Petroleum Science 22 (2025) 430–441

Contents lists available at ScienceDirect

Petroleum Science

journal homepage: www.keaipublishing.com/en/journals/petroleum-science

Original Paper

Real-time drilling torque prediction ahead of the bit with just-in-time learning

Kan-Kan Bai [a], Mao Sheng [a, *], Hong-Bao Zhang [a, b], Hong-Hai Fan [a], Shao-Wei Pan [b]

[a] National Key Laboratory of Petroleum Resources and Engineering, China University of Petroleum (Beijing), Beijing, 102249, China
[b] Sinopec Research Institute of Petroleum Engineering, Beijing, 102206, China

ARTICLE INFO

Article history:
Received 21 June 2023
Received in revised form
11 December 2024
Accepted 14 December 2024
Available online 15 December 2024

Edited by Jia-Jia Fei

Keywords:
Drilling torque prediction
Just-in-time learning
Digital twin
Machine learning

ABSTRACT

The digital twin, as the decision center of the automated drilling system, incorporates physical or data-driven models to predict the system response (rate of penetration, down-hole circulating pressure, drilling torques, etc.). Real-time drilling torque prediction aids in drilling parameter optimization, drill string stabilization, and comparing the discrepancy between observed signal and theoretical trend to detect down-hole anomalies. Due to their inability to handle huge amounts of time series data, current machine learning techniques are unsuitable for the online prediction of drilling torque. Therefore, a new way, the just-in-time learning (JITL) framework and local machine learning model, are proposed to solve the problem. The steps in this method are: (1) a specific metric is designed to measure the similarity between time series drilling data and scenarios to be predicted ahead of bit; (2) parts of drilling data are selected to train a local model for a specific prediction scenario separately; (3) the local machine learning model is used to predict drilling torque ahead of bit. Both the model data test results and the field data application results certify the advantages of the method over the traditional sliding window methods. Moreover, the proposed method has been proven to be effective in drilling parameter optimization and pipe sticking trend detection. Finally, we offer suggestions for the selection of local machine learning algorithms and real-time prediction with this approach based on the test results.

1. Introduction

Digital twin technology (Qi et al., 2021; Tao et al., 2022; Sottet and Pruski, 2023) is on the rise in the drilling industry for monitoring and optimizing processes. By creating simulations that replicate equipment states and rig-formation interactions using numerical models, digital twin technology predicts system responses under varying input parameters. These prediction capabilities aid in decision-making for future drilling operations, ultimately improving efficiency and mitigating risks.

The main parameters reflecting the drilling process are weight on bit (WOB), rotating speed (RPM), mud-flow in (MFI), etc., which are controllable parameters, and several parameters could indicate the response of the rig-formation system, such as rate of penetration (ROP), surface torque, standpipe pressure, mud-flow out, etc., which are uncontrollable parameters. Investigations on controllable and uncontrollable parameters can help to model the interacting mechanism of the rig-formation system. The simulation and prediction methods of ROP (Hegde et al., 2017; Barbosa et al., 2019; Najjarpour et al., 2020; Negara and Saad, 2020; Ren et al., 2022) and standpipe pressures (Erge and van Oort, 2021, 2022; Elmgerbi et al., 2022) etc., are well studied by previous research. However, the studies of drilling torque prediction are few (Marquez, 2021; Song et al., 2022).

During the drilling of a well, energy is applied by surface equipment to the rock through the drill string and bit, creating a channel between the reservoir fluid and surface pipelines. The surface torque measured at the top drive or rotary table is the sum of the torque of the drill bit when breaking rock and the friction between drilling strings with the wellbore and drilling fluids. During drilling, improving the ROP is an important goal of drilling optimization. Field experience shows that increasing WOB can generally enhance ROP. However, as WOB increases, torque will also increase. High torque indicates a high working load of drilling tools both on the surface and down-hole; the abnormal trend of the surface torque signal may indicate a formation change on the

* Corresponding author.
E-mail address: shengmao@cup.edu.cn (M. Sheng).

https://doi.org/10.1016/j.petsci.2024.12.014

down-hole complex (such as pipe sticking, bit wear, etc.). Furthermore, torque is a crucial variable in the computation of mechanical specific energy (MSE) (Teale, 1965; Oyedere and Gray, 2020). The energy used for rock breaking during drilling is known as MSE, and it is a useful indicator of both drilling efficiency and down-hole conditions. In the digital twin system, predicting the surface torque to be drilled ahead of the bit can help the drilling technicians optimize the drilling parameters and provide risk warning by comparing the signal trend of surface torque with the predicted torque. Thus, it is crucial to estimate surface torque in real time during drilling.

At present, predicting surface torque ahead of the bit could be achieved by physical models, such as the soft string model (Johancsik et al., 1984) and the stiff string model (Ho, 1988). The axial force, side force, and torque distribution of drill strings are theoretically calculated, and the surface torque is the sum of the torque along the drill string. Gao (1994) derived the integral mechanical model from the basic equations of down-hole strings. Wu and Juvkam-Wold (1995) established the torque and drag (T&D) prediction model of the down-hole string, considering the buckling effect. McSpadden and Newman (2002), Menand et al. (2006), and Mitchell et al. (2015) studied the influences of pipe joints on T&D. Huang and Gao (2019, 2021) fully considered the coupling effects of joint and buckling effects, then introduced the overall stress model based on descriptions of local models. Samuel and Huang (2020) proposed a dynamic torque and drag model that takes into account the impacts of viscous forces, drill fluid, tubular vibration, and velocity-dependent friction forces, overcoming the shortcomings in previous models. However, the traditional physical model of the pipe string fails to fully consider factors such as mud viscosity, stick-slip, buckling, whirl, key-way effect, etc., and this method is not suitable for real-time deployment due to low efficiency and poor flexibility. In addition, these physical models require down-hole torque as input for bottom-up calculations, and the down-hole torque cannot be obtained without down-hole measuring tools. In the field, surface torque is mostly monitored using a rule-based approach, in which experienced experts define a certain value range to determine if the signal is normal. The rule-based method often provides inaccurate warnings as it cannot consider the influence of controllable parameters for surface torque. Therefore, a robust and flexible method to predict and monitor surface torque is needed.

The development of modern data science, especially machine learning, provides powerful tools for modeling high-dimensional, nonlinear, and complex processes (Liu et al., 2020). Oyedere and Gray (2020) used multiple machine learning algorithms to predict drilling torque-on-bit, and the results indicate that machine learning algorithms can effectively model the nonlinear relationship between drilling features. Marquez (2021) used the sliding window method and the XGBoost algorithm to update the model every 60 m and predict the drilling torque ahead of the bit. Nevertheless, the data from a short window restricts the temporary model's capacity to predict during sliding. Bai et al. (2022) studied a hybrid model by combining the physical model and machine learning methods to predict drilling torque in real time with high stability and interpretability. Song et al. (2022) proposed a BP-LSTM network to predict surface torque. The historical data points of 12 wells were used to train an offline model, and 3 wells were tested with remarkable performance. However, the impracticality of the offline model is evident due to its inability to adjust to dynamic down-hole environmental factors.

This paper proposes a just-in-time learning (JITL) model for predicting drilling torque accurately and reliably, optimizing drilling parameters, and detecting anomalies in real-time. The JITL framework was particularly designed for dynamically selecting training samples from historical data to support model training. The local machine learning model was used to establish the relationship between controllable drilling parameters and surface torque. Our model was validated by field data compared to the traditional sliding window method for real-time drilling torque prediction. It is significant to improve drilling efficiency and reduce risks.

2. Methodology

2.1. Just-in-time learning-based framework

Just-in-time learning (Cybenko, 1996), also known as lazy learning (Galván et al., 2011) and local weighted learning (Yin et al., 2016), is a representative local learning framework that advances in dealing with nonlinear, time-varying behavior. JITL is widely used in soft sensor modeling and process monitoring (Jin et al., 2019, 2020), which can better reflect the local dynamics in engineering (Liu and Gao, 2015; Li et al., 2021; Guo et al., 2020; Jiang and Ge, 2022; Zhang et al., 2022), but is rarely used in petroleum engineering. With historical data in the database, local models are dynamically built by the data most similar to the feature for prediction. Different from traditional global modeling methods that use all historical data to train the model, JITL features two aspects. (1) All available modeling data is stored in the database, and the local model is trained using queried local data from the database. Only parts of the samples in the database that are most similar to the query features are used to build the model, instead of all historical datasets. Therefore, JITL is an online, non-parametric modeling technology. (2) The local model constructed will be discarded after predicting the corresponding features until the next prediction, and then a new local model will be built.

The JITL-based drilling torque prediction framework is shown in Fig. 1. In the historical database, X_i refers to input features for torque prediction, such as depth, WOB, hook load, RPM, MFI, etc., and TQ_i is the label corresponding to X_i. $X_{h,i}$ refers to the input features for torque prediction in a specific scenario. When drilling to depth h and a $X_{h,i}$ is determined, the samples whose input features are similar to $X_{h,i}$ are selected for local model building. The similar features in historical data and predictions are marked with the same color in the diagram. Once a prediction is finished, another cycle of sample selection and local model training is to be conducted.

2.2. Similarity measuring metrics

In the JITL framework, the similarity measuring metric directly impacts the quality of the local model. The most commonly used similarity measuring metrics are Euclidean distance and cosine similarity (Zhang and Rasmussen, 2001; Liu et al., 2019; Kirişci, 2023).

For multi-dimensional vectors A and B, the formulas for calculating Euclidean distance and cosine similarity are:

$$\mathrm{dist}(A,B)=\sqrt{\sum_{i=1}^{k}(A_i-B_i)^2} \tag{1}$$

where, $\mathrm{dist}(A,B)$ denotes the distance between the k-dimensional vector A and B,

$$\cos(A,B)=\frac{A\cdot B}{\|A\|\cdot\|B\|} \tag{2}$$

where, $\cos(A,B)$ denotes the cosine value of the angle between vector A and B.

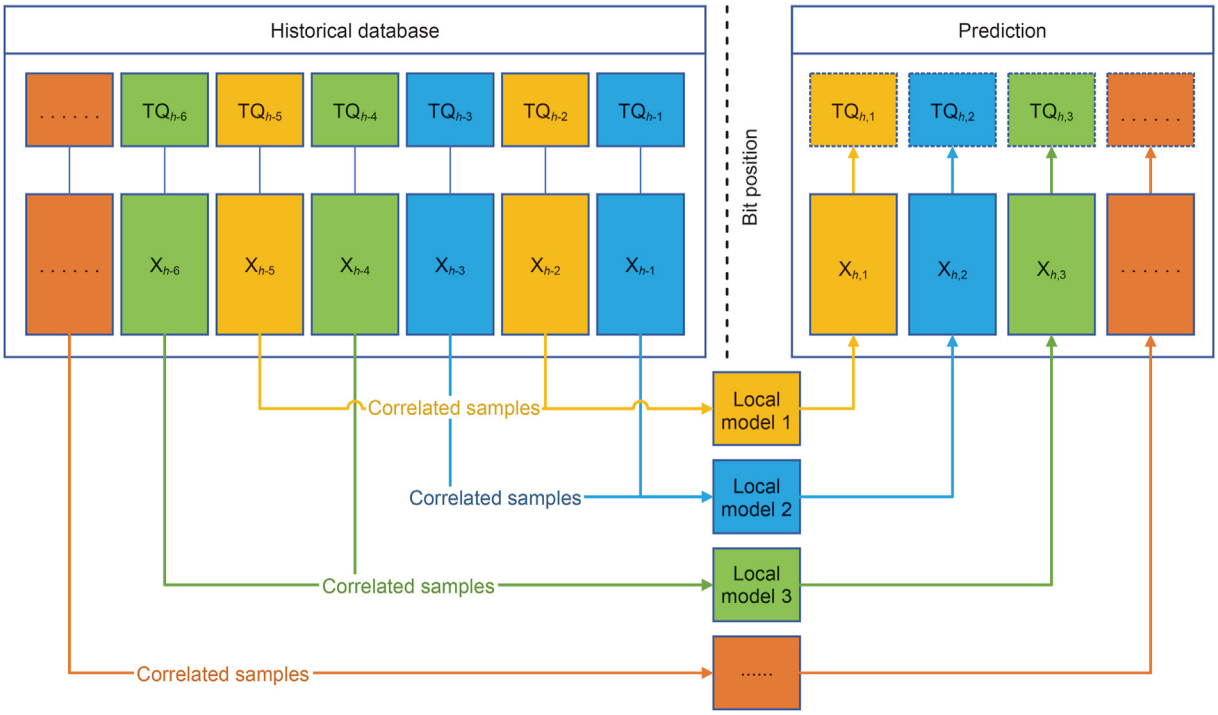

Fig. 1. Diagram of JITL based torque prediction.

As shown in Fig. 2, the Euclidean distance measures the absolute distance between points in space, which is directly related to the feature values of each point. Cosine similarity measures the included angle of the space vector, which can reflect the commonality of each dimension between the two vectors. However, in some cases, such as vectors [1, 2, 3, 4] and [4, 3, 2, 1], their included angles differ greatly, but their distances are relatively close. Another example is the vectors [1, 2, 3, 4] and [2, 4, 6, 8], where the included angle is 0 but the distance is large. From this point of view, Euclidean distance and cosine similarity measure the similarity from different perspectives, and the combination of the two methods is expected to provide an integral measure of similarity.

In this paper, cosine similarity and Euclidean distance are utilized for similarity measurement on normalized data. This approach ensures that the selected related samples have high similarity both in terms of direction and magnitude. The diagram of the similarity measuring metric is shown in Fig. 3.

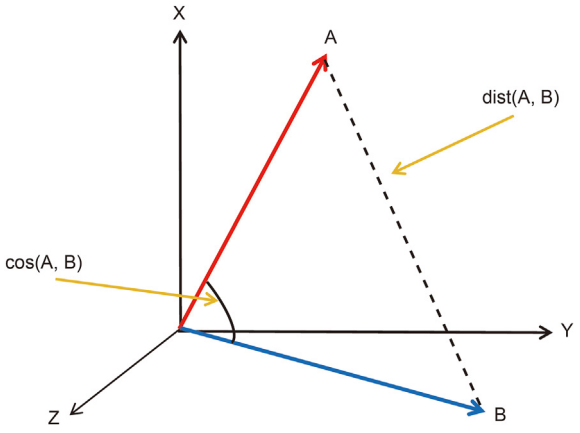

Fig. 2. Diagram of Euclidean distance similarity and cosine similarity.

2.3. Online prediction methods

Online prediction methods continuously update models in real time during ongoing processes to adapt to changing conditions. As shown in Fig. 4(a), in the online prediction process based on JITL, newly transmitted field data is added to the historical database when drilling to the next predetermined well depth (D_i), and then correlated samples are selected in accordance with the procedure in Section 2.2 to build different local models for each test sample to predict. The traditional online prediction method in the drilling industry is the sliding window (SW) method (Marquez, 2021; Yang et al., 2022). As shown in Fig. 4(b), the SW method dynamically extracts drilling data within the current depth window to construct a model and predict all the test samples. When drilling to the next predetermined depth of the well, a new model will be built based on the data from the new window and used to predict, and the historical data before this window will be discarded.

Overall, the following two factors reflect the distinctions between the JITL and SW methods, which are also the benefits and innovations of the JITL method. One of them is that the modeling data is different. The SW method (Fig. 4(b)) employs data within a sliding fixed well depth window as modeling data, while the JITL method (Figs. 1 and 4(a)) selects portions of historical data that are similar to the test sample. Another is that the model's prediction

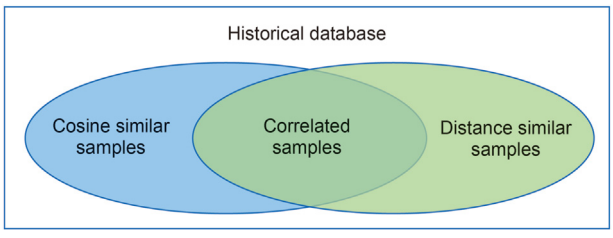

Fig. 3. Diagram of the similarity measuring metric.

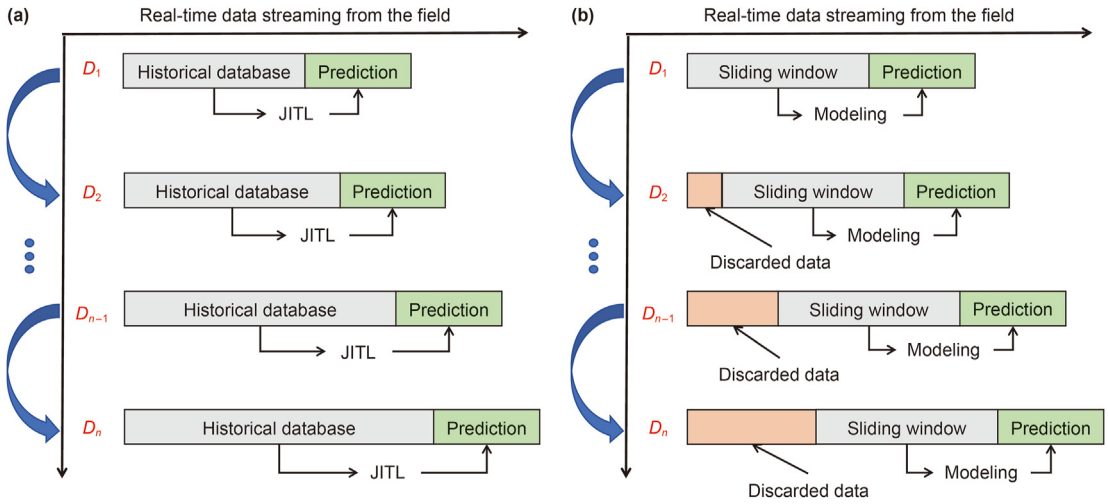

Fig. 4. Diagram of the online prediction methods. (**a**) JITL-based online prediction; (**b**) sliding window online prediction.

range is different. For the test samples at a certain distance ahead of the bit, the SW method (Fig. 4(b)) built a model to predict every one of them, while the JITL approach (Fig. 1) built a local model that only predicts the corresponding sample, and every test sample has a corresponding local model to predict it.

2.4. Machine learning methods for regression

According to the characteristics of JITL, there are three principles for the selection of machine learning models: (1) high capacity for small data modeling is necessary, because the data needed to build local models is generally limited, which leads to algorithms for big data processing such as deep neural networks not being suitable. (2) Amounts of the local model will be trained in real-time, the computing complexity of the machine learning algorithm should be low, and non-parameter methods are preferred, because there is no need to train high hyper-parameters. (3) For industrial deployment, the robustness of the method is important.

Therefore, Linear Regression (LR), Gaussian Process Regression (GPR), and Support Vector Machine (SVM) are evaluated for local model building in the JITL framework.

2.4.1. Multiple linear regression

LR (Narula and Wellington, 2007) features a simple structure, rapid modeling, and good interpretability, which is expressed as,

$$y_i = k_0 + \sum_{i=1}^{m} k_i x_i \tag{3}$$

where, x_i is the explanatory variable, y_i is the dependent variable, k_i is the slope of the line, and k_0 is the intercept.

The model can be regarded as an ordinary normal linear model. Under the normal assumption, for the feature matrix $\boldsymbol{X}$, if $\boldsymbol{X}^T\boldsymbol{X}$ is full rank, the least squares estimate of parameter k is,

$$\widehat{k} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T y \tag{4}$$

and the estimated value of y is,

$$\widehat{y} = \boldsymbol{X}\widehat{k} \tag{5}$$

2.4.2. Gaussian process regression

GPR (Williams and Rasmussen, 2006; Chen et al., 2022) is a non-parametric machine learning method and provides an uncertain description of prediction, which is conducive to ensuring robustness in deployment.

The Gaussian process is a set of random variables in a joint Gaussian distribution. In essence, the Gaussian process is a multivariate Gaussian distribution, which can be expressed as follows,

$$f(x) \sim GP\left[m(x), k(x,x') + {\sigma_n}^2 I\right] \tag{6}$$

where, GP is Gaussian distribution. $m(x)$ is the predicting expectation with input x. x is the feature vector. x' is the feature vector of the test set. ${\sigma_n}^2 I$ is Gaussian random noise matrix. $k(x,x')$ is a covariance function that represents the dependency between x and x'. The Radial Basis Function (RBF kernel) is used as the covariance function in this paper,

$$K(x,x') = \exp\left(-\frac{1}{2\sigma^2}\|x - x'\|^2\right) \tag{7}$$

where, σ is the signal variance.

With the prior defined in Eq. (6), the Gaussian process regression is designed to learn the function from the training data. For given new points x_*, the joint distribution of the training outputs y and the predictive outputs y_* can be given as follows,

$$\begin{bmatrix} y \\ y_* \end{bmatrix} \sim N\left\{0, \begin{bmatrix} \boldsymbol{K}(X,X) + {\sigma_n}^2 I & \boldsymbol{K}(X,x_*) \\ \boldsymbol{K}(x_*,X) & \boldsymbol{K}(x_*,x_*) \end{bmatrix}\right\} \tag{8}$$

where, $\boldsymbol{K}(X,X)$ represents the covariance matrix of training data. $\boldsymbol{K}(x_*,x_*)$ represents the variance of test point. $\boldsymbol{K}(X,x_*) = \boldsymbol{K}(x_*,X)^T$ represents the covariance matrix between the test data and the input of the training data set.

The mean value of the prediction $\overline{\widehat{y}}$ with input x_* is,

$$\overline{\widehat{y}} = \boldsymbol{K}(x_*,X)[\boldsymbol{K}(X,X) + {\sigma_n}^2 I]^{-1} y \tag{9}$$

The variance of prediction $\mathrm{cov}(\widehat{y})$ is,

$$\mathrm{cov}(\widehat{y}) = k(x_*,x_*) - \boldsymbol{K}(x_*,X)[\boldsymbol{K}(X,X) + {\sigma_n}^2 I]^{-1}\boldsymbol{K}(X,x_*) \tag{10}$$

According to statistics, to ensure that a 95% probability of the

observations will be within the confidence interval, the standard deviation of the predictions is always multiplied by a coefficient of 1.96, and the confidence interval is,

$$\left[\overline{\widehat{y}} - 1.96 \times \sqrt{\mathrm{cov}(\widehat{y})}, \overline{\widehat{y}} + 1.96 \times \sqrt{\mathrm{cov}(\widehat{y})}\right] \tag{11}$$

2.4.3. Support vector machines

With the absence of the curse of dimensionality, better generalization due to structural risk minimization, less prone to overfitting, and fewer training features, SVM performs well in robustness (Awad et al., 2015; Olatunji and Micheal, 2017).

SVM maps the input vector to the high-dimensional feature space to conduct prediction, and the formula is,

$$f(x) = W^T \varphi(x) + b \tag{12}$$

where, W is the weight vector, $\varphi(x)$ is the feature mapping function and b is the bias.

Since SVM formulations are done within the context of convex optimization theory, the general methodology is to initiate the formulation of the problem as a constrained optimization problem, subsequently formulate the Lagrangian, and then take the conditions for optimality to ultimately solve the problem in the dual space of Lagrange multipliers. The estimated value is,

$$\widehat{y} = \sum_{i=1}^{m} \alpha' k(x, x') + b \tag{13}$$

where, α' is Lagrange multiplier term.

2.5. Hyperparameter optimization

For SVM and LR, hyperparameter optimization involves systematically exploring various combinations of hyperparameters, such as through grid search, to identify the most effective settings. This procedure demands significant computational resources and may be inefficient. Thus, we predetermine multiple hyperparameter combinations for the model. Each combination is then assessed via k-fold cross-validation. Ultimately, we select the most appropriate parameter combination to establish the model.

The hyperparameter of GPR is $\Theta = \{\sigma, \sigma_n\}$. By utilizing the negative logarithmic likelihood function, we can calculate the partial derivative of this hyperparameter. Subsequently, the conjugate gradient method can be applied to minimize this partial derivative, thereby attaining the optimal hyperparameter.

2.6. Evaluating metrics

The root mean square error (RMSE) and prediction error variance (PEV) are selected to evaluate the performance of the models.

$$\mathrm{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \widehat{y}_i)^2} \tag{14}$$

where, m is the number of samples; i is the index of a sample; y_i is the observation of surface torque at a certain depth; $\widehat{y}_i$ is the prediction of surface torque at a certain depth.

$$\mathrm{PEV} = \frac{1}{m}\sum_{i=1}^{m}(e_i - \widehat{e}_i)^2 \tag{15}$$

where, $e_i = y_i - \widehat{y}_i$ is the prediction error; $\widehat{e}_i = \frac{1}{m}\sum_{i=1}^{m} e_i$ is the mean of the prediction error. RMSE indicate the accuracy of models, and PEV indicate the stability of prediction error.

3. Experiments and results

3.1. Dataset

The raw data used in this paper was collected from two ultra-deep and horizontal wells in a field in China. The collected data is divided into static and dynamic sections. Static data includes drilling logs, wellbore structure, bottom hole assembly, bit selection, etc. Dynamic data is collected by physical sensors on the rig once every 1 s, including time, depth, WOB, hook load, RPM, MFI, surface torque, and so on. Among them, depth, WOB, hook load, RPM, and MFI are the input features of the model, and surface torque is the output of the model. The description of features is shown in Table 1.

In this paper, we focus on torque prediction in rotary drilling. Firstly, we extract the data in rotary drilling condition from the mud logging data, following a specific extraction rule.

$$\mathrm{WOB} > 0 \text{ and } \mathrm{RPM} > 0 \text{ and } \mathrm{MFI} > 0 \tag{16}$$

Then, the extracted data are normalized according to Eq. (17),

$$x_i' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{17}$$

where, x_i' is the normalized value in row i of a certain column in the data; x_i is the value in row i of the column before normalization; $x_{\min}$ is the minimum value of the column; $x_{\max}$ is the maximum value of the column.

Next, calculate the Euclidean distance and cosine similarity between the normalized query sample and the database samples according to Eqs. (1) and (2), and sort them according to the similarity to extract the samples with high similarity as the training set. Finally, build a local model. The workflow is shown in Fig. 5.

3.2. Effects of local modeling data sample numbers and regression methods

The number of training data points for the local model is critical to predicting performance. A large number of training data may incorporate dissimilar samples that could mislead the model training, while a small number of training data may decrease the prediction accuracy of the local model (Zhao et al., 2010). To explore the influence of different training samples on prediction, 19425 samples from the drilled hole with a length of 3505 m are used as historical data, and SVM, GPR, and LR are used for the experiment.

It can be seen from Fig. 6 that with the increase in training sample number, the RMSE of the three models first decreases and then increases, with the optimal number of train data points in this well section being 240. Compared with the other two models, SVM has the best prediction performance.

The prediction and observation are compared in Fig. 7. It is

Table 1
Description of features.

No.	Feature name	Unit	Data range
1	Depth	m	0–8500
2	WOB	kN	0–300
3	Hook load	kN	300–3000
4	RPM	r/min	0–105
5	MFI	L/s	0–73

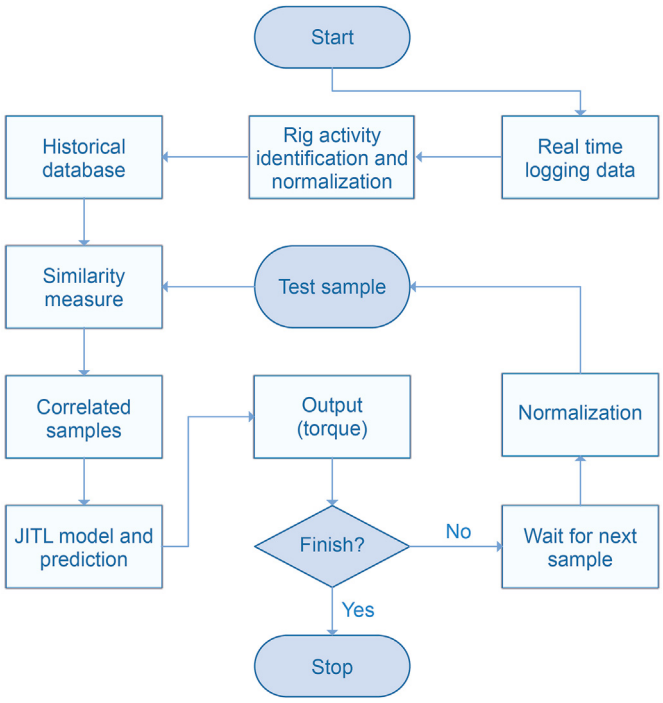

Fig. 5. Workflow of torque prediction with JITL.

shown that with the increase of the train data number, the prediction of the model gradually shifts downward. This is caused by the larger training set containing more shallow formation data that is dissimilar to prediction features.

Considering the influence of the different numbers of samples in the historical database, five experiments were conducted on different well sections. The experimental results are shown in Table 2.

It is shown by Table 2 that the number of optimal train data generally increases with the increase in database sample, but the minimum cosine similarity and the maximum Euclidean distance are relatively stable, which are 0.9998 and 0.1 respectively. To reduce the workload of optimal train data fine-tuning in real-time prediction, fixed thresholds of the cosine similarity and the Euclidean distance are used to select the train data for the local model, which are 0.9998 and 0.1 respectively.

3.3. Performance test of the sliding window method and the JITL method

To verify the performance of the JITL method, the JITL framework and the SW framework of the three machine learning methods are compared respectively. The experimental results are shown in Table 3 and Fig. 8.

As shown in Tables 3 and it is obvious that the JITL model has better prediction performance than the SW models, the RMSE of SVM, GPR, and LR with the JITL framework are 12.65%, 13.61%, and 14.32% lower than the SW method. In addition, the JITL method is more stable than the SW method, with PEV decreasing by 11.06%, 13.74%, and 25.23% for the three local machine learning models. Moreover, the JITL model will require more training time than the SW model since each test sample must be compared to every sample in the historical database to determine similarity. To prevent the training time from growing as the number of database samples rises, we can categorize the samples in the historical database.

It can be seen from Fig. 8 that at 3505.1 m, the torque increases significantly, and the JITL model can still maintain high accuracy while the SW model cannot. This indicates that the JITL model's predictive capability in rapidly changing down-hole conditions is notably superior to that of the SW model.

3.4. Performance test of different prediction distances

As drilling advances, the model should reliably forecast surface torque ahead of the drill bit. Therefore, we employed SVM for simulating and predicting drilling outcomes at depths of 3100 and 3400 m. The outcomes are detailed in Table 4 and Fig. 9, indicating the method's efficacy in accurately predicting within the next 40 m. However, as the prediction distance extends, the error incrementally rises, attributable to variations in stratigraphic conditions.

Therefore, by updating predictions at specified intervals ideally less than 40 m and continuously integrating real-time data from the field into the historical database, we can ensure the accuracy of the height.

3.5. Comparison of real-time prediction of the SW method and the JITL method

To test the accuracy and robustness of the JITL method in real-time drilling operations, we utilized SVM as an example to

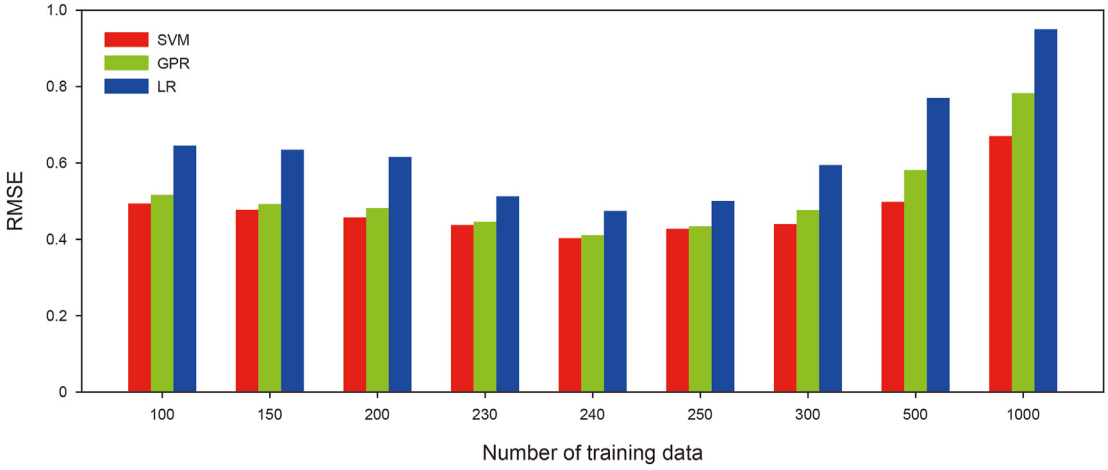

Fig. 6. RMSE of different numbers of train data points and different regression methods.

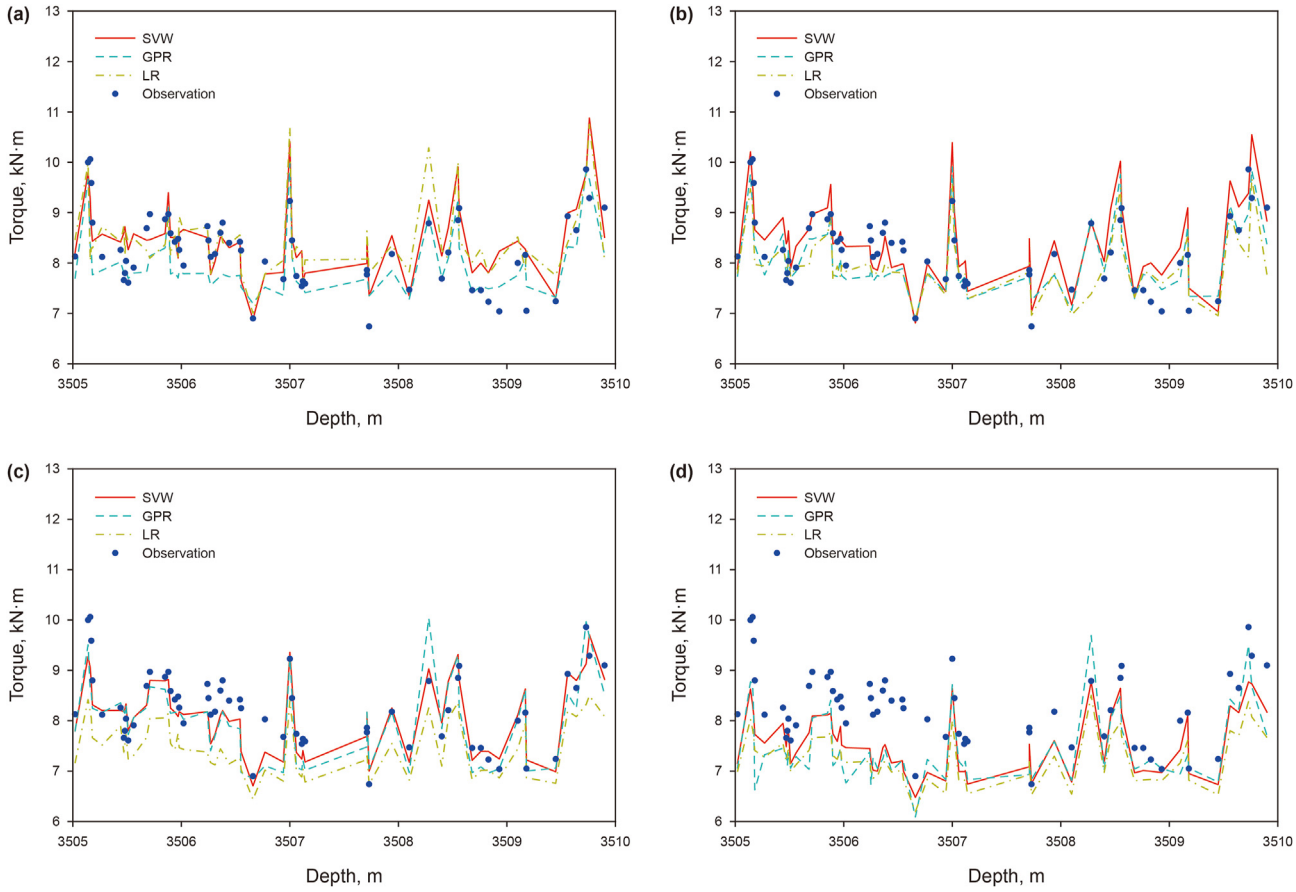

Fig. 7. Predictions of different numbers of train samples and different regression methods. (**a**) the number of train samples is 100; (**b**) the number of train samples is 240; (**c**) the number of train samples is 500; (**d**) the number of train samples is 1000.

Table 2
Experimental results in different well sections.

Number of samples in the historical database	2833	7786	10313	19452	26652	32042
The optimal number of train data for local model	18	25	105	240	172	288
Minimum cosine similarity	0.99986	0.99984	0.99977	0.99979	0.99985	0.99979
Maximum Euclidean distance	0.091	0.093	0.089	0.102	0.096	0.099

Table 3
RMSE, PEV and training time of the SW model and the JITL method.

No.	Method	RMSE	PEV	Training time, s
1	JITL SVM	**0.4023**	**0.1646**	13.71
2	SW SVM	0.4606	0.1908	0.33
3	JITL GPR	0.4107	0.1702	21.66
4	SW GPR	0.4777	0.1973	6.41
5	JITL LR	0.4739	0.2036	9.02
6	SW LR	0.5531	0.2671	**0.02**

predict the drilling torque in the next 40 m. Additionally, new field data was incorporated into the historical database every 40 m to replicate the ongoing drilling process. Table 5 and Fig. 10 display the entire well's (400–8500 m) experimentally compared results between the JITL method and the SW method.

Table 5 illustrates that the JITL method outperforms the SW method, showing a 23.03% decrease in RMSE and a 25.75% decrease in PEV, with sustained accuracy and robustness. As depicted in Fig. 10, the JITL method exhibits greater reliability and accuracy compared to the SW method. However, a small fraction of predictions show significant errors (Fig. 10(a)), potentially attributed to insufficient similar data.

4. Applications and discussion

4.1. Drilling parameters optimization

During drilling, all parameter combinations within a reasonable range are taken as test features and predicted, and the predictions can guide the optimization of drilling parameters and assist field decision-making. SVM is used as the modeling algorithm in this experiment to simulate the parameter optimization of the next step when drilling to 8200 m. Taking WOB & RPM and WOB & MFI as examples, we predict the test samples under different parameter combinations, and the feedback results are shown in Fig. 11.

On the one hand, Fig. 11 shows that torque grows in proportion to WOB, which is in line with field cognition and physical principles. On the other hand, since the drilling has now reached the horizontal part and the PDM drill is providing RPM, which makes RPM positively linked with MFI, the impact of RPM (Fig. 11(a)) and

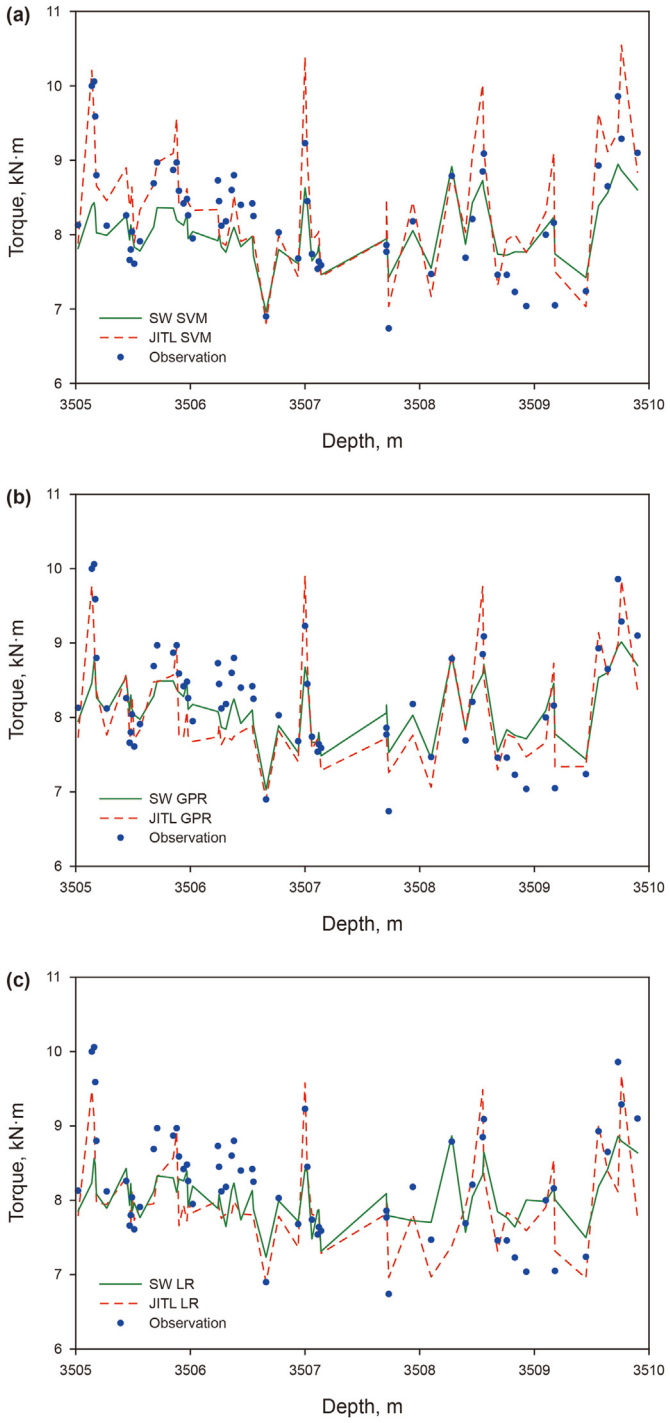

Fig. 8. Comparison between the SW method and the JITL method of different machine learning algorithms.

Table 4
RMSE, PEV of different prediction distances.

Test set, m	Prediction distance, m	RMSE	PEV
3100–3180	10	0.8093	0.3131
	40	0.8264	0.6264
	80	1.1085	0.7662
3400–3480	10	0.5667	0.3176
	40	0.6972	0.7057
	80	1.1275	0.9471

MFI (Fig. 11(b)) on torque is essentially the same. With the torque tolerance of the top drive as 20 kN·m, the parameter combinations corresponding to the red squares are reminded that the surface torque will be too high and should be avoided. The current drilling parameter combination shown by the black pentagram in Fig. 11(b) is located in this area. Therefore, it is recommended to reduce WOB and increase MFI at this time to prevent high torque generation and ensure drilling efficiency.

4.2. Anomaly detection

The confidence interval provided by GPR helps detect anomalies (Pang et al., 2014, 2017). Based on the idea of the digital twin, the GPR model trained using normal drilling data simulates the virtual drilling system operating normally. The predicted confidence interval encompasses all potential torque values within normal operations. If the observed torque surpasses the GPR model's confidence interval and persists for a specific duration, it indicates a high probability of an abnormality in the actual system.

The samples before drilling to 4660 m are used as the database to predict the surface torque of the section to be drilled. The observed surface torque will be generated successively during the drilling process. It can be seen from Fig. 12 that at 4702 m, the torque suddenly increases and continues to exceed the confidence interval. Through checking the drilling log of the well, the top drive was stopped at 4702.84 m, and the drilling tool was stuck. The results of the analysis show that this method is effective for anomaly detection.

4.3. Discussion

Our method was confirmed to be effective for real-time drilling torque prediction, which helps optimize drilling parameters to avoid high torque and detect pipe sticking trends. Due to filtering modeling data through similarity, our method is more accurate than the traditional sliding window method. In real-time prediction testing of the entire well, our method reduces RMSE by 23.03% and PEV by 25.75% compared with the sliding window method.

The data-driven drilling torque prediction models can be divided into the following two types: offline models and online models. The BP-LSTM network (Song et al., 2022), which uses huge data from more than a dozen wells in the same area to build a model and predict the torque of other wells, is the representative offline model. However, because the offline model cannot adapt itself to the changing down-hole environment (formation property, bit wear, etc.) (Alali et al., 2021), its practicability is restricted. Online models have good adaptability and can perform real-time correction according to changes in the down-hole environment, upon which our study was based. The most commonly used online prediction method for drilling torque is the sliding window method (Marquez, 2021), which can dynamically extract drilling data within the current depth window for modeling and prediction. However, the data from a show window restricts the temporary model's capacity to predict during sliding. Thus, in the experiment, the sliding window method and our model were compared in Sections 3.3 and 3.5. The comparison in Table 3 and the error distribution in Fig. 10(b) demonstrate that our method is more effective.

Using similar drilling parameters generally leads to comparable responses in drilling systems. Our approach involves dynamically selecting historical data samples that closely resemble the test sample for modeling. This enables us to address significant changes in parameters such as WOB or RPM by identifying historical data with similar WOB or RPM. This method is particularly effective in situations where drilling parameters experience sudden and drastic

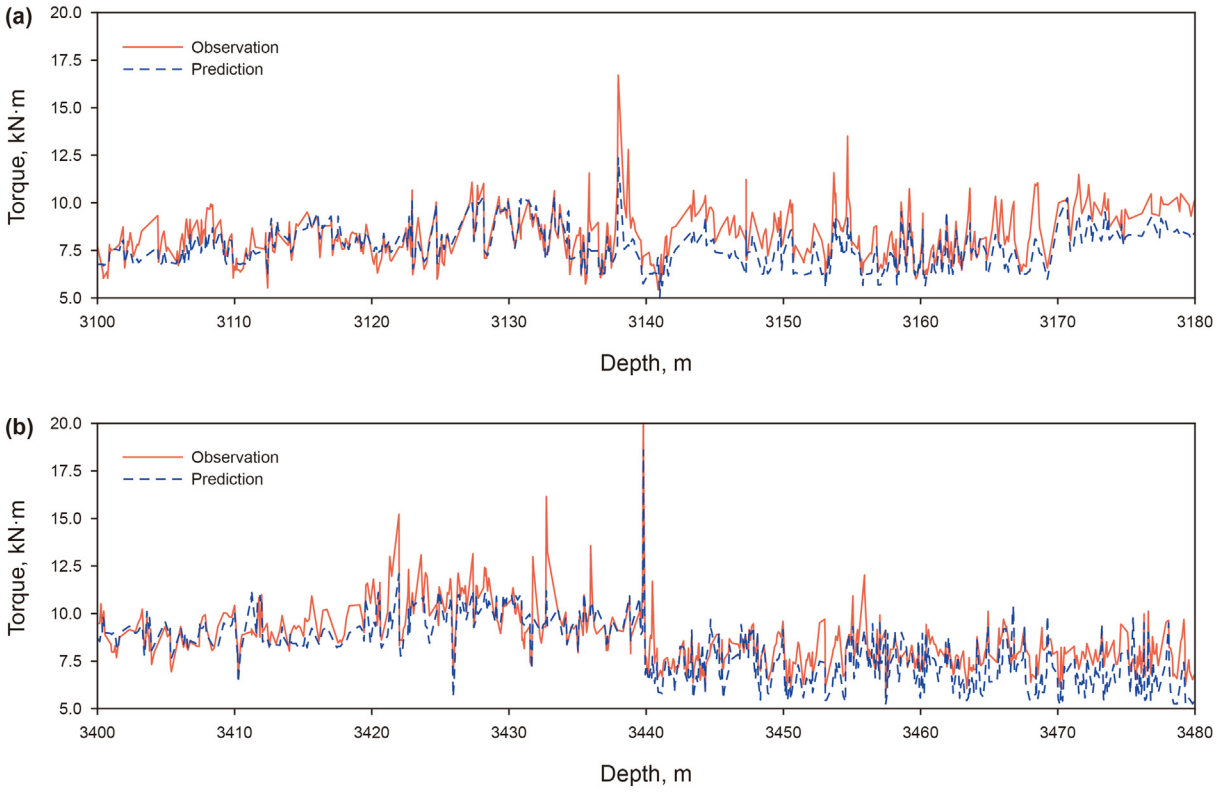

Fig. 9. Prediction performance test of different prediction distances. (**a**) from 3100 to 3180 m; (**b**) from 3400 to 3480 m.

Table 5
RMSE and PEV of different methods in real-time prediction.

No.	Method	RMSE	PEV
1	JITL	1.2445	1.9281
2	SW	1.6169	2.5968

fluctuations, as shown in Fig. 8 where the depth is at 3505.1 m. In contrast, the traditional sliding window method has limited prediction accuracy as it only considers data within a fixed display window, potentially missing important similar samples for effective modeling.

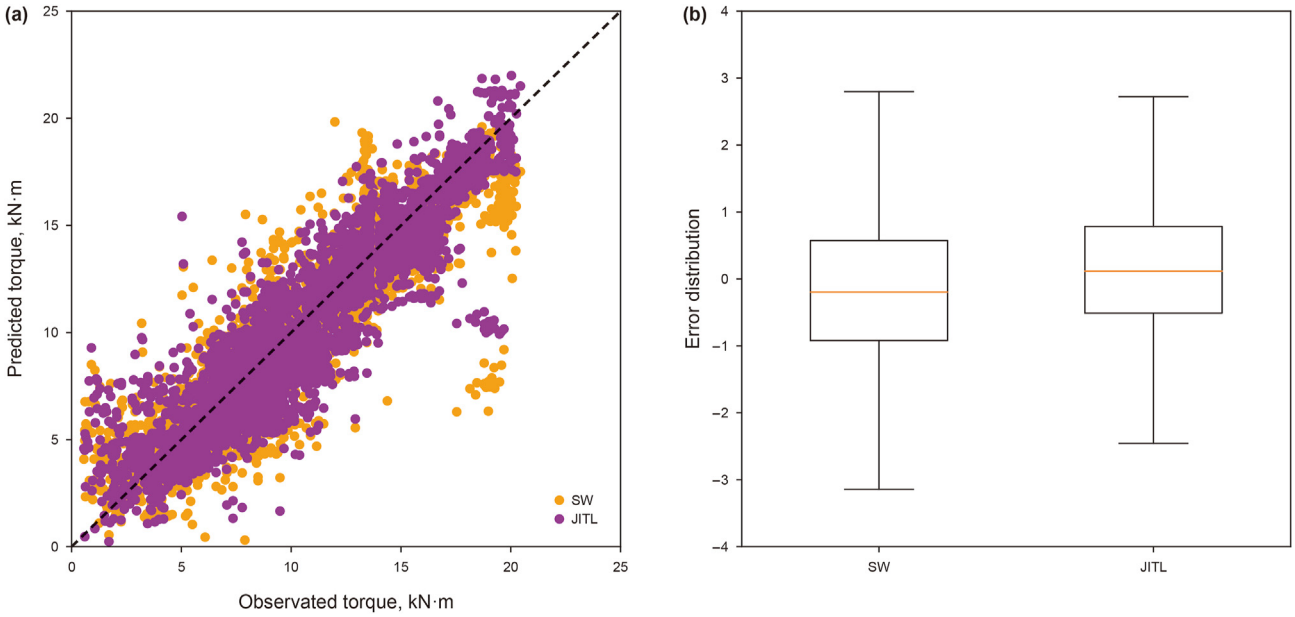

Fig. 10. Comparison between different methods in real-time prediction. (**a**) scatter diagram of prediction and observation; (**b**) box plot of the error distribution.

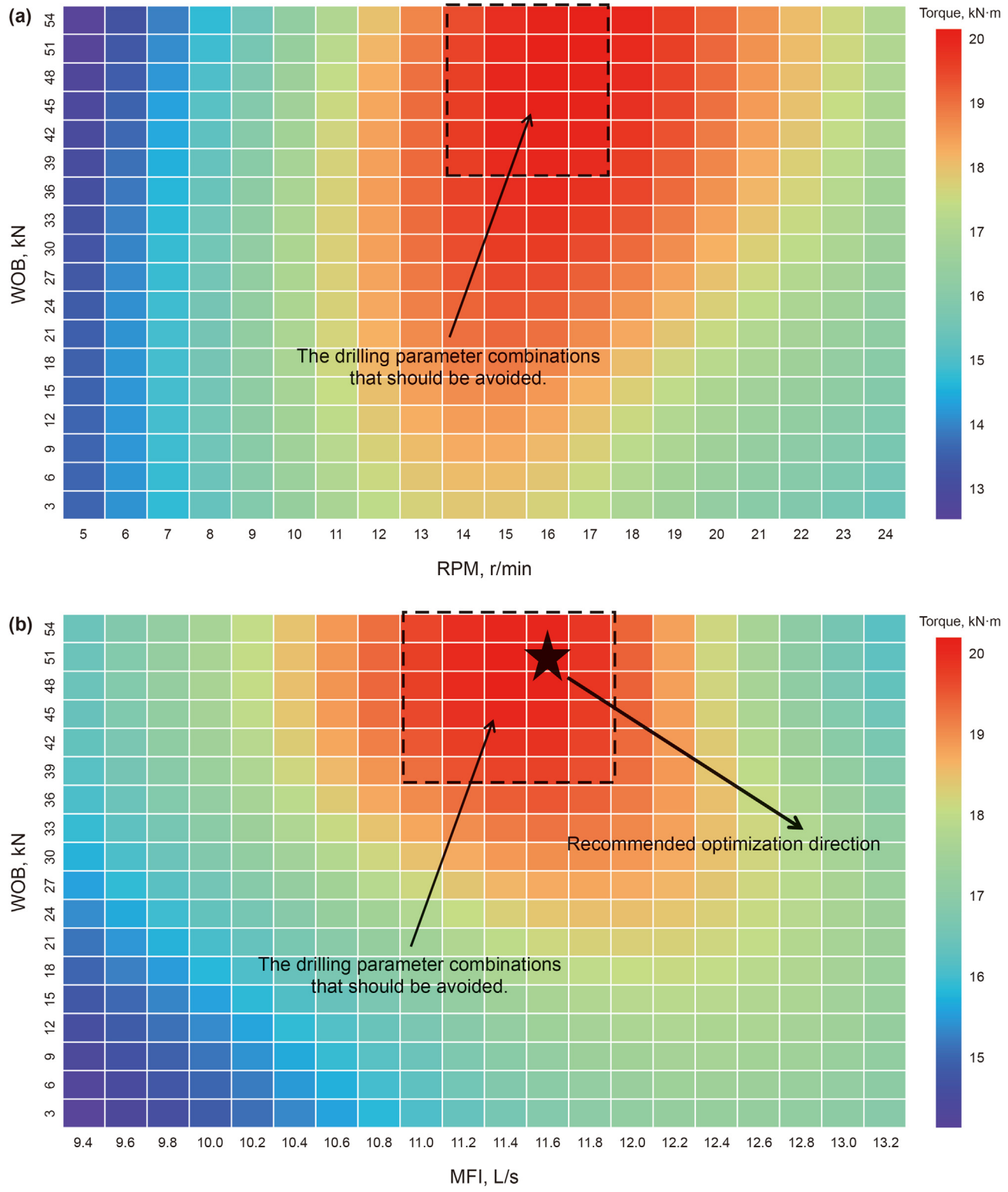

Fig. 11. Heat map of WOB & RPM (**a**) and WOB & MFI (**b**) optimization. The current drilling parameters are: The depth is 8200 m, the WOB is 50 kN, the RPM is 16 r/min, the MFI is 11.6 L/s, and the hook load is 2783 kN.

Therefore, ensuring similarity between the modeling data and the test sample is essential for accurate predictions. Furthermore, while the choice of machine learning algorithms influences predictive performance, it is not a significant factor, as shown in Fig. 7. Thus, instead of solely depending on the most accurate SVM, selecting suitable machine learning algorithms based on their specific purpose is recommended. For example, anomaly detection can benefit from utilizing the confidence interval provided by GPR.

Our approach is limited in two aspects. One is that the prediction accuracy will be impacted if there aren't enough similar

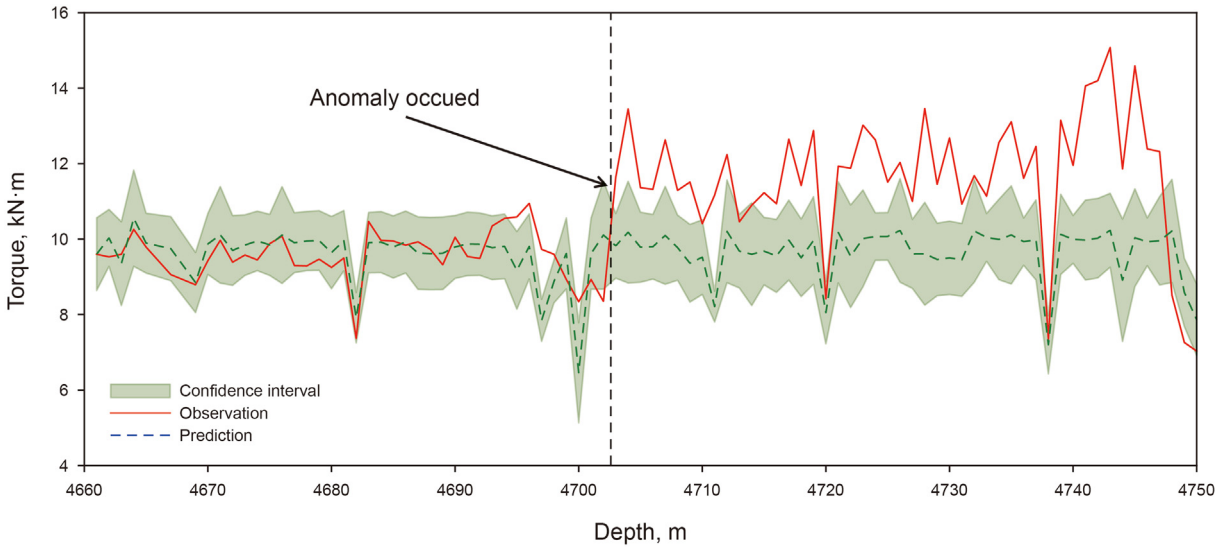

Fig. 12. Diagram of anomaly detection.

samples in the historical database. A shortage of training data will lead to an increase in inaccuracy, as seen by the results in Fig. 6. The other one is slow prediction speed, as shown by the findings in Table 3. Since, in order to assess similarity, each test sample needs to be compared to every sample in the historical database. To overcome these limitations, research will be done in the future on combining off-set well data with target well data using multi-source information fusion to enrich the historical database and improve prediction performance. In addition, we will investigate the efficient classification of databases, including the relative size of drilling parameters, formation information, etc., to stop the prediction time from increasing as the number of database samples increases.

5. Conclusions

In this work, a local modeling and prediction method of surface torque while drilling based on JITL is proposed. This paper applies JITL technology to the drilling process for the first time to realize the dynamic prediction of surface torque ahead of the drill bit. The prediction model can guide the optimization of drilling parameters, and quickly detect torque anomalies by comparing the predicted trend with the actual value. The main points in this study are as follows:

(1) Compared with the sliding window model, JITL has higher accuracy and better robustness. The RMSE of SVM, GPR, and LR with the JITL framework are 12.65%, 13.61%, and 14.32% lower than the SW method, and the PEV decreased by 11.06%, 13.74%, and 25.23%. In real-time prediction testing of the entire well, the JITL method reduces RMSE by 23.03% and PEV by 25.75% compared with the sliding window method.

(2) In the framework, SVM performs best and is recommended for parameter optimization, GPR allows for confidence interval estimation useful for anomaly detection, and LR is ideal for enhancing time efficiency.

(3) This method demonstrates high prediction accuracy within a 40-m range. It is advisable to update the predictions at predetermined intervals, ideally less than 40 m, to maintain height accuracy. Real-time data from the field continually feeds into the historical database for improved performance.

CRediT authorship contribution statement

Kan-Kan Bai: Writing – review & editing, Writing – original draft, Visualization, Validation. **Mao Sheng:** Supervision, Resources, Methodology, Conceptualization. **Hong-Bao Zhang:** Writing – original draft, Supervision, Resources, Methodology, Data curation, Conceptualization. **Hong-Hai Fan:** Resources, Methodology, Conceptualization. **Shao-Wei Pan:** Writing – original draft, Data curation.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors gratefully acknowledge the financial support from the Natural Science Foundation of China (Grant numbers: U23B6010 and 52122401).

References

Alali, A.M., Abughaban, M.F., Aman, B.M., Ravela, S., 2021. Hybrid data driven drilling and rate of penetration optimization. J. Petrol. Sci. Eng. 200, 108075. https://doi.org/10.1016/j.petrol.2020.108075.

Awad, M., Khanna, R., Awad, M., Khanna, R., 2015. Support vector regression. Efficient learning machines: Theories, concepts, and applications for engineers and system designers 67–80. https://doi.org/10.1007/978-1-4302-5990-9_4.

Bai, K.K., Fan, H.H., Zhang, H.B., Zhou, F., Tao, X.G., 2022. Real time torque and drag analysis by combining of physical model and machine learning method. In: SPE/AAPG/SEG Unconventional Resources Technology Conference, Houston, Texas (P. D031S064R005). URTEC-3723045-MS. https://doi.org/10.15530/urtec-2022-3723045.

Barbosa, L.F.F., Nascimento, A., Mathias, M.H., de-Carvalho-Jr, J.A., 2019. Machine learning methods applied to drilling rate of penetration prediction and optimization-A review. J. Petrol. Sci. Eng. 183, 106332. https://doi.org/10.1016/j.petrol.2019.106332.

Chen, M.H., Yu, C.H., Gao, J.L., Yu, K., Lin, S., Guo, G.D., Li, J., 2022. Quantum algorithm for Gaussian process regression. Phys. Rev. 106 (1), 012406. https://doi.org/10.1103/PhysRevA.106.012406.

Cybenko, G., 1996. Just-in-time learning and estimation. Identification, Adaptation, Learning 423–434. https://doi.org/10.1007/978-3-662-03295-4_11.

Elmgerbi, A., Chuykov, E., Thonhauser, G., Nascimento, A., 2022. Machine learning techniques application for real-time drilling hydraulic optimization. In: International Petroleum Technology Conference. OnePetro. https://doi.org/10.2523/

IPTC-22662-MS.
Erge, O., van-Oort, E., 2021. Hybrid physics-based and data-driven modeling for improved standpipe pressure prediction. In: SPE/IADC International Drilling Conference and Exhibition. OnePetro. https://doi.org/10.2118/204094-MS.
Erge, O., van-Oort, E., 2022. Combining physics-based and data-driven modeling in well construction: hybrid fluid dynamics modeling. J. Nat. Gas Sci. Eng. 97, 104348. https://doi.org/10.1016/j.jngse.2021.104348.
Galván, I.M., Valls, J.M., García, M., Isasi, P., 2011. A lazy learning approach for building classification models. Int. J. Intell. Syst. 26 (8), 773–786. https://doi.org/10.1002/int.20493.
Gao, D.L., 1994. Prediction and Control of Wellbore Trajectory. China University of Petroleum Press, Dongying (in Chinese).
Guo, F., Xie, R., Huang, B., 2020. A deep learning just-in-time modeling approach for soft sensor based on variational autoencoder. Chemometr. Intell. Lab. Syst. 197, 103922. https://doi.org/10.1016/j.chemolab.2019.103922.
Hegde, C., Daigle, H., Millwater, H., Gray, K., 2017. 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. https://doi.org/10.1016/j.petrol.2017.09.020.
Ho, H.S., 1988. An improved modeling program for computing the torque and drag in directional and deep wells. In: SPE Annual Technical Conference and Exhibition. Houston, Texas. https://doi.org/10.2118/18047-MS.
Huang, W.J., Gao, D.L., 2019. Combined effects of wellbore curvature, connector, and friction force on tubular buckling behaviors. SPE J. 24, 2083–2096. https://doi.org/10.2118/195680-PA.
Huang, W.J., Gao, D.L., 2021. Local-integral coupling model of tubular strings with connectors and its application in periodic sticking analyses. SPE J. 26 (6), 3410–3423. https://doi.org/10.2118/204463-PA.
Jiang, X.Y., Ge, Z.Q., 2022. Improving the performance of just-in-time learning-based soft sensor through data augmentation. IEEE Trans. Ind. Electron. 69 (12), 13716–13726. https://doi.org/10.1109/TIE.2021.3139194.
Jin, H.P., Pan, B., Chen, X.G., Qian, B., 2019. Ensemble just-in-time learning framework through evolutionary multi-objective optimization for soft sensor development of nonlinear industrial processes. Chemometr. Intell. Lab. Syst. 184, 153–166. https://doi.org/10.1016/j.chemolab.2018.12.002.
Jin, H.P., Li, J.G., Wang, M., Qian, B., Yang, B., Li, Z., Shi, L.X., 2020. Ensemble just-in-time learning-based soft sensor for mooney viscosity prediction in an industrial rubber mixing process. Adv. Polym. Technol. 2020, 1–14. https://doi.org/10.1155/2020/6575326.
Johancsik, C.A., Friesen, D.B., Dawson, R., 1984. Torque and drag in directional wells-prediction and measurement. J Pet Technol 36, 987–992. https://doi.org/10.2118/11380-PA.
Kirişci, M., 2023. New cosine similarity and distance measures for Fermatean fuzzy sets and TOPSIS approach. Knowl. Inf. Syst. 65 (2), 855–868. https://doi.org/10.1007/s10115-022-01776-4.
Li, X.C., Mba, D., Lin, T.R., Yang, Y.J., Loukopoulos, P., 2021. Just-in-time learning based probabilistic gradient boosting tree for valve failure prognostics. Mech. Syst. Signal Process. 150, 107253. https://doi.org/10.1016/j.ymssp.2020.107253.
Liu, D.H., Chen, X.H., Peng, D., 2019. Some cosine similarity measures and distance measures between q-rung orthopair fuzzy sets. Int. J. Intell. Syst. 34 (7), 1572–1587. https://doi.org/10.1002/int.22108.
Liu, X.Y., Zhou, L., Chen, X.H., Li, J.Y., 2020. Lithofacies identification using support vector machine based on local deep multi-kernel learning. Petrol. Sci. 17, 954–966. https://doi.org/10.1007/s12182-020-00474-6.
Liu, Y., Gao, Z., 2015. Enhanced just-in-time modelling for online quality prediction in BF ironmaking. Ironmak. Steelmak. 42 (5), 321–330. https://doi.org/10.1179/1743281214Y.0000000229.
Marquez, F.J., 2021. Drilling optimization applying machine learning regression algorithms. In: Offshore Technology Conference. OnePetro. https://doi.org/10.4043/30934-MS.
McSpadden, A., Newman, K., 2002. Development of a stiff-string forces model for coiled tubing. In: SPE/ICoTA Coiled Tubing Conference and Exhibition. Houston, Texas. https://doi.org/10.2118/74831-MS.
Menand, S., Sellami, H., Tijani, M., Stab, O., Dupuis, D., Simon, C., 2006. Advancements in 3D drillstring mechanics: from the bit to the topdrive. In: IADC/SPE Drilling Conference. Miami, Florida, USA. SPE-98965-MS. https://doi.org/10.2118/98965-MS.
Mitchell, R.F., Bjorset, A., Grindhaug, G., 2015. Drillstring analysis with a discrete torque/drag model. SPE Drill. Complet. 30 (1), 5–16. https://doi.org/10.2118/163477-PA.
Najjarpour, M., Jalalifar, H., Norouzi-Apourvari, S., 2020. The effect of formation thickness on the performance of deterministic and machine learning models for rate of penetration management in inclined and horizontal wells. J. Petrol. Sci. Eng. 191, 107160. https://doi.org/10.1016/j.petrol.2020.107160.
Narula, S.C., Wellington, J.F., 2007. Multiple criteria linear regression. Eur. J. Oper. Res. 181 (2), 767–772. https://doi.org/10.1016/j.ejor.2006.06.026.
Negara, A., Saad, B., 2020. Combining insight from physics-based models into data-driven model for predicting drilling rate of penetration. In: International Petroleum Technology Conference. OnePetro. https://doi.org/10.2523/IPTC-20090-MS.
Olatunji, O.O., Micheal, O., 2017. Prediction of sand production from oil and gas reservoirs in the Niger Delta using support vector machines SVMs: a binary classification approach. In: SPE Nigeria Annual International Conference and Exhibition. OnePetro. https://doi.org/10.2118/189118-MS.
Oyedere, M., Gray, K., 2020. Torque-on-bit (TOB) prediction and optimization using machine learning algorithms. J. Nat. Gas Sci. Eng. 84, 103623. http://www.elsevier.com/locate/jngse.
Pang, J.Y., Liu, D.T., Liao, H.T., Peng, Y., Peng, X.Y., 2014. Anomaly detection based on data stream monitoring and prediction with improved Gaussian process regression algorithm. In: 2014 International Conference on Prognostics and Health Management. IEEE, pp. 1–7. https://doi.org/10.1109/ICPHM.2014.7036394.
Pang, J.Y., Liu, D.T., Peng, Y., Peng, X.Y., 2017. Anomaly detection based on uncertainty fusion for univariate monitoring series. Measurement 95, 280–292. https://doi.org/10.1016/j.measurement.2016.10.031.
Qi, Q.L., Tao, F., Hu, T.L., Anwer, N., Liu, A., Wei, Y.L., Wang, L.H., Nee, A.Y.C., 2021. Enabling technologies and tools for digital twin. J. Manuf. Syst. 58, 3–21. https://doi.org/10.1016/j.jmsy.2019.10.001.
Ren, C.J., Huang, W.J., Gao, D.L., 2022. Predicting rate of penetration of horizontal drilling by combining physical model with machine learning method in the China Jimusar Oil Field. SPE J. 1–24. https://doi.org/10.3390/en15093037.
Samuel, R., Huang, W.J., 2020. Dynamic torque and drag model. In: SPE Annual Technical Conference and Exhibition. OnePetro. https://doi.org/10.2118/201629-MS.
Sottet, J.S., Pruski, C., 2023. Data and model harmonization research challenges in a nation wide digital twin. Systems 11 (2), 99. https://doi.org/10.3390/systems11020099.
Song, X.Z., Zhu, S., Li, G.S., Zeng, Y.J., Guo, H.J., Hu, Z.J., 2022. Prediction of hook load and rotary drive torque during well-drilling using a BP-LSTM network. Journal of China University of Petroleum (Edition of Natural Science) 46 (2), 76–84. https://doi.org/10.3969/j.issn.1673-5005.2022.02.007 (in Chinese).
Tao, F., Xiao, B., Qi, Q.L., Cheng, J.F., Ji, P., 2022. Digital twin modeling. J. Manuf. Syst. 64, 372–389. https://doi.org/10.1016/j.jmsy.2022.06.015.
Teale, R., 1965. The concept of specific energy in rock drilling. Int. J. Rock Mech. Min. Sci. Geomech. Abstracts 2 (1), 57–73. https://doi.org/10.1016/0148-9062(65)90022-7.
Williams, C.K., Rasmussen, C.E., 2006. Gaussian Processes for Machine Learning. MIT Press Cambridge, Cambridge, MA, USA.
Wu, J., Juvkam-Wold, H.C., 1995. The effect of wellbore curvature on tubular buckling and lockup. ASME. J. Energy Resour. Technol 117 (3), 214–218. https://doi.org/10.1115/1.2835343.
Yang, R.Y., Qin, X.Z., Liu, W., Huang, Z.W., Shi, Y., Pang, Z.Y., Zhang, Y.Q., Li, J.B., Wang, T.Y., 2022. A physics-constrained data-driven workflow for predicting Coalbed methane well production using artificial neural network. SPE J. 27 (3), 1531–1552. https://doi.org/10.2118/205903-PA.
Yin, S., Xie, X.C., Sun, W., 2016. A nonlinear process monitoring approach with locally weighted learning of available data. IEEE Trans. Ind. Electron. 64 (2), 1507–1516. https://doi.org/10.1109/TIE.2016.2612161.
Zhang, J., Rasmussen, E.M., 2001. Developing a new similarity measure from two different perspectives. Inf. Process. Manag. 37 (2), 279–294. https://doi.org/10.1016/S0306-4573(00)00027-3.
Zhang, Y., Jin, H.P., Liu, H.P., Yang, B., Dong, S.L., 2022. Deep semi-supervised just-in-time learning based soft sensor for mooney viscosity estimation in industrial rubber mixing process. Polymers 14 (5), 1018. https://doi.org/10.3390/polym14051018.
Zhao, Y.P., Sun, J.G., Zou, X.Q., 2010. Reducing samples for accelerating multikernel semiparametric support vector regression. Expert Syst. Appl. 37 (6), 4519–4525. https://doi.org/10.1016/j.eswa.2009.12.058.