minimize the well delivery time.

To understand how to maximize ROP, it is necessary to comprehend the parameters influencing it. These parameters consist of operational, geological, and fluid parameters (Ashrafi et al., 2019). There are two types of factors that should be considered when analyzing ROP; controllable and uncontrollable factors. Controllable factors such as drilling parameters (e.g. mud weight (MW), revolutions per minute (RPM), weight on bit (WOB), flow rate (Q), etc.). On the other hand, some examples of uncontrollable factors including geological and environmental factors that cannot be altered such as formation strength, depth, state of stress, etc. (Sikes, 1936; Hossain and Al-Majed, 2015; Osgouei, 2007). The challenge with predicting and optimizing ROP is that there is not a clear connection between the aforementioned factors affecting ROP. Thus, it is difficult to create a model that comprehends all these factors (Darbor et al., 2017; Akgun, 2007). Nonetheless, stochastic and non-linear relationships can be used to include the effect of these parameters (Moradi et al., 2010). Researchers have explored physics-based models to model the effect of these parameters (Bingham, 1965; Motahhari et al., 2010; Warren, 1987; Hareland and Rampersad, 1994; Winters et al., 1987). However, the necessity of more data in these models and the reliability on the lithology have contributed to minimizing the efficiency of these physics-based models (Mendes et al., 2007; Hedge et al., 2017). Therefore, there is a need for models with less possibility of error in predicting ROP. Artificial intelligence has shown promising findings in prediction and classification. That is why, researchers have explored the use of artificial intelligence as a reliable tool to predict and optimize ROP (Yilmaz and Kaynar, 2011).

Classic regression models have been used to forecast ROP for a long time (Moraveji and Naderi, 2016; Yagiz et al., 2009). Bourgoyne and Young (1974) used multiple regression method to predict ROP. They realized that ROP is affected by eight parameters; depth, formation compaction and strength, rotation speed, bit diameter and weight, hydraulics, and differential pressure. This model was created around the drillability of the formation and it depends on bit design, geology, and drilling parameters (Soares and Gray, 2019). Due to the complex non-linear relationship between the parameters affecting ROP, using this model can be challenging to represent the actual ROP (Nascimento et al., 2015; Ricardo et al., 2007). The connection

between rock strength and drilling parameters have been identified by Mohammad et al. (2014) using laboratory experiments. Despite the usefulness of this model, most variables have to be estimated from experimental work, making it difficult to be applied in the field. Moraveji and Naderi (2016) utilized the response surface method to estimate ROP. Ultimately, many regression models fall short in predicting ROP due to the non-linearity, data noise, and/or the scarcity of the data (Gan et al., 2019).

Shi et al. (2016) used extreme learning machine (ELM) to forecast ROP with a quick leaning rate and produced decent results. Adoko and Gokceoglu (2017) used Bayesian estimation to estimate ROP within the rock mass. Another application presented by Basarir et al. (2014), they used adaptive neuro-fuzzy estimation methods to forecast ROP. Most results showed that data-driven models surpassed physical-based models in the prediction of ROP (Hedge et al., 2017). Ahmed et al. (2019) examined the potential of ELM, least-square support vector regression (LSSVR), artificial neural networks (ANN), and support vector regression (SVR) in the prediction of ROP. In addition, a hybrid ANN model was implemented by Ashrafi et al. (2019) and showed that hybrid ANN surpassed the conventional ANN models. Yagiz and Krahan (2015) used three optimization methods to predict ROP and compared the findings of the three methods. Moreover, using particle swarm optimization (PSO) for drilling optimization was introduced by Hedge and Gray (2018). They deduced a great improvement in drilling efficiency. Hybrid support vector regression was introduced by Gan et al. (2019) to predict ROP and they compared their model with eight ROP models in the literature. The results were promising after applying the model in Central China.

The aim of this work is to develop a machine learning model to predict ROP with the minimum number of parameters. The goal is to have an easy model that can be used in the field to optimize the drilling operation and save time and money. To achieve that goal, data from more than 3000 wells drilled worldwide were used to train nineteen machine learning models and the model with the highest accuracy was selected.

Data and Methods

In this section, the process of data collection, data processing, machine learning models' developments will be explained.

Data

Data collection is a very time-sensitive process. Data of key drilling parameters were collected from daily drilling reports, final wells' reports, the literature, etc. The data covered many areas around the world, including but not limited to, the middle east, North and South America, Australia, China, UK, Norway, Indonesia, and Russia. The data went through processing steps where all outliers were removed from all key drilling parameters. The parameters that were selected to predict ROP in m/hr are WOB in Ton, Q in L/Min, and RPM; while the

formation strength was implemented using the unconfined compressive strength (UCS) in psi. Eq. (1) was used to calculate UCS (Chang et al., 2006).

$$UCS = 1.35 \times \left(\frac{304.8}{\Delta t_c} \right)^{2.6} \quad \text{Eq. 1}$$

Where UCS is in MPa, and Δt_c is the compressional wave travel time in $\mu\text{sec}/\text{ft}$. The goal of this study is to create a model that can be used easily to predict ROP with a minimum number of inputs. Having three operation parameters (WOB, Q, and RPM) and one formation strength parameter (UCS) will make it easy to optimize ROP. **Figure 1 (Appendix A)** shows the distribution of the data; while **Table 1 (Appendix A)** shows the summary of statistics of the data

Developments of the Machine Learning Models

Nineteen machine learning algorithms were used in this study; four linear regression models, three tree models, six support vector machine (SVM) models, two ensemble models, and four Gaussian process regression (GPR) models. **Table 2** shows a summary of the machine learning algorithms used in this study. The details of each algorithm are beyond the scope of this paper and can be found in MATLAB (2019).

To assess the accuracy of each model, two factors were used. The first one is R^2 , which is a measure of the goodness of the fitted model. The second one is related to the error, which is called the root mean square error (RMSE). RMSE is the square root of the mean squared error (MSE). The model with the lowest RMSE and the highest R^2 was selected. Eq. (2) and Eq. (3) were used to calculate R^2 and RMSE, respectively.

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad \text{Eq. 2}$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad \text{Eq. 3}$$

Where N is the number of data, \hat{y}_i is the predicted data, \bar{y} is the mean of the real data, and y_i is the real data point.

Cross-Validation

Cross-validation is a vital step in the process of training any machine learning model. It assures that all data points are implemented in the training process. Moreover, it ensures the model's generalization for new data and avoids overfitting the model. There are many ways to do cross-validation. In this work, 5-fold cross-validation was used to confirm the validity of the created models. The idea of 5-fold cross-validation is that the data will be randomly divided into 5-folds (5 equal-sized sets). Next, four folds will be used for training and the other one will be used for testing. This process is conducted five times, each time a different fold is used for testing and the other four used for training. After conducting five iterations, the average of these iterations was taken and used to assess the models. This

will ensure all of the representative data points will be taken into account in the process of creating the models. On the other hand, dividing the data into training and testing sets may lead to missing some important data in the training set (Alpaydin, 2014).

Results and Discussion

The main goal of this paper is to have a model that can accurately predict ROP with the minimum number of parameters. All nineteen models were trained and the RMSE and R^2 were calculated for each model. **Figure 2** shows the results of R^2 and RMSE for all models. As shown in **Figure 2**, all linear regression models did not do very well and produced relatively high RMSE and low R^2 when compared to the other models. These results should be expected due to the non-linear and complex relationship between parameters influencing ROP. On the other hand, tree and GPR models did the best among the other models, with the exponential GPR being the best in terms of having the lowest RMSE (0.96) and having the highest R^2 (0.91). Thus, the exponential GPR was selected to train the network and all the other models were ignored.

Figure 3 shows the predicted and true ROP with the row number of the data used in this study. **Figure 3** clearly illustrates that ROP predictions are consistent with the different range of data used in this work. **Figure 4** shows the predicted and true ROP. The red line is where a perfect prediction is expected, and most of the data are distributed around the red line, with some being a little off the red line. However, the model is generally doing quite good at predicting ROP.

The created exponential GPR model can be used to optimize ROP in an easy and efficient way. Only three operational parameters (WOB, Q, and RPM) are required along with the UCS of the formation. While UCS is uncontrollable, WOB, Q, and RPM are very easy to control and alter to optimize ROP and decrease the well delivery time.

Comparison with Previously Developed Models

Models to predict ROP were previously developed, in this section, a discussion of the new exponential GPR model developed in this study and the models developed in previous studies will be executed. Starting with the models developed for the Rumaila Field in South Iraq to predict ROP using linear regression (Al-Hameedi et al., 2017a; Al-Hameedi et al., 2017b). These models used data from the Rumaila field and developed models to predict ROP using RPM, WOB, and Q. These models are easy to use in the specific area where the data collected. However, they will not be accurate if they were used in different areas. On the other hand, Alkinani et al. (2019) developed a recurrent neural network model to predict ROP using the following eight parameters:

1. Q in L/Min
2. Mud weight (MW) in gm/cc
3. Nozzles total flow area (TFA) in inch²
4. Plastic viscosity (PV) in cp
5. RPM

6. UCS in psi
7. WOB in Ton
8. Yield point (Yp) in lb/100ft²

While this model has proven a good potential to predict ROP, it is not easy to apply due to the need for operational and fluid data. However, the exponential GPR model developed in this study is easy to implement with only three operational parameters, which can be altered easily to optimize ROP and save time and money. The drilling personnel can simply use this model to predict and optimize ROP by altering Q, WOB, and/or RPM.

Conclusion

Data from over 3000 wells were collected from many sources worldwide. Nineteen machine learning algorithms were utilized to train models to predict the ROP from WOB, Q, RPM, and UCS. Cross-validation was carried out to ensure adequate representation of all data points in the training process. The exponential GPR model was selected among the other machine learning algorithms since it resulted in the highest R^2 and the lowest RMSE. The exponential GPR model was compared to previously developed models, and the pros and cons were presented. The exponential GPR model can be utilized to estimate ROP in an effective and easy way utilizing WOB, Q, RPM, and UCS. In the same vein, the exponential GPR model can be used in reverse to adjust WOB, Q, and RPM to optimize and achieve the required ROP.

To sum it up, machine learning and other artificial intelligence methods have proven their applicability to solve complicated problems that cannot be described in analytical and physics-based models. Because of the large data availability in the oil and gas industry, artificial intelligence can play an important role in revolutionizing the industry by cutting costs and decreasing the non-productive time.

Acknowledgment

The authors would like to thank Basra Oil Company from Iraq for providing us with various real field data.

References

1. Adoko, A.C., Gokceoglu, C., 2017. Bayesian prediction of TBM penetration rate in rock mass. *Eng. Geol.* 226, 245–256.
2. Ahmed, O.S., Adeniran, A.A., Samsuri, A., 2019. Computational intelligence based prediction of drilling rate of penetration: a comparative study. *J. Petrol. Sci. Eng.* 172, 1–12.
3. Akgun, F. (2002, January 1). How to Estimate the Maximum Achievable Drilling Rate without Jeopardizing Safety. *Society of Petroleum Engineers*. doi:10.2118/78567-MS.
4. Akgun, F., 2007. Drilling rate at the technical limit. *Int. J. Petrol. Sci. Technol.* 1, 99–118.
5. Al-Hameedi, A. T., Dunn-Norman, S., Alkinani, H. H., Flori, R. E., & Hilgedick, S. A. (2017a, August 28). Limiting Drilling Parameters to Control Mud Losses in the Dammam

- Formation, South Rumaila Field, Iraq. American Rock Mechanics Association.
6. Al-Hameedi, A.T., Dunn-Norman, S., Alkinani, H.H., Flori, R.E., and Hilgedick, S.A. 2017b. Limiting Drilling Parameters to Control Mud Losses in the Shuaiba Formation, South Rumaila Field, Iraq. Paper AADE-17- NTCE- 45, 2017 AADE National Technical Conference, Houston, Texas, April 11-12, 2017. Available from www.AADE.org.
 7. Alkinani, H. H., Al-Hameedi, A. T., Dunn-Norman, S., Flori, R. E., Al-Alwani, M. A., & Mutar, R. A. (2019, August 28). Dynamic Neural Network Model to Predict the Rate of Penetration Prior to Drilling. American Rock Mechanics Association.
 8. Alpaydin, E. (2014). Introduction to machine learning. Cambridge, MA: The MIT Press.
 9. Ashrafi, S.B., Anemangely, M., Sabah, M., Ameri, M.J., 2019. Application of hybrid artificial neural networks for predicting rate of penetration (ROP): a case study from Marun oil field. *J. Petrol. Sci. Eng.* 175, 604–623.
 10. Basarir, H., Tutluoglu, L., Karpuz, C., 2014. Penetration rate prediction for diamond bit drilling by adaptive neuro-fuzzy inference system and multiple regressions. *Eng. Geol.* 173, 1–9.
 11. Basra Oil Company (2017). Various Daily Drilling Reports, Final Reports, Drilling Time Analysis, and Tests for Several Drilled Wells, Rumaila Field, Iraq.
 12. Bingham, G., 1965. A new approach to interpreting rock drillability. Technical Manual Reprint. *Oil Gas J.* 93, 1965.
 13. Bourgoyne A.T., Millheim K.K., Chenevert M.E., and Young F.S. 1986. Applied Drilling Engineering. Society of Petroleum Engineers Text Book Series, Vol. 1, Richardson, TX.
 14. Bourgoyne, A.T., Young, F.S., 1974. A multiple regression approach to optimal drilling and abnormal pressure detection. *SPE. J.* 14 (4), 371–384.
 15. Chang, C., Zoback, M. D., & Khaksar, A. (2006). Empirical relations between rock strength and physical properties in sedimentary rocks. *Journal of Petroleum Science and Engineering*, 51(3), 223–237. <https://doi.org/https://doi.org/10.1016/j.petrol.2006.01.003>.
 16. Darbor, M., Faramarzi, L., Sharifzadeh, M., 2017. Performance assessment of rotary drilling using non-linear multiple regression analysis and multilayer perceptron neural network. *Bull. Eng. Geol. Environ.* 1–13.
 17. Gan, C., Cao, W.-H., Wu, M., Chen, X., Hu, Y.-L., Liu, K.-Z., ... Zhang, S.-B. (2019). Prediction of drilling rate of penetration (ROP) using hybrid support vector regression: A case study on the Shennongjia area, Central China. *Journal of Petroleum Science and Engineering*, 181, 106200. doi: 10.1016/j.petrol.2019.106200.
 18. Hareland, G., & Rampersad, P. R. (1994, January 1). Drag - Bit Model Including Wear. Society of Petroleum Engineers. doi:10.2118/26957-MS.
 19. Hedge, C., Daigle, H., Millwater, H., Gray, K., 2017. Analysis of rate of penetration (ROP) prediction in drilling using physics-based and data-driven models. *J. Petrol. Sci. Eng.* 159, 295–306.
 20. Hedge, C., Gray, K., 2018. Evaluation of coupled machine learning models for drilling optimization. *J. Nat. Gas. Sci. Eng.* 56, 397–407.
 21. Hossain, M.E., and Al-Majed, A.A. 2015. Fundamentals of Sustainable Drilling Engineering. Scrivener Publishing LLC. 27 Feb.
 22. MATLAB Statistics and Machine Learning Toolbox™ User's Guide Release 2019a, The MathWorks, Inc., Natick, Massachusetts, United States.
 23. Mendes, J.R.P., Fonseca, T.C., Serapião, A., 2007. Applying a genetic neuro-model reference adaptive controller in drilling optimization. *World Oil* 29–36.
 24. Moradi, H., Bahari, M.H., Naghibi Sistani, M.B., Bahari, A., 2010. Drilling rate prediction using an innovative soft computing approach. *Sci. Res. Essays* 5.
 25. Moraveji, M.K., Naderi, M., 2016. Drilling rate of penetration prediction and optimization using response surface methodology and bat algorithm. *J. Nat. Gas. Sci. Eng.* 31, 829–841.
 26. Motahhari, H.R., Hareland, G., James, J., 2010. Improved drilling efficiency technique using integrated PDM and PDC bit parameters. *J. Can. Petrol. Technol.* 49 (10), 45–52.
 27. Nascimento, A., Kutas, D.T., Elmgerbi, A., Thonhauser, G., Mathias, M.H., 2015. Mathematical modeling applied to drilling engineering: an application of Bourgoyne and Young ROP model to a presalt case study. *Math. Probl. Eng.*, 631290.
 28. Osgouei R.E. 2007. Rate of Penetration Estimation Model for Directional and Horizontal Wells. Master Thesis, the Graduate School, Middle East Technical University.
 29. Ricardo, J., Mendes, P., Fonseca, T. C., and Serapiao, A. B. S. 2007. Applying a Neuro-model Reference Adaptive Controller in Drilling Optimization. *World Oil Magazine*, 228: 29-38.
 30. Shi, X., Liu, G., Gong, X.L., Zhang, J.L., Wang, J., Zhang, H.N., 2016. An efficient approach for real-time prediction of rate of penetration in offshore drilling. *Math. Probl. Eng.*, 3575380.
 31. Sikes, S.T. (1936), Drilling Speeds. API Drilling and Production Practice pp. 54-59.
 32. Soares, C., Gray, K., 2019. Real-time predictive capabilities of analytical and machine learning rate of penetration (ROP) models. *J. Petrol. Sci. Eng.* 172, 934–959.
 33. Warren, T., 1987. Penetration rate performance of roller cone bits. *SPE Drill. Eng.* 2 (01), 9–18.
 34. Winters, W. J., Warren, T. M., & Onyia, E. C. (1987, January 1). Roller Bit Model With Rock Ductility and Cone Offset. Society of Petroleum Engineers. doi:10.2118/16696-MS.
 35. Yagiz, S., Gokceoglu, C., Sezer, E., Iplikci, S., 2009. Application of two non-linear prediction tools to the estimation of tunnel boring machine performance. *Eng. Appl. Artif. Intell.* 22, 808–814.
 36. Yilmaz, I., Kaynar, O., 2011. Multiple regression, ANN (RBF, MLP) and ANFIS models for prediction of swell potential of clayey soils. *Expert Syst. Appl.* 38 (5), 5958–5966.

Appendix A

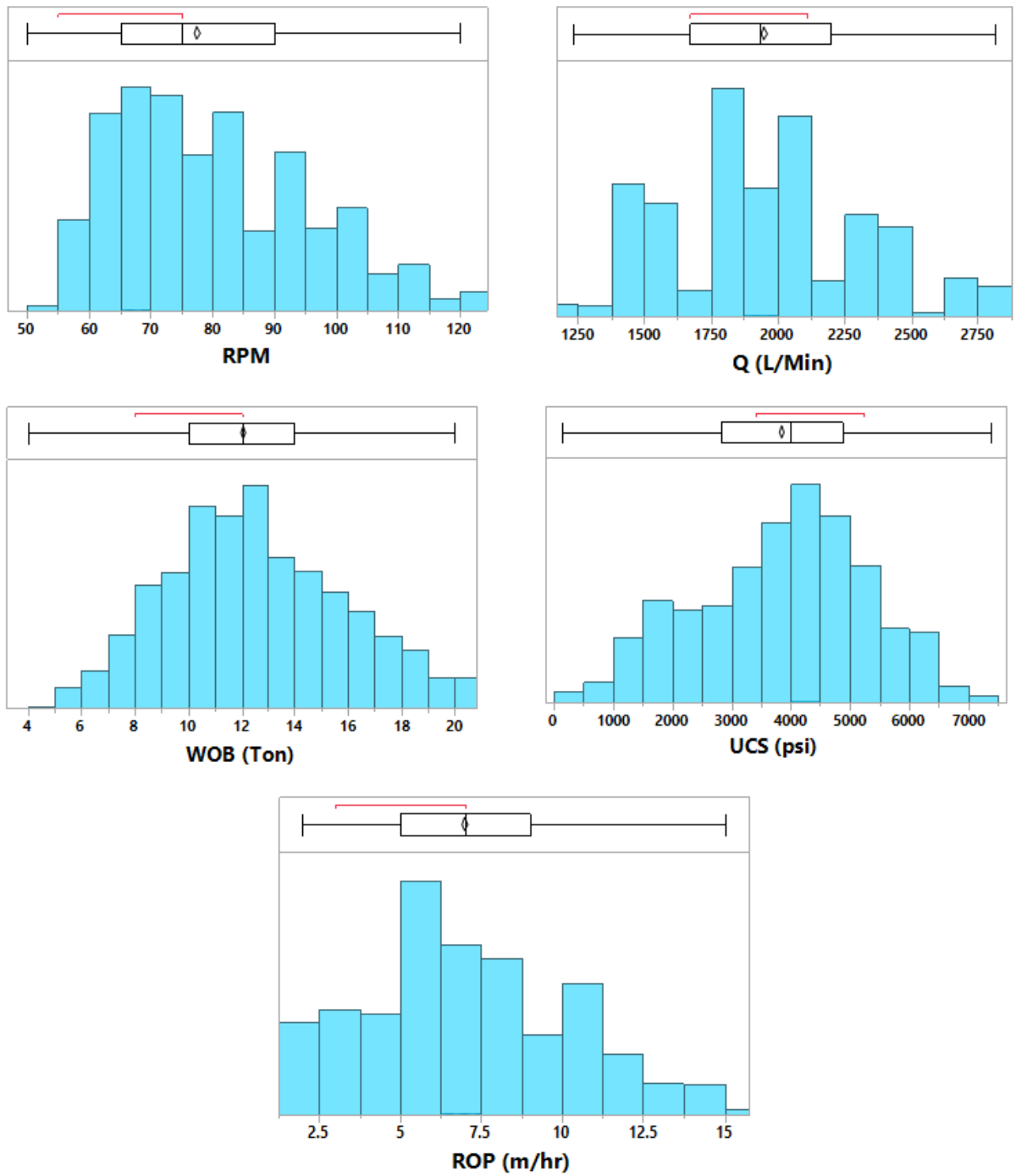


Figure 1. Data Distribution

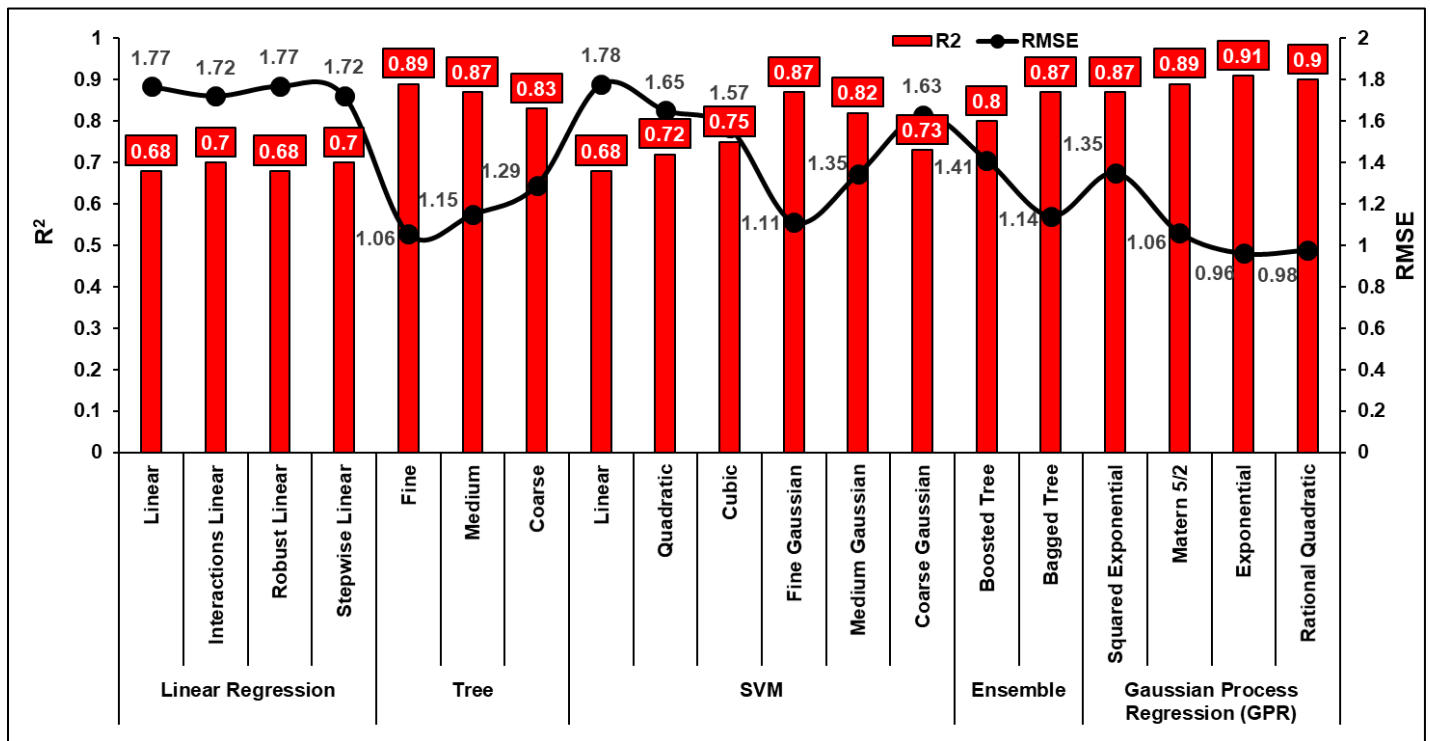


Figure 2. Summary of the R² and RMSE Results of All Algorithms

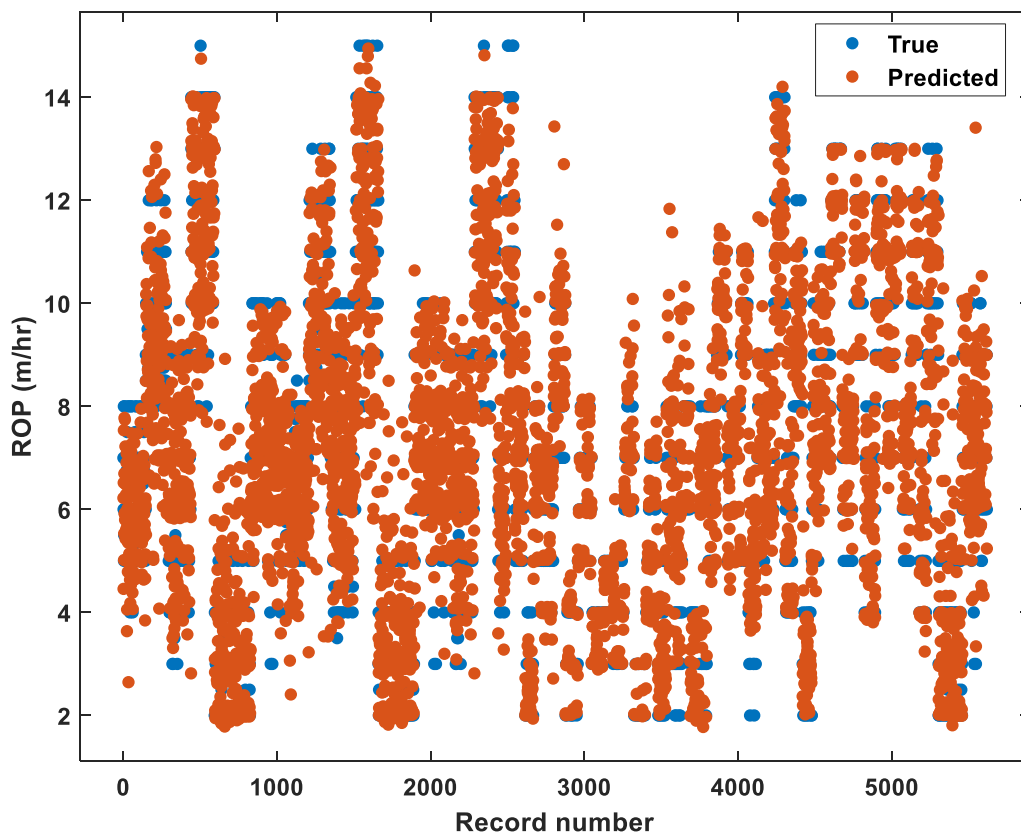


Figure 3. Predicted ROP vs Row Number for the Exponential GPR

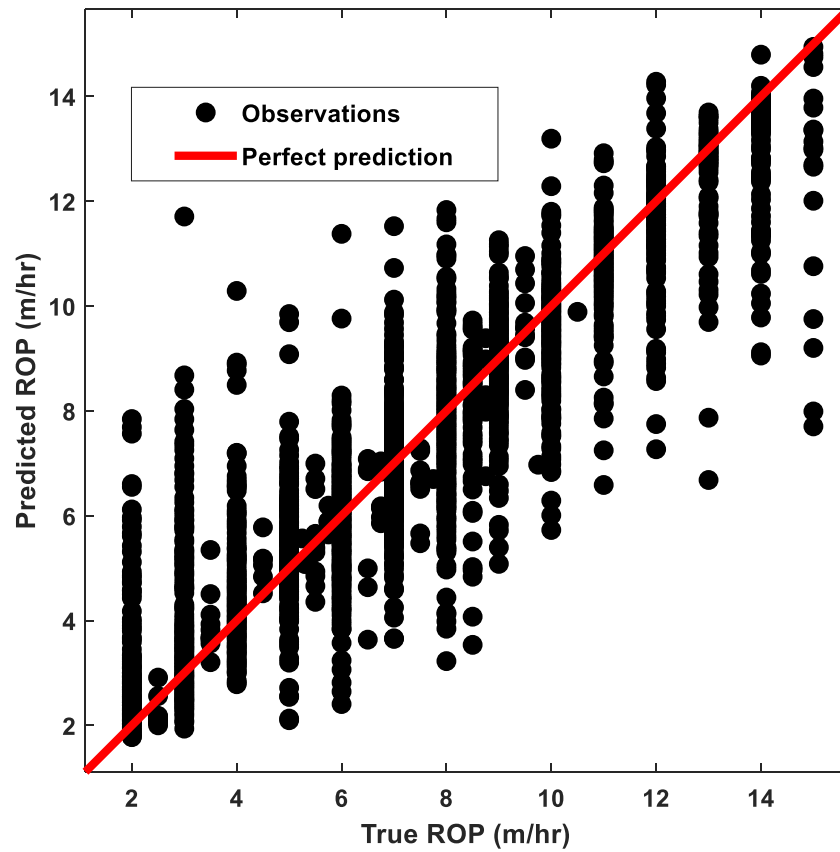


Figure 4. Predicted and True ROP for the Exponential GPR

Table 1. Summary of Statistics

Parameter	Mean	Median	Standard Deviation	Min	Max
RPM	77.5	75	15.5	50	120
Q (L/Min)	1950	1936	361.8	1232	2816
WOB (Ton)	12	12	3.3	4	20
UCS (psi)	3839	4002	1436	140	7374
ROP (m/hr)	7	7	3	2	15

Table 2. Machine Learning Algorithms Used in this Study

Model Type	Algorithm
Linear Regression	Linear
	Interactions Linear
	Robust Linear
	Stepwise Linear
Tree	Fine
	Medium
	Coarse
SVM	Linear
	Quadratic
	Cubic
	Fine Gaussian
	Medium Gaussian
	Coarse Gaussian
Ensemble	Boosted Tree
	Bagged Tree
Gaussian Process Regression (GPR)	Squared Exponential
	Matern 5/2
	Exponential
	Rational Quadratic