Prediction of Rate of Penetration Using Gaussian Process Regression and Bayesian Optimization for Drilling Process

Kanghui Zeng[1,2,3], Min Wu[1,2,3,4] ⋆, Chengda Lu[2,3,4], Xiao Yang[1,2,3], and Zhejiaqi Ma[2,3,4]

[1] School of Future Technology, China University of Geosciences, Wuhan 430074, China
[2] Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems, Wuhan 430074, China
[3] Engineering Research Center of Intelligent Technology for Geo-Exploration, Ministry of Education, Wuhan 430074, China
[4] School of Automation, China University of Geosciences, Wuhan 430074, China

Abstract. Rate of penetration prediction is essential for improving drilling efficiency due to its crucial contribution in the optimization of operational parameters. Accurate rate of penetration prediction enables better decision-making and reduces drilling costs, which helps obtain optimal operational parameters. This article proposes a new prediction model that combines gaussian process regression and bayesian optimization methods. Firstly, the interquartile range method and the Savitzky-Golay filtering methods are used to denoise data. Appropriate input variables are identified based on spearman correlation analysis to reduce the model redundancy. Secondly, the gaussian process regression model tuned by bayesian optimization is established to predict the rate of penetration. Finally, the public data sourced from the UTAH FORGE Well 58-32 dataset are used to validate the proposed model. The results indicate that the proposed model offers reliable prediction accuracy and serves as a valuable reference for enhancing the rate of penetration during the drilling process.

Keywords: Rate of penetration; Gaussian process regression; Bayesian optimization; Drilling process.

1 Introduction

Rate of Penetration (ROP) plays a crucial role in enhancing drilling efficiency and minimizing drilling cost, which reflects the depth drilled per unit time. Accurate mathematical models relating rate of penetration to influential factors are vital for precise ROP prediction [1]. The influence of varying drilling variables on ROP, coupled with its nonlinear, complex, and random characteristics,

⋆ Corresponding author: Min Wu(wumin@cug.edu.cn)

makes it challenging to mitigate the effects of non-fixed variables like geological conditions and equipment features [2].

Existing ROP prediction models can be categorized into two groups: traditional and machine learning models [3]. Physics-based traditional models including the well-known bourgoyne and young (BY) models [4] have limitations due to their dependency on data analyse methods, bottom hole assembly requirements and geological properties, resulting in suboptimal practical results [5].

To overcome these limitations, researchers have adopted machine learning models such as back-propagation neural network (BPNN) [6], artificial neural networks (ANN) [7], random forests [8], and extreme learning machines (ELM) [9]. These models offer superior function approximation capabilities. Gan *et al.* [10] optimized a support vector regression model using a hybrid bat algorithm, which outperformed traditional methods. Similarly, Ahmed *et al.* [11] confirmed the feasibility of intelligent algorithms in ROP prediction by analyzing the accuracy of various models. However, conventional machine learning models often neglect the data distribution characteristics of feature variables during the drilling process.

Considering that the distribution of key feature variables typically follows a gaussian distribution, it is reasonable to view the drilling process as a gaussian process. This insight leads us to the utilization of gaussian process regression (GPR) as a more fitting model for ROP prediction. Within the gaussian process regression framework, the accuracy of predictions is heavily reliant on the selection of hyperparameters such as kernel functions [12]. To address this dependency, a bayesian optimization algorithm is introduced for optimizing these hyperparameters, thereby improving the predictive precision of the GPR model for ROP prediction [13].

In this article, a new rate of penetration prediction method is proposed using bayesian-optimized gaussian process regression. Firstly, the interquartile range method and the Savitzky-Golay filtering methods are used to denoise data. Then, a new bayesian-optimized gaussian process regression (BO-GPR) prediction model is established for drilling process, which is a non-parametric, probabilistic model that provides both predictions and uncertainty estimates. Meanwhile, the bayesian optimization is utilized to optimize the hyperparameters of the GPR model to enhance its predictive performance. Simulation results are conducted using data collected from the UTAH FORGE Well 58-32 drilling site.

2 Process description and scheme design

This section details the drilling process, analyzes the drilling characteristics and provides a scheme of ROP prediction.

2.1 Process description and characteristic analysis

The overview diagram of the geological drilling process is depicted in Fig. 1. During the drilling process, the drill-string applies sufficient weight-on-bit and rotary speed to the drill bit to ensure continuous rock breaking. Drill mud is pumped into the hollow drill pipe via mud pumps, exits through the drill bit, and carries the rock cuttings formed from the broken rock back to the mud pit through the annulus.

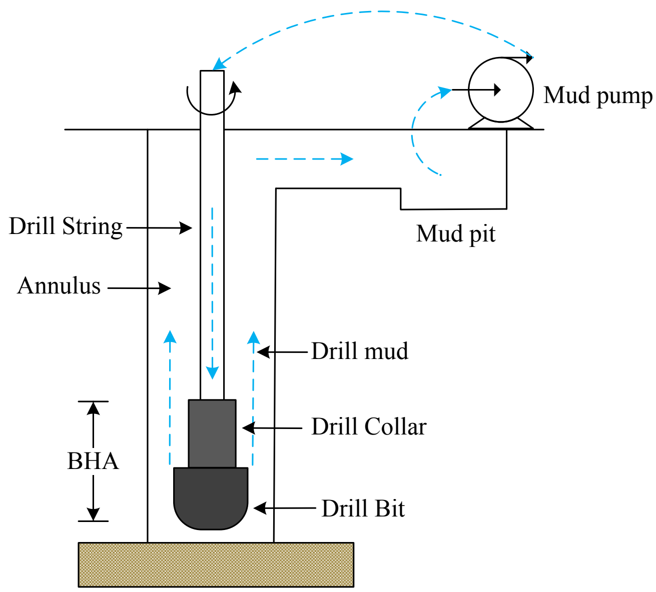

Fig. 1. Overview diagram of geological drilling process

The drilling variables are listed in Table 1. Drilling variables are divided into operational variables and state variables. Operational variables include W_b, R, F_{in}, and F_{out}. W_b denotes the pressure utilized for rock breaking, R denotes the rotary speed of the drill rod, and F_{out} and F_{in} denote the input and output flow rate of the mud pump per unit time. The adjustment of F_{in} is utilized to cool the drill bit and transport rock cuttings back to the surface. State variables include D, P_p, F_{ROP}, T_{in}, and T_{out}.

The design of reliable prediction modeling for the drilling process also needs to consider the characteristics of the geological drilling process. The analysis of drilling process characteristics can provide a design basis for prediction modeling, and improve the efficiency and safety of the drilling process. The description of drilling characteristics are as follows.

(1) Significant data noise: Factors like uneven soil hardness, information loss, and sensor errors contribute to substantial high and low frequency noise in drilling data. This manifests as abnormal values, missing data, and sudden changes. Such noise can disrupt parameter balance, leading to incorrect data rule learning and impacting the accuracy of mechanisms and machine learning models.

Table 1. Drilling variables

Variable	Description	Unit
W_b	Weight on bit	KN
R	Rotary speed	r/min
F_{in}	Flow in rate	m^3/min
F_{out}	Flow out rate	m^3/min
W_p	Wellhead Pressure	KPa
D	Depth	m
P_p	Pump pressure	KPa
S_t	Surface torque	KPa
V_{ROP}	Rate of penetration	m/h
T_{in}	Temperature in	
T_{out}	Temperature out	
H_l	Hook load	Kg

(2) Intense parameter coupling: Drilling variables are highly interconnected. For instance, WOB, rotary speed, and torque show strong correlation, while variables like flow rate and depth significantly influence standpipe pressure. Changes in one parameter induce variations in others due to the inherent mechanics of drilling.

(3) Stochastic and uncertain nature: Drilling processes are inherently stochastic and uncertain due to the unpredictable nature of the geological formations being drilled. This unpredictability manifests as random variations in the drilling data, which can be challenging to model using conventional methods. However, these random variations often exhibit Gaussian-like properties, i.e., they tend to follow a normal distribution.

2.2 Prediction scheme design

Based on the characteristics analysis of the drilling process, our prediction scheme is divided into two stages. Fig. 2 depicts the framework of the proposed method.

In the first stage, the outliers and noise are removed to enhance the quality of the data. Furthermore, spearman correlation analysis is employed to select the most relevant variables due to the inherent coupling in the data. This step helps in reducing the dimensionality of the data and focusing on the most informative variables, thereby improving the efficiency and accuracy of the subsequent modeling stage.

In the second stage, the BO-GPR prediction model is utilized to tackle the uncertainty and randomness in the data. This model leverages the features selected in the first stage to predict the ROP. The BO-GPR model is particularly

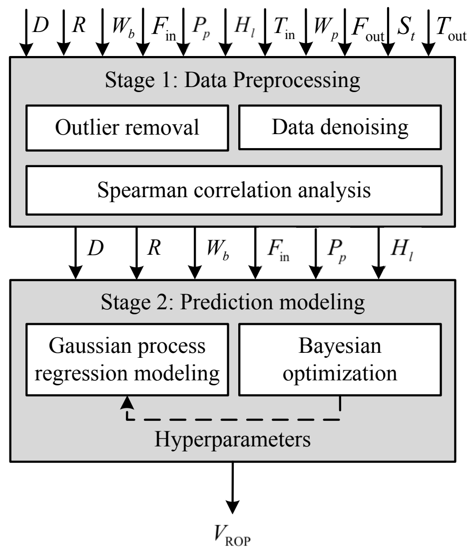

Fig. 2. Framework of the ROP prediction model

suitable for this application due to its ability to handle the non-linearity, noise, and stochastic nature of the drilling data.

3 Prediction model of rate of penetration

This section outlines the required data preprocessing and modeling methods for a precise ROP prediction model.

3.1 Data preprocessing and feature selection

The drilling site environment is complex, with various interferences that lead to some white noise in the data signals received by the measuring instruments. In order to enhance the quality of drilling data, it is essential to remove outliers and minimize noise. The Interquartile Range (IQR) method and Savitzky-Golay filtering technique are employed for this purpose. The IQR method is utilized for outlier detection and removal, while the Savitzky-Golay filter smooths the data by reducing high-frequency noise without significantly altering the shape and width of the signal. This ensures that the filtered data closely resemble the original waveform [14].

In addition, it is important to perform correlation analysis to minimize model redundancy and enhance model precision. The Spearman correlation coefficient method is used for this analysis. The the spearman correlation coefficient between V_{ROP} and drilling variable X is expressed as

$$\rho(V_{\mathrm{ROP}}, X) = 1 - \frac{6\sum(p_i - q_i)^2}{n(n^2-1)}, \tag{1}$$

where p_i represents the rank of V_{ROP}, q_i represents the rank of drilling variable X, and n represents the number of observations. A higher value of ρ signifies a stronger correlation between the two variables. The variables with the highest correlation coefficients with V_{ROP} will be selected for subsequent regression analysis. Table 2. summarizes the spearman correlation between V_{ROP} and drilling variables.

Table 2. Spearman correlation between V_{ROP} and drilling variables.

drilling variables	Spearman correlation coefficient
D	-0.78
H_l	-0.74
F_{in}	0.64
W_b	-0.60
R	0.51
P_p	-0.50

3.2 Gaussian Process Regression model

The drilling process is inherently stochastic and uncertain due to unpredictable geological formations, leading to data variations that often exhibit Gaussian-like properties. This makes GPR suitable for modeling drilling data. Combining GPR with bayesian optimization allows for effective handling of uncertainties by optimizing hyperparameters, improving the model's accuracy and reliability. Thus, the BO-GPR model is ideal for predicting ROP during drilling process.

GPR is a kernel-based machine learning approach based on statistical learning and Bayesian theory [15]. This approach is not only suited to handle complex regression problems such as small sample size, high dimensionality, and nonlinearity, but also has strong learning and generalization abilities. As a non-parametric method, this method organizes data in a way that any given subset of the data always follows a multivariate Gaussian distribution.

In GPR, the mean function [base function $\mu(X)$] and covariance function [kernel function $k(X, X')$] are used to represent the real process, where X and X' indicate different inputs, and $f(X)$ represents the output. The mean function and the covariance function are expressed as

$$\mu(X) = E(f(X)), \tag{2}$$

$$k(X, X') = E\left[(f(X) - \mu(X))\,(f(X') - \mu(X))\right]. \tag{3}$$

GPR is expressed as

$$f(X) \sim G_P(\mu(X), k(X, X')), \tag{4}$$

where E represents the expected function; G_P represents the Gaussian process. In the ideal scenario, a Gaussian process can be considered noise-free. In this case, the joint prior distribution of the training output y_t and the test output $f_\star$ can be represented as

$$\begin{bmatrix} y_t \\ f_\star \end{bmatrix} \sim N\left(\begin{bmatrix} \mu(X_t) \\ \mu(X_\star) \end{bmatrix}, \begin{bmatrix} K(X_t, X_t) & K(X_t, X_\star) \\ K(X_\star, X_t) & K(X_\star, X_\star) \end{bmatrix}\right), \tag{5}$$

$$K(X_\star, X_\star) = \begin{bmatrix} k(x_1^\star, x_1^\star) & k(x_1^\star, x_2^\star) & \dots & k(x_1^\star, x_n^\star) \\ k(x_2^\star, x_1^\star) & k(x_2^\star, x_2^\star) & \dots & k(x_2^\star, x_n^\star) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n^\star, x_1^\star) & k(x_n^\star, x_2^\star) & \dots & k(x_n^\star, x_n^\star) \end{bmatrix}, \tag{6}$$

where N represents Gaussian distribution.

The drilling process involves multiple process variables and strong nonlinear complex characteristics. At the same time, feature variables such as W_b mainly follow a Gaussian distribution, suggesting that the drilling process can be viewed as a Gaussian process. Therefore, adopting the GPR method to construct a ROP prediction model can lead to enhanced predictive performance [16].

After feature selection, 6 drilling variables: W_b, R, F_{in}, D, P_p and H_l are utilized as input variables of the GPR model. The V_{ROP} is the target variable of the model.

However, for the ROP prediction in actual drilling processes, noise-free predictions are often inaccurate because they do not consider the randomness in the observations. Therefore, in this scenario, Gaussian white noise, reflecting the randomness in the observed results, is incorporated into the Gaussian process regression model. Transforming the noise-free form into a noisy form, the V_{ROP} prediction model is expressed as

$$V_{\text{ROP}} = f(X) \sim G_P(\mu(X), k(X, X')) + \varepsilon, \tag{7}$$

where ε represents Gaussian white noise which obeys $\varepsilon \sim N\left(0, \sigma_n^2\right)$, data set X is the model inputs

$$X = \{W_b, R, F_{\text{in}}, D, P_p, H_l\}. \tag{8}$$

Then, the prior distribution of V_{ROP} is expressed as

$$V_{\text{ROP}} \sim N\left(\mu(X), \kappa(X, X') + \sigma_n^2 I_n\right), \tag{9}$$

where I_n represents the n-dimensional identity matrix. Introducing new model inputs as $\hat{X}$, and the corresponding predicted output as $\widehat{V_{\text{ROP}}}$. Assuming that $\widehat{V_{\text{ROP}}}$ and V_{ROP} follow the joint Gaussian distribution, their joint prior distribution can be described as

$$\begin{bmatrix} V_{\text{ROP}} \\ \widehat{V_{\text{ROP}}} \end{bmatrix} \sim N\left(\begin{bmatrix} \mu(X) \\ \mu(\hat{X}) \end{bmatrix}, \begin{bmatrix} K(X, X) + \sigma_n^2 I_n & K\left(X, \hat{X}\right) \\ K\left(\hat{X}, X\right) & K\left(\hat{X}, \hat{X}\right) \end{bmatrix}\right). \tag{10}$$

In Gaussian process regression methods, the accuracy of predictions largely depends on the choice of the kernel function $k(X, X')$ used to represent the covariance. The functional forms of the mean (base) function and covariance (kernel) function can be considered as Gaussian process hyperparameters, which also include length scale, signal variance, and noise variance. To achieve better prediction results, optimization of these hyperparameters is necessary, involving the selection of optimal mean (base) function and covariance (kernel) function, length scale, signal variance, noise variance, among others.

3.3 Model hyperparameter optimization

Machine learning algorithms involve various hyperparameters, and optimizing these is crucial because they greatly influence model behavior. Hyperparameters can be set manually or automatically, with manual methods being time-consuming and potentially ineffective. Recently, Bayesian optimization (BO) has become widely used for hyperparameter tuning. BO optimizes the objective function by iteratively updating the posterior distribution based on new sample points until it closely matches the true distribution. In essence, it uses insights from previous parameters to better adjust the current ones [17].

Bayesian optimization offers several advantages over grid search or random search:

(1) Utilization of Gaussian processes: Bayesian optimization adopts Gaussian processes and considers past parameter information for continuous updates, unlike grid or random search methods.

(2) Efficiency in parameter tuning: Bayesian optimization typically demands fewer iterations compared to grid search, making it a more efficient choice„ particularly with a large number of variables.

(3) Global optimization potential: In the case of non-convex problems, Bayesian parameter optimization is more likely to identify the global optimum instead of getting trapped in local optima, which can be a limitation of grid or random search methods.

As an effective global optimization algorithm, Bayesian optimization can obtain the optimal solution $x^\star$ of the objective function $f(X)$ within a given range of variables by first obtaining the initial distribution of $f(X)$ through random sampling and calculation, and then optimizing the objective function internally. Its expression is as follows

$$x^\star = \arg\min_{x \in D} f(x) . \tag{11}$$

where x is hyperparameters; D is the search space of x; $f(x)$ is the objective function. The Bayesian optimization-based Gaussian process regression model is known for its robust flexibility and ability to capture nonlinear relationships in the data, so it's used in this article.

4 Simulation and results analysis

In this section, the simulation environment is described, and the application of our model within this setting is presented. The results are then discussed. The performance of the proposed model is assessed by using a public drilling dataset to simulate ROP prediction through Bayesian Optimized Gaussian Process Regression (BO-GPR). Additionally, the proposed model is compared with four widely used models, including Random Forest, Decision Tree, and BP.

4.1 Simulation setup

The data used in this article is sourced from the Utah FORGE geothermal well 58-32, available at https://gdr.openei.org. The dataset includes 27 drilling performance variables and comprises 7,311 observations. Due to the presence of many outliers and noisy data, data preprocessing and feature selection were necessary.

The boxplot of the drilling dataset is shown in Fig. 3, which clearly demonstrates whether each drilling variable contains outliers, based on which we can remove these outliers.

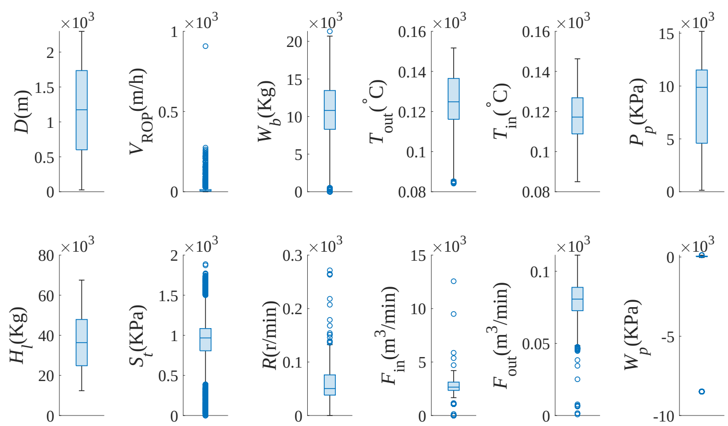

Fig. 3. Boxplot of the drilling dataset.

The actual data, which contains many peaks and mutations, is denoised using the Savitzky-Golay filtering method, as shown in the Fig. 4. The data curve after applying the Savitzky-Golay filter is smoother and clearer in terms of curve profile compared to the original data. It can be seen that the Savitzky-Golay filter effectively removes noise interference from the original signal, and the denoised data still retains the original data's variation characteristics. It is noteworthy that the smoothing factor we set is 0.05, as we found that the larger the smoothing factor, the worse the regression effect.

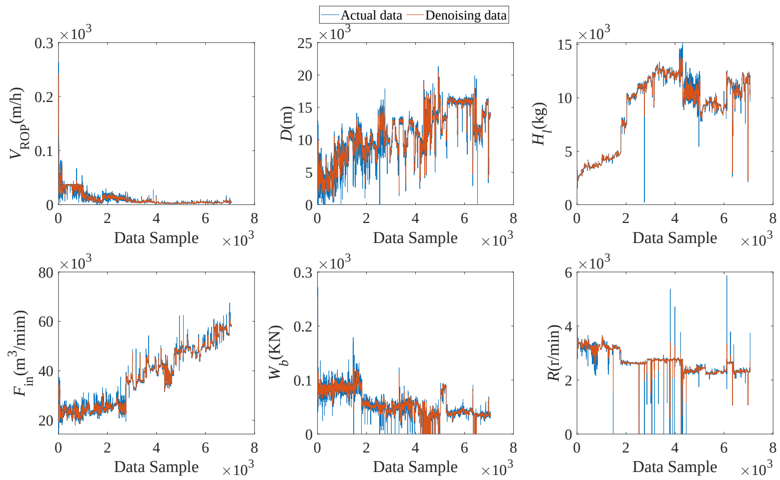

Fig. 4. Data denoising of actual drilling process variables.

After data preprocessing, drilling variables are selected using the spearman correlation coefficient [18]. The correlation coefficients between the drilling variables and V_{ROP} are shown in Fig. 5, which indicates that V_{ROP} is strongly influenced by D, H_l, F_{in}, W_b, R, and P_p, with correlation coefficients of -0.76, -0.73, 0.61, -0.58, 0.57, and -0.45, respectively. In contrast, V_{ROP} is only weakly influenced by the other features.

Therefore, this article selects the following features or drilling variables in drilling process:

Target variable γ : V_{ROP}.

drilling variables X_i: D, H_l, F_{in}, W_b, R and P_p.

4.2 Prediction results

After preprocessing and feature selection, split the data into 80% training and 20% test sets. Use 20% of the training set as a validation set for each fold in 5-fold cross-validation to enhance model generalization [19].

The ROP prediction results of the model proposed in this article are illustrated in Fig. 6. The prediction curve closely aligns with the actual drilling data, demonstrating the model's capability to accurately reflect the real-world drilling process. This alignment not only validates the effectiveness of our approach but also underscores its potential as a reliable tool in drilling operations.

To further illustrate the model's performance, Fig. 7 provides a detailed view of the prediction error. The error range falls between [-1, 1], which is within the

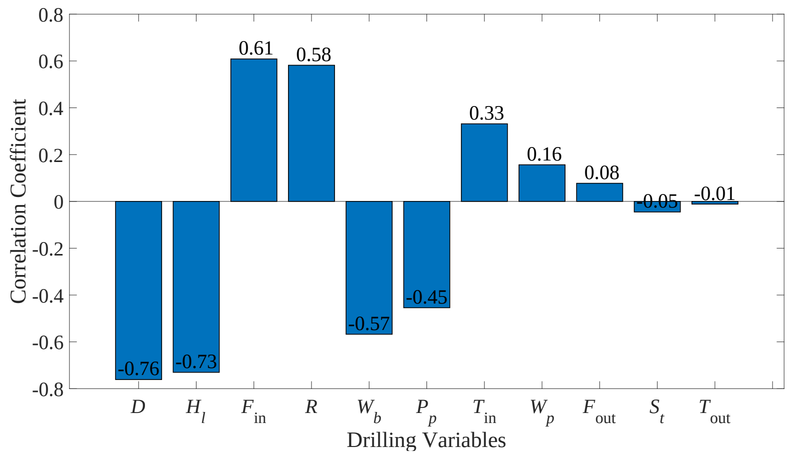

Fig. 5. The correlation coefficients between the drilling variables and V_{ROP}.

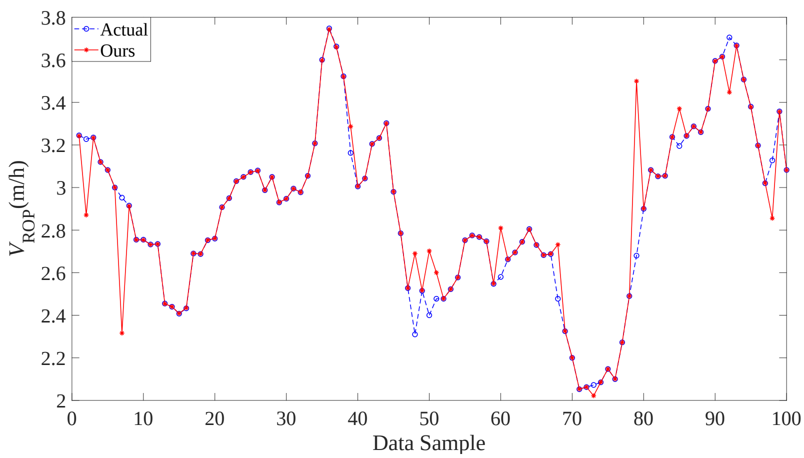

Fig. 6. ROP prediction results of the proposed method.

acceptable limits for drilling operations. This narrow error margin further attests to the model's precision and its ability to provide reliable predictions.

Therefore, the proposed model exhibits excellent performance in predicting the ROP, with a high degree of accuracy and a satisfactory error range. These results highlight the model's robustness and its potential to enhance efficiency and accuracy in drilling operations.

4.3 Analysis and discussion

The performance of the prediction models was evaluated based on three criteria:

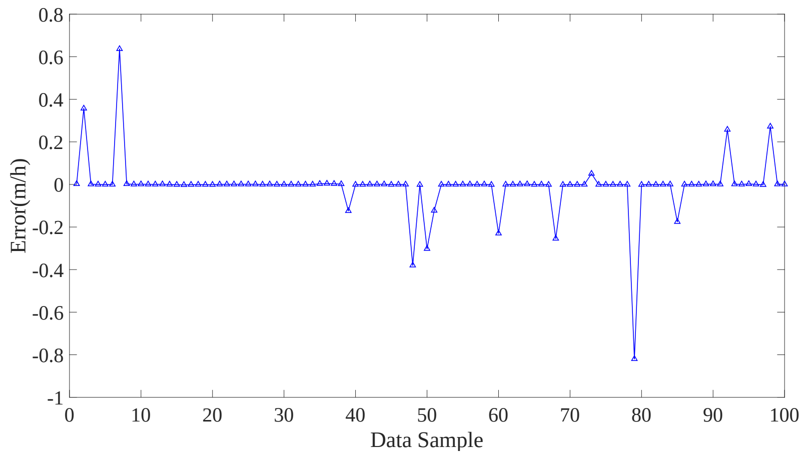

Fig. 7. Prediction error of the proposed method.

(1) coefficient of determination (R^2):

$$R^2 = 1 - \frac{\sum_i^n (y_i - \hat{y}_i)^2}{\sum_i^n (y_i - \bar{y}_i)^2} \tag{12}$$

(2) Root Mean Squared Error (RMSE):

$$R_e = \sqrt{\frac{1}{n}\sum_i^n (y_i - \hat{y}_i)^2} \tag{13}$$

(3) Mean Absolute Error (MAE):

$$M_e = \frac{1}{n}\sum_i^n |y_i - \hat{y}_i| \tag{14}$$

where y_i is the actual value of V_{ROP}; $\hat{y}_i$ is the prediction of V_{ROP}; and $\bar{y}_i$ is the average value of V_{ROP}.

To verify whether the BO-GPR model has superiority over other optimized intelligent models, the random forest model [18], decision Tree model [14], and BP neural network model [6] were selected for comparative analysis. The optimal hyperparameters of each model are shown in Table 3. After using the identical training dataset to train all models, the simulation results of each model are obtained by predicting the identical test dataset. The comparison of prediction results of each model are shown in Fig. 8, which indicates that the predictive values of our BO-GPR model exhibit a consistent trend with the actual data.

Furthermore, the correspondence between the actual test data and prediction results of each model is visually depicted in Fig. 9. Comparison of aforementioned evaluation criteria of different models is shown in Table 4., which indicates that our BO-GPR model exhibits superior prediction accuracy and model robustness,

Table 3. The optimal hyperparameters of each model.

Model	Model hyperparameters	variables values
BO-GPR	Sigma	22.288
	BasisFunction	Zero
	KernelFunction	Isotropic Exponential
	KernelScale	0.005057
BO-RF	EnsembleMethod	Bag
	NumLearners	160
	MinLeafSize	1
	NumberOfPredictorsToSample	4
Decision Tree	MinLeafSize	5
BP	Train function	Trainlgdx
	Transfer functions	tansig, purelin
	hidden layer	15

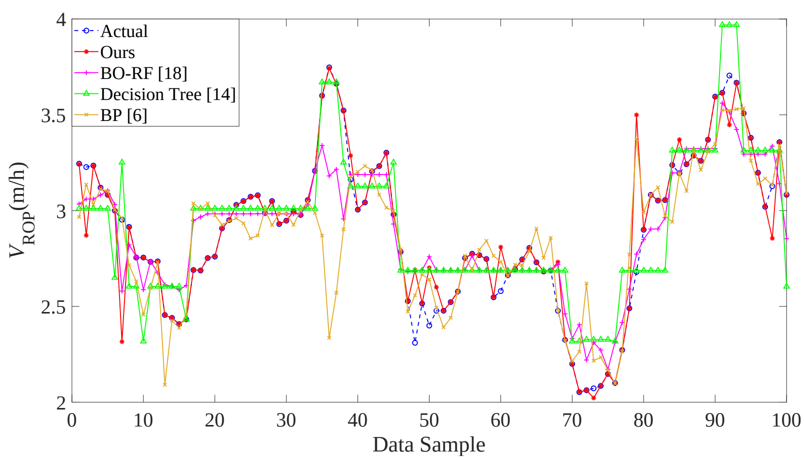

Fig. 8. Comparison of prediction results.

with best values across every evaluation criteria, outperforming other three intelligent models. Therefore, it can be concluded that the proposed BO-GPR model is suitable to predict ROP in geological drilling process.

5 Conclusion

To precisely predict the ROP in complex geological drilling processes, this article introduces a novel Bayesian optimization-based Gaussian Process Regression (BO-GPR) model. The main advantages of the proposed model are as follows:

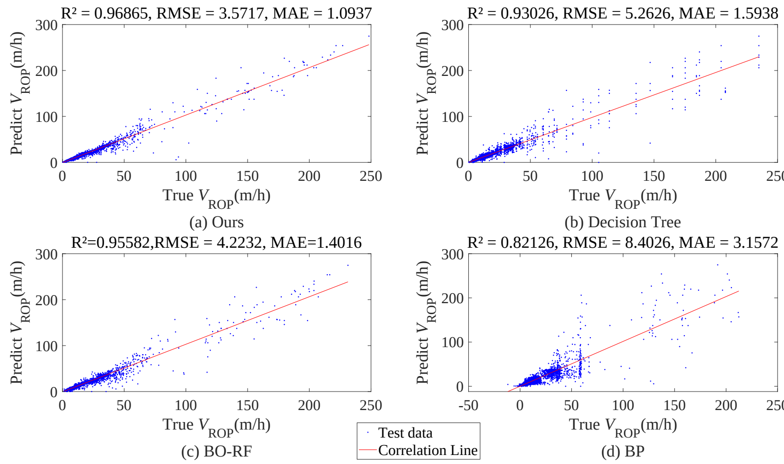

Fig. 9. Actual vs predicted ROP using each model.

Table 4. Comparison of evaluation criteria of each model

Model	R^2	M_r	M_a
Ours	0.969	3.572	1.094
BO-RF [18]	0.956	4.223	1.402
Decision Tree [14]	0.930	5.263	1.594
BP [6]	0.838	8.045	3.182

(1) Themodel effectively handles the inherent stochasticity and uncertainties in drilling data by optimizing hyperparameters, thereby capturing the underlying data structure and providing more accurate and reliable predictions of the drilling rate.

(2) By using robust data preprocessing techniques like the IQR method and Savitzky-Golay filtering, and selecting input variables based on Spearman correlation analysis, the model reduces data noise and redundancy, thereby improving predictive accuracy.

(3) The model shows superior performance compared to other widely used ROP models, as demonstrated by a case study using public data from the UTAH FORGE Well 58-32 dataset.

In the future, we will persist in refining our method and integrate more efficient algorithms to further enhance ROP prediction, laying a foundation for future ROP optimization.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 62333019 and 62273317, the 111 project under Grant

B1704, and the Fundamental Research Funds for the Central Universities, China University of Geosciences.

References

1. Soares, C., Gray, K.: Real-time predictive capabilities of analytical and machine learning rate of penetration (rop) models. Journal of Petroleum Science and Engineering **172**, 934–959 (2019). https://doi.org/10.1016/j.petrol.2018.08.083
2. Barbosa, L.F.F., Nascimento, A., Mathias, M.H., de Carvalho, J.A.: Machine learning methods applied to drilling rate of penetration prediction and optimization - a review. Journal of Petroleum Science and Engineering **183**, 106332 (2019). https://doi.org/https://doi.org/10.1016/j.petrol.2019.106332
3. Hegde, C., Daigle, H., Millwater, H., Gray, K.: Analysis of rate of penetration (rop) prediction in drilling using physics-based and data-driven models. Journal of Petroleum Science and Engineering **159**, 295–306 (2017). https://doi.org/10.1016/j.petrol.2017.09.020
4. Bourgoyne, A.T., J., Young, F.S., J.: A Multiple Regression Approach to Optimal Drilling and Abnormal Pressure Detection. Society of Petroleum Engineers Journal **14**(04), 371–384 (08 1974). https://doi.org/10.2118/4238-PA
5. Negara, A., Saad, B.: Combining insight from physics-based models into data-driven model for predicting drilling rate of penetration. In: IPTC International Petroleum Technology Conference. pp. 5–9 (01 2020). https://doi.org/10.2523/IPTC-20090-MS
6. Liao, X., Khandelwal, M., Yang, H., Koopialipoor, M., Bhatawdekar, R.: Effects of a proper feature selection on prediction and optimization of drilling rate using intelligent techniques. Engineering with Computers **36**, 499–510 (2020). https://doi.org/10.1007/s00366-019-00711-6
7. Bataee, M., Irawan, S., Kamyab, M.: Artificial neural network model for prediction of drilling rate of penetration and optimization of parameters. Journal of the Japan Petroleum Institute **57**(2), 65–70 (2014). https://doi.org/10.1627/jpi.57.65
8. Delavar, M.R., Ramezanzadeh, A., Gholami, R., Sanei, M.: Optimization of drilling parameters using combined multi-objective method and presenting a practical factor. Computers & Geosciences **175**, 105359 (2023). https://doi.org/10.1016/j.cageo.2023.105359
9. Yang, X., Wu, M., Lu, C., Li, W., Chen, L., Du, S.: Prediction of rate of penetration based on drilling conditions identification for drilling process. Neurocomputing **579**, 127439 (2024). https://doi.org/10.1016/j.neucom.2024.127439
10. Gan, C., Cao, W., Wu, M., Chen, X., Zhang, S.: 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 (01). https://doi.org/10.1016/j.petrol.2019.106200
11. Ahmed, O.S., Adeniran, A.A., Samsuri, A.: Computational intelligence based prediction of drilling rate of penetration: A comparative study. Journal of Petroleum Science and Engineering **172**, 1–12 (2019). https://doi.org/10.1016/j.petrol.2018.09.027
12. Alali, Y., Harrou, F., Sun, Y.: Optimized gaussian process regression by bayesian optimization to forecast covid-19 spread in india and brazil: A comparative study. In: 2021 International Conference on ICT for Smart Society (ICISS). pp. 1–6 (2021). https://doi.org/10.1109/ICISS53185.2021.9532501

13. Kaya, E.Y., Ali, I., Ceylan, Z., Ceylan, S.: Prediction of higher heating value of hydrochars using bayesian optimization tuned gaussian process regression based on biomass characteristics and process conditions. Biomass and Bioenergy **180**, 106993 (2023). https://doi.org/10.1016/j.biombioe.2023.106993
14. Yehia, T., Gasser, M., Ebaid, H., Meehan, N., Okoroafor, E.R.: Comparative analysis of machine learning techniques for predicting drilling rate of penetration (rop) in geothermal wells: A case study of forge site. Geothermics **121**, 103028 (2024). https://doi.org/10.1016/j.geothermics.2024.103028
15. Schulz, E., Speekenbrink, M., Krause, A.: A tutorial on gaussian process regression: Modelling, exploring, and exploiting functions. Journal of Mathematical Psychology **85**, 1–16 (2018). https://doi.org/10.1016/j.jmp.2018.03.001
16. Wu, X., Lai, X., Hu, J., Chen, X., Cao, W., Wu, M.: Rate of penetration prediction using hybrid gaussian process regression model with ard structure in geological drilling process. In: 2022 13th Asian Control Conference (ASCC). pp. 607–612 (2022). https://doi.org/10.23919/ASCC56756.2022.9828292
17. Yao, X., Wu, D.: Prediction of mechanical properties of recycled concrete using bayesian optimization-based gaussian process regression method. Science Technology and Engineering **23**(7), 2968–2975 (2023). https://doi.org/10.12404/j.issn.1671-1815.2023.23.07.02968
18. Ben, A.M.A., Madarász, T.: Applying machine learning to predict the rate of penetration for geothermal drilling located in the utah forge site. Energies **15**(12), 4288 (2022). https://doi.org/10.3390/en15124288
19. Zhou, Y., Chen, X., Wu, M., Cao, W.: A novel modeling and drilling optimization method with suitable constraints in geological well. Control Engineering Practice **122**, 105062 (2022). https://doi.org/10.1016/j.conengprac.2021.105062