Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería

Tech Science Press

https://doi.org/10.32604/rimni.2026.083039

ARTICLE

Intelligent Well Trajectory Design Based on Optimal Offset Well and RL-PSO

Jiaqi Zhang[1,*], Guo Wang[1], Song Deng[2], Duo Yuan[1] and Xiaoli Hu[1]

[1]Digital & Intelligent Technology R&D Center, Sinopec Research Institute of Petroleum Engineering Co., Ltd., Beijing, China
[2]School of Petroleum and Natural Gas Engineering, Changzhou University, Changzhou, China

*Corresponding Author: Jiaqi Zhang. Email: jiaqizhang0221@outlook.com

Received: 27 March 2026; Accepted: 01 July 2026; Published: 06 November 2026

ABSTRACT: Well trajectory design is a fundamental component of drilling design. Conventional well trajectory parameter selection relies heavily on expert experience, requiring repeated manual adjustment, resulting in low design efficiency. Traditional trajectory optimization algorithms require pre-defined parameter ranges and fail to fully leverage the offset well data. To address these limitations, this paper presents a proof-of-concept intelligent well trajectory design scheme that fully exploits the formation and offset well data by integrating reinforcement learning and optimization theory. An optimal offset well screening strategy and a reinforcement learning-based PSO (RL-PSO) algorithm are proposed to realize automatic trajectory design without manual parameter configuration. First, multiple features including formation similarities are encoded and quantified; the entropy weight method is applied to calculate the weight of each feature, and the optimal offset well is selected through a multi-dimensional comprehensive evaluation system. Second, based on the trajectory profile and parameter range derived from the optimal offset well, reinforcement learning is adopted to dynamically adjust the inertia weight and acceleration coefficients of the PSO algorithm for adaptive trajectory parameter optimization. Experimental results indicate that compared with traditional design methods, the proposed RL-PSO framework, driven by optimal offset well data can generate more economical well trajectories and improve design efficiency under the studied experimental conditions, providing a feasible intelligent solution for automated well trajectory design.

KEYWORDS: Intelligent well trajectory design; optimal offset well selection; reinforcement learning; particle swarm optimization

1 Introduction

Well trajectory design is a fundamental part of drilling design [1], and the scientific rigor and rationality of the design scheme directly determine the safety, cost and duration of drilling operations. As oil and gas exploration and development extend to deep and unconventional formations, geological complexity has increased dramatically, placing higher demands on the precision, automation and intelligent level of well trajectory design. Traditional trajectory design methods fail to fulfill the development needs of modern oil and gas engineering.

In on-site situations, trajectory parameters are predominantly determined by expert experience [2], leading to complex calculations and inconsistent design quality. Studies on complex well trajectory design presented in papers [3,4] demonstrate that engineers must repeatedly adjust trajectory parameters manually, resulting in heavy workloads, susceptibility to human error, and potential safety risks. Besides, the

Copyright © 2027 The Authors. Published by Tech Science Press. This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

expert-experience-dominated design paradigm lacks uniform standards, causing significant variability in scheme quality.

Against this background, the automatic and intelligent design of well trajectory has become an urgent demand in the field of oil and gas drilling engineering. In recent years, researchers have carried out studies on optimal design of the well trajectory [5–8]. Most existing studies preset value ranges for key parameters such as trajectory profile shape, kick-off point position, and dog-leg severity, then establish trajectory equations using mathematical models. Particle swarm optimization, genetic algorithms and other optimization algorithms are employed to solve for optimal well trajectory parameters. The optimization objectives include minimizing segment length and friction torque, while parameter boundary intervals are set as constraint conditions. Although such methods can produce feasible design schemes, they rely on manually predefined parameter bounds and underutilize historical offset well datasets, limiting fully automatic intelligent design. To improve the efficiency of well trajectory design and meet the economic benefits requirements, an intelligent well trajectory design algorithm has become increasingly urgent.

Offset well data constitutes the primary basis for intelligent drilling design [9]. Such data contain information including geological stratification, pressure distribution and complex drilling conditions [10], which can provide references for regional geology and engineering construction in new well design, and serve as important data support to realize intelligent design of well trajectory [11]. Applying machine learning algorithms and data mining techniques to offset well data [5] can transform engineering experience into a knowledge base, thereby enabling a data-driven intelligent design workflow for well trajectories.

To address the deficiencies of existing methods, this study proposes an intelligent well trajectory design method based on optimal offset well selection and an RL-PSO algorithm. It aims to enhance the automation and intelligence well trajectory design and fully exploit the value of offset well data. The main contributions are as follows: character encoding combined with the dynamic time warping (DTW) algorithm [12] is adopted to quantitatively characterize formation information, and a multi-dimensional evaluation system is established for offset well selection; an automatic trajectory design procedure is constructed in which design parameters are automatically extracted from optimal offset wells without manual setting; integrating techniques from reinforcement learning and particle swarm optimization algorithm are applied to get the optimal design parameters for well trajectory design.

2 Reinforcement Learning-Particle Swarm Optimization Algorithm

Reinforcement learning [13], together with supervised learning and unsupervised learning, constitutes the three main learning paradigms of machine learning [14]. It aims at learning through the interaction between an agent and the environment. A reinforcement learning system comprises: agent, environment, state, reward, action, policy, and goal. The agent is the decision-making entity; the environment refers to everything external to the agent that can influence it; the state represents the current status of the agent; the reward is the feedback signal obtained after the agent executes an action; the action denotes the behavior taken by the agent; the policy is the mapping from states to actions; and the goal is to find the optimal policy corresponding to the maximum long-term cumulative reward.

The fusion of reinforcement learning with optimization algorithms is based on the mechanism by which dynamically adjusts algorithmic behavior based on environmental reward and penalty feedback [15]. Specifically, the convergence speed and objective optimization efficiency of the optimization algorithm are defined as the state space of reinforcement learning, while key algorithm parameters serve as actionable strategies. The agent perceives the current search state of the optimizer and adaptively adjusts these parameters, achieving closed-loop collaborative optimization [16].

The overall framework of the RL-PSO algorithm [17] is divided into two layers (Fig. 1): the lower layer is the PSO layer, responsible for searching and iterating well trajectory design parameters; the upper layer is the reinforcement learning layer, which dynamically adjusts the key PSO parameters (ω, c_1, c_2). Through agent–environment interaction, the parameter adjustment strategy is optimized to prevent the algorithm from converging prematurely to local optima. The main steps are as follows: ① Initialize the particle swarm and reinforcement learning agent; ② The PSO algorithm searches for feasible solutions according to the current parameters and calculates the fitness value of each particle; ③ The reinforcement learning agent perceives the current search state of the particle swarm, executes parameter adjustment actions, and updates the parameters; ④ The environment feeds back corresponding rewards based on the optimization performance of the particle swarm after parameter adjustment; ⑤ The agent updates its decision-making policy according to the reward signal; ⑥ Repeat steps ② to ⑤ until the preset convergence condition is met.

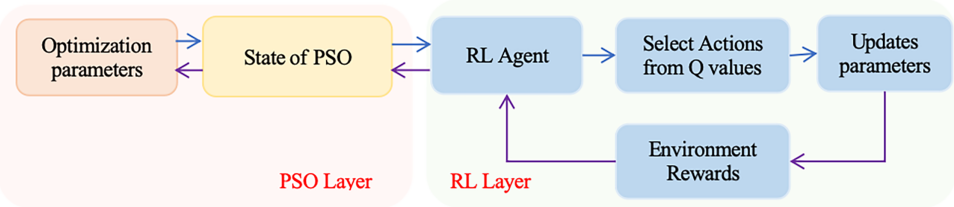


Figure 1: Reinforcement learning-particle swarm optimization algorithm.

3 Multi-Dimensional Comprehensive Evaluation for Offset Well Selection

Optimal offset well screening lays the foundation for intelligent trajectory design by providing reference parameter ranges. Wellhead or target distance is the most commonly used criterion for offset well selection, wells with the closest distances are selected as references intuitively. However, under complex working conditions, single-factor decision-making can hardly meet the accuracy requirements of drilling design, and the nearest well does not always provide the highest reference value [18]. By systematically reviewing relevant literature and incorporating the engineering experience of on-site drilling experts, this paper establishes a multi-dimensional comprehensive evaluation system for offset well selection, adopting the entropy weight method to score and rank candidate wells. Detailed implementation steps are shown in Fig. 2.

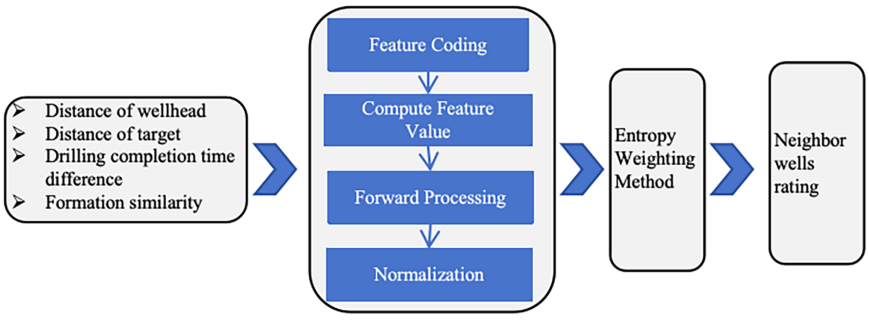


Figure 2: Multi-dimensional comprehensive evaluation system of offset wells.

3.1 Feature Processing of Offset Wells

In this paper, the evaluation dimensions for offset wells encompass three aspects: spatial distance, drilling completion time difference, and formation horizon information. Taking the target well as reference,

each offset well is characterized by four features: wellhead distance, target distance, drilling completion time difference, and formation similarity.

As features vary in dimension and data type, feature encoding and min–max normalization are adopted to unify them into scaled quantitative data [19], which supports subsequent comparison and difference calculation.

The coding and calculation methods for each feature are as follows:

- Wellhead distance and target distance: The horizontal projection displacement of the connecting line between the offset well and the current well is adopted.
- Drilling completion time difference: The number of natural days from the drilling completion date of the selected offset well to the current date.
- Formation similarity: Character coding is first applied to encode formation horizons, and then the distance between codes is calculated as the formation similarity value using the DTW algorithm.

The first three features are calculated using simple numerical methods, and this paper focuses on method for calculating formation similarity. The character encoding rule specifies that text is encoded via alphanumeric combinations. For formation features: uppercase letters denote formation group, digits denote formation member, and lowercase letters denote formation sub-member. Assuming the formation group information of a well is [Pingyuan, Minghuazhen, Guantao, Dongying, Shahejie], and the member information is [1st member of Shahejie, 2nd member of Shahejie, upper submember of 3rd member of Shahejie, lower submember of 3rd member of Shahejie], the character code of this formation horizon is denoted as [P, M, G, D, S1, S2, S3a, S3c]. The formation similarity between two wells is characterized by the minimal alignment distance solved by the DTW algorithm from encoded formation sequences, with Levenshtein distance [20] adopted as the elemental distance metric during DTW calculation. Operation costs are defined as {uppercase letter: 10, digit: 5, lowercase letter: 1}, reflecting the hierarchical importance of formation levels: group-level differences are geologically most significant (cost 10), member-level differences are intermediate (cost 5), and sub-member differences are minor (cost 1).

Given formation information of two wells: #1: [Pingyuan, Minghuazhen, Guantao, Dongying, 1st member of Shahejie, upper submember of 2nd member of Shahejie], #2: [Pingyuan, Minghuazhen, Guantao, Dongying, 1st member of Shahejie, 2nd member of Shahejie, upper submember of 3rd member of Shahejie, lower submember of 3rd member of Shahejie]. Table 1 shows the DTW distance matrix of formation sequence between two wells. We can see that the columns and rows represent formation coding sequence of well No. 1 (#1) and well No. 2 (#2), respectively, the value of each table unit is the distance between single formation coding. The number in the upper-right corner of Table 1 represents the result of the DTW algorithm, which denotes the cumulative distance given by Eq. (1).

Table 1: DTW distance between two wells.

				#2				
#1	**P**	**M**	**G**	**D**	**S1**	**S2**	**S3a**	**S3c**
P	0^{77}	10^{77}	10^{87}	10^{97}	15^{112}	15^{127}	16^{143}	16^{159}
M	10^{77}	0^{77}	10^{77}	10^{87}	15^{102}	15^{117}	16^{133}	16^{145}
G	10^{67}	10^{77}	0^{67}	10^{77}	15^{92}	15^{107}	16^{123}	16^{129}
D	10^{57}	10^{77}	10^{77}	0^{77}	15^{92}	15^{100}	16^{107}	16^{113}
S1	15^{47}	15^{62}	15^{77}	15^{92}	0^{80}	5^{85}	6^{91}	6^{97}
S2a	16^{32}	16^{48}	16^{64}	16^{80}	6^{86}	1^{87}	10^{97}	6^{103}

Note: #1: {P, M, G, D, S1, S2a}; #2: {P, M, G, D, S1, S2, S3a, S3c}.

$$g_{ij} = min\{g_{i-1,j} + d_{ij}, g_{i-1,j-1} + 2d_{ij}, g_{i,j-1} + d_{ij}\} \tag{1}$$

where g_{ij} is the cumulative distance, d_{ij} is Levenshtein distance, the formation similarity between well No. 1 (#1) and well No. 2 (#2) is the upper-right corner value of DTW distance matrix, which is 159.

3.2 Comprehensive Scoring of Offset Wells

Comprehensive scoring for offset wells is conducted based on multi-factor evaluation [21]. Common methods include the analytic hierarchy process, fuzzy comprehensive evaluation, and multi-attribute decision-making, which typically rely on subjective expert judgment for weight assignment, making it difficult to objectively reflect the real contribution of each influencing factor. In this paper, we use entropy weight method to score and rank offset wells. The entropy weight method is an objective weighting method that determines weights according to the dispersion of feature data and avoids direct subjective assignment of weights. This is suitable for the proposed offset-well selection problem because the four features have different scales and information variability. With a solid theoretical foundation, low computational cost, and freedom from subjective bias, the entropy weight method is well suited for field deployment scenarios that prioritize engineering efficiency and transparency. The main steps are as follows:

(1) After calculating and processing the differences of each feature between offset wells and the target well, a standardized feature difference matrix can be constructed as $\mathbf{Y} = (y_{ij})_{n*m}$, where i denotes the ith offset well, j denotes the jth feature, y_{ij} denotes the feature difference value, n denotes the number of offset wells, and m denotes the number of evaluation features.

(2) Convert the standardized feature matrix into a probability matrix by Eq. (2):

$$p_{ij} = \frac{y_{ij}}{\sum_{i=1}^{n} y_{ij}} \tag{2}$$

(3) Then the entropy value of jth feature is $e_j = -k \sum_{i=1}^{n} p_{ij} ln(p_{ij})$, $k = 1/ln(n)$; the diversity coefficient is $d_j = 1 - e_j$.

(4) The entropy weight of jth feature is given by Eq. (3).

$$w_j = \frac{d_j}{\sum_{j=1}^{m} d_j} \tag{3}$$

(5) The comprehensive score of the ith offset well can be got by Eq. (4).

$$A_i = \sum_{j=1}^{m} w_j \dot{y}_{ij} \tag{4}$$

After scoring offset wells by the multi-dimensional comprehensive evaluation system, the optimal offset wells can be determined according to the comprehensive scores, so as to obtain the initial parameters for the well trajectory optimization algorithm and avoid artificial subjective setting.

4 Intelligent Well Trajectory Design Based on RL-PSO

Research on the optimal design of well trajectories has become relatively mature [22]. The general procedure establishes mathematical models with objectives such as minimizing the wellbore length or friction drag, constrained by inclination, azimuth, etc. Machine learning and optimization algorithms

are then employed to achieve optimized well trajectory design parameters. Such methods impose high requirements on the selection of optimization algorithms, and the initialization of parameters relies heavily on expert experience. The proposed intelligent design method eliminates reliance on human experience by automatically deriving design parameters from offset well data. By inputting only the coordinates of wellhead and target of the well to be designed, the well trajectory design scheme can be automatically generated.

Fig. 3 gives the steps of intelligent optimization design algorithm for well trajectory, it mainly consists of three components: offset well selection, recommendation of constraint intervals for the design parameters of well trajectory, and calculation of optimal design parameters.

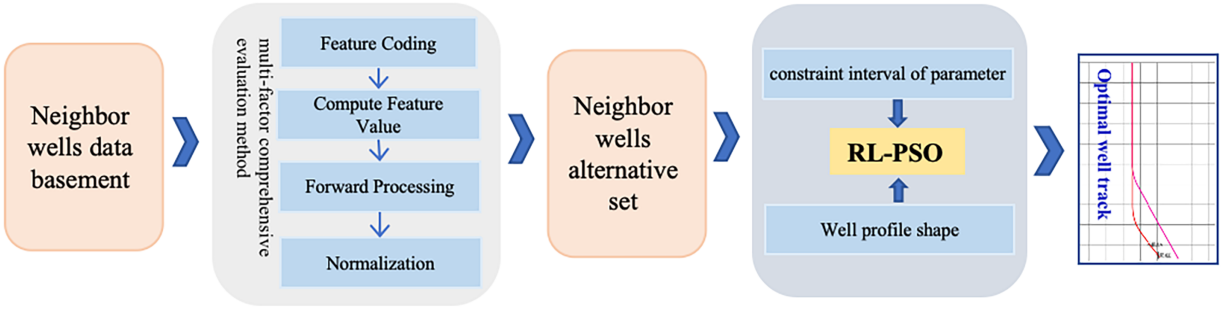


Figure 3: Intelligent well trajectory design flowchart.

4.1 Mathematical Model of Well Trajectory Design

To design a well trajectory automatically, offset wells are first selected based on the multi-dimensional evaluation system to construct an alternative set. The voting method is used to count the most frequently used trajectory profile shape in the set. In case of tied votes, the profile from the highest-scoring offset well is selected as the baseline shape for the target well. When the scores of multiple offset wells are close, the final selection should be checked by performing a ranking perturbation analysis. If the selection frequency of the top K wells is no less than 80% with minor ranking fluctuations, the original selection result is reliable. Combined with the selected trajectory profile shape, the corresponding design scheme can be solved by well trajectory mathematical models such as the cylindrical helix model and the spatial circular arc model.

Due to differences in the mathematical models corresponding to various trajectory profiles, the optimization parameters must be set accordingly. For instance, the optimal solution for a 2D three-segment trajectory can be obtained by only optimizing the kickoff point and dogleg severity, whereas a 2D five-segment trajectory requires five independent optimization parameters. The parameters of all wells are extracted from offset well set, and the upper and lower bounds of the parameter range are defined by the maximum and minimum values of each key parameter, which are used as constraints for the trajectory optimization design model.

Most studies construct optimization models with objectives such as minimizing well trajectory length or reducing friction and torque. Considering both engineering practicality and computational efficiency, this paper takes the minimum well trajectory length as the optimization objective. The single-objective formulation is adopted here for three reasons: (1) minimizing trajectory length is strongly correlated with reducing drilling cost and time, making it a practically meaningful proxy for overall economy; (2) the dogleg severity is bounded within the offset well parameter range, inherently limiting excessive curvature; (3) the single-objective formulation enables efficient closed-form computation within the RL-PSO framework.

Based on the constraint ranges of kickoff point and dogleg severity determined from the offset wells, combined with the mathematical models corresponding to the trajectory profile, a constrained optimization

equation is established. Then with an optimization algorithm, the optimal design parameters can be obtained. Taking a 2D five-segment trajectory (Fig. 4) as an example, the track design involves five independent parameters, including kickoff point, dogleg severity of first build-up section and second build-up section, terminal inclination angle, terminal hold-section length. In this case, according to on-site experience and design specifications, dogleg severity of second build-up section, terminal inclination angle and terminal hold-section length are predefined as fixed values, then the mathematical model can be given by Eq. (5).

$$f(x) = \begin{cases} minL_{SUM}(Z,K) = L_{OZ} + L_{ZA} + L_{AB} + L_{BC} + L_{CT} \\ h_{min} < Z < h_{max} \\ k_{min} < K < k_{max} \end{cases} \tag{5}$$

where L_{SUM} represents the total length of the well trajectory, which is a set of equations with respect to the kickoff point and dogleg severity; Z denotes the kickoff point, and h_{min}, h_{max} represent the minimum and maximum values of the kickoff point, respectively; K denotes the dogleg severity, and k_{min}, k_{max} represent the minimum and maximum values of the dogleg severity, respectively.

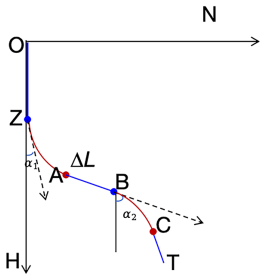


Figure 4: A 2D five-segment well trajectory.

Mathematical model of Eq. (5) can be solved using optimization theory algorithms. Commonly used optimization algorithms include particle swarm optimization, genetic algorithm, simulated annealing, etc. Traditional methods tend to suffer from premature convergence, and their fixed parameters make it difficult to achieve sufficient solution accuracy and stability. The RL-PSO algorithm can dynamically adjust the optimization strategy and key parameters according to the search state, effectively improving global search ability and convergence performance, making it more suitable for well trajectory design. So in this paper, a particle swarm optimization algorithm based on reinforcement learning is adopted.

4.2 Intelligent Well Trajectory Design Based on RL-PSO

Based on the offset well selection method and trajectory mathematical model, the RL-PSO approach is applied to solve the optimization problem, the key parameters for well trajectory design are entirely determined by the algorithm without manual presets, thereby effectively reducing the intervention of human experience. Q-learning reinforcement learning is employed to solve the trajectory design model to obtain the key parameters (i.e., kickoff point and dogleg severity) of the well to be designed.

4.2.1 RL-PSO Agent

The reinforcement learning agent dynamically adjusts the inertia weight ω and acceleration coefficients c_1, c_2 of the PSO algorithm, improving the optimization accuracy and efficiency of well trajectory design and achieving the goal of minimum footage.

The reinforcement learning agent is designed as follows:

(1) State space design: Footage optimization efficiency S_1, particle swarm convergence speed S_2, and population diversity S_3 are selected as the three features of the state space to construct the state space, which are calculated by Eq. (6).

$$\begin{cases} S_1 = \dfrac{L_{gbest}(0) - L_{gbest}(t)}{L_{gbest}(0)} \\ S_2 = L_{gbest}(t-1) - L_{gbest}(t) \\ S_3 = \dfrac{1}{N}\sum_{i=1}^{N} \|x_i - \bar{x_i}\|^2 \end{cases} \tag{6}$$

where $L_{gbest}(0)$ represents the optimal well footage of the initial population, $L_{gbest}(t)$ denotes the optimal well footage of the population at the tth iteration; N represents the population size; x_i represents the position vector of the ith individual; $\bar{x_i}$ represents the average position of all particles within the population.

(2) Action space design: The specific operations of the inertia weight ω and acceleration coefficients c_1, c_2 in the PSO algorithm are taken as actions, and the action space is a discrete set of parameter adjustment amounts. The value range of the inertia weight ω is [0.1, 0.9], with three adjustment actions set: {increase by 0.1, remain unchanged, and decrease by 0.1}. For the acceleration coefficients c_1 and c_2, their value ranges are both [0.5, 2.5], and three adjustment actions are set: {increase by 0.2, remain unchanged, and decrease by 0.2}. The agent selects the optimal action from the action space according to the current state to adjust parameters dynamically.

(3) The reward function is $R = k_1R_1 + k_2R_2 + k_3R_3$, where R_1 is the reward for optimizing footage; R_2 is the reward for convergence speed; R_3 is the reward for population diversity; k_1, k_2, k_3 are weights for each reward item, the selected weights are based on the priority of minimizing total trajectory length while maintaining convergence stability and population diversity. Specifically, k_1 is assigned the largest value because trajectory length improvement is the primary objective; k_2 and k_3 are smaller auxiliary weights because convergence speed and population diversity are used to improve the optimization process, $k_1 + k_2 + k_3 = 1$.

The three reward items can be calculated by Eqs. (7)–(9).

$$R_1 = \begin{cases} \dfrac{L_{gbest}(t-1) - L_{gbest}(t)}{L_{gbest}(t-1)}, L_{gbest}(t) < L_{gbest}(t-1) \\ -\dfrac{L_{gbest}(t) - L_{gbest}(t-1)}{L_{gbest}(t-1)}, L_{gbest}(t) \geq L_{gbest}(t-1) \end{cases} \tag{7}$$

$$R_2 = \begin{cases} S_2, S_2 > 0 \\ -0.5, S_2 \leq 0 \end{cases} \tag{8}$$

$$R_3 = \begin{cases} 1 - |S_3 - 0.4|, 0.2 < S_3 < 0.6 \\ 0.1, S_3 \leq 0.2 \ or \ S_3 \geq 0.6 \end{cases} \tag{9}$$

Here we assume that the reasonable range of population diversity is [0.2, 0.6], it is determined empirically based on observed population behavior during preliminary experimental runs. When S_3 fell below 0.2, the particle swarm exhibited premature convergence to local optima, while S_3 exceeding 0.6 resulted in excessively slow convergence due to over-exploration.

4.2.2 Q-Learning Reinforcement Learning Based PSO Algorithm

A particle swarm optimization algorithm based on Q-learning is proposed. Q-learning is a simple and classic reinforcement learning method, it learns by interacting with the environment, trying different actions and recording their rewards in a Q-table, and gradually finds the best action strategy to achieve the optimal goal. The main steps of this algorithm are as follows:

(1) Parameter initialization:

①. Determine the number of particles: set the number of particles in the particle swarm as N;

②. Initialize the position of particles: each particle represents a possible solution, the position of ith particle is $x_i = \{Z_i, K_i\}$, corresponding to randomly generated values of the kickoff point and dogleg severity;

③. Initialize particle velocities: the velocity vector of the ith particle is v_i, which is generated randomly;

④. Initialize individual optimal positions and global optimal position: set the initial position of each particle x_i as its initial optimal individual position $PBest_i$, calculate the fitness value of each particle, i.e., the well trajectory length. The position corresponding to the shortest trajectory length is selected as the global optimal position $GBest$.

⑤. State discretization for Q-table initialization: to implement tabular Q-learning, we adopt a uniform binning discretization strategy to convert continuous states into discrete indices, as follows: the three continuous state variables S1, S2, and S3 are scaled to $[0,1]$; each normalized state variable $(\hat{S1}, \hat{S2}, \hat{S3})$ is discretized using equal-width binning over [0, 1] with five intervals. Therefore, the total number of discrete states is $5 \times 5 \times 5 = 125$. For the action space, three actions are defined for each of ω, c_1, and c_2: increase, maintain, and decrease. Thus, the total number of actions is $3 \times 3 \times 3 = 27$, and the Q-table is defined as a matrix $Q \in \mathbb{R}^{125\times27}$.

⑥. Set reinforcement learning Q-Learning parameters: learning rate η, discount factor δ, and reward weights k_1, k_2, k_3;

⑦. Initialize the Q-table for Q-Learning: the Q-value of taking action a under state s, usually initialized to small random values and updated during the learning process;

(2) The reinforcement learning agent perceives the current state s of the particle swarm, selects the optimal action based on the Q-table and ϵ-greedy strategy, then adjusts ω, c_1, c_2 of the PSO algorithm. The ω has three adjustment actions: {increase by 0.1, remain unchanged, and decrease by 0.1}. c_1 and c_2 have three adjustment actions: {increase by 0.2, remain unchanged, and decrease by 0.2}. ϵ-greedy strategy follows an exponential decay schedule: $\epsilon(t) = \max(\epsilon_{\min}, \epsilon_0 \times \alpha^t)$, where the initial exploration rate $\epsilon_0 = 0.3$ guarantees 30% random exploration at the start, the decay factor $\alpha = 0.995$ gradually suppresses random sampling, and the minimum threshold $\epsilon_{\min} = 0.01$ limits exploration to merely 1% in later iterations for concentrated local exploitation.

(3) Update position and velocity of particles by Eq. (10).

$$\begin{cases} v_i^{t+1} = w \cdot v_i^t + c_1 r_1 (PBest_i - x_i^t) + c_2 r_2 (GBest - x_i^t) \\ x_i^{t+1} = x_i^t + v_i^{t+1} \end{cases} \tag{10}$$

(4) The reward generated by the environment in response to the agent's action is calculated via the reward function Eqs. (7)–(9), based on which the Q-table is updated and the agent's decision-making strategy is optimized. The corresponding update formula is given by Eq. (11).

$$Q(s,a) \leftarrow Q(s,a) + \eta \left[R + \delta \max_{a'} Q(s',a') - Q(s,a) \right] \tag{11}$$

where s' is the new state after executing the action, and a' is the optimal action under the new state.

(5) Termination criterion judgment: the algorithm terminates and outputs the global optimal position if the iteration reaches the preset maximum number or the global optimal fitness value $f(GBest)$ satisfies the predefined minimum fitness threshold ϵ; otherwise, the procedure returns to step (2) and continues the next iterative calculation.

(6) With the optimal design parameters (kickoff point and dogleg severity) obtained from the proposed optimization algorithm, well trajectory can be performed.

5 Instance Verification

Taking the Bonan Oilfield as an instance, 10 wells in this field are selected as the dataset for intelligent well trajectory design, among which 9 wells serve as reference wells and 1 well acts as the well to be designed. Due to confidentiality requirements, the original data like coordinates are not provided, we only give the necessary parameters of well trajectory, shown in Table 2.

Table 2: Basic parameters of well trajectory for 10 wells in Bonan oilfield.

Well Name	Kickoff Point (m)	Dogleg (°/30 m)	Trajectory Profile
#0	3000	3.6/1.5	Vertical–Build–Hold–Drop–Hold
#1	2300	1.5	Vertical–Build–Hold
#2	3100	3.6/1.5	Vertical–Build–Hold–Drop–Hold
#3	2300	3.6/1.5	Vertical–Build–Hold–Drop–Hold
#4	2950	3.6/1.5	Vertical–Build–Hold–Drop–Hold
#5	2400	3.6/1.5	Vertical–Build–Hold–Drop–Hold
#6	2910	6.11/6.11	Vertical–Build–Build–Tangent
#7	2700	4.5/4.5	Vertical–Build–Hold–Build–Tangent
#8	2650	5.5/5.5	Vertical–Build–Hold–Build–Tangent
#9	2400	4.5	Vertical–Build–Hold

5.1 Feature Processing of Offset Wells

First we calculate the wellhead distance, target distance, drilling completion time difference and formation difference between the well to be designed (#0) and other offset wells, shown in Table 3. Then feature preprocessing is performed on the difference values of offset wells. For negatively correlated features, reciprocal processing is adopted to comply with the principle that a larger value corresponds to a higher priority. If the difference value is zero, the degree of difference is defined as 10 to avoid singularity, after that, normalization is performed on each feature by min–max normalization algorithm. The post-processed results of feature differences are presented in Table 4.

Table 3: Difference values of offset wells.

Well Name	Wellhead Distance (m)	Target Distance (m)	Time Distance (m)	Formation Difference
#1	635.67	219.31	1200	153
#2	0	631.34	15	0
#3	0	972.67	40	0
#4	0	390.12	32	162
#5	0	718.47	46	0
#6	7681.52	7232.99	1500	153
#7	5054.04	4096.46	1541	153
#8	5054.04	4554.84	1560	153
#9	1118.26	1275.81	1308	159

Table 4: Post-processed difference values of offset wells.

Well Name	Wellhead Distance (m)	Target Distance (m)	Time Distance (m)	Formation Difference
#1	0.00	0.99	0.00	0.00
#2	1.00	0.32	0.99	1.00
#3	1.00	0.20	0.37	1.00
#4	1.00	0.54	0.46	0.00
#5	1.00	0.28	0.32	1.00
#6	0.00	0.00	0.00	0.00
#7	0.00	0.02	0.00	0.00
#8	0.00	0.02	0.00	0.00
#9	0.00	0.14	0.00	0.00

5.2 Optimal Offset Wells Selection

The scores of offset wells are calculated by entropy weight algorithm. First, the probability matrix (Table 5) is calculated by Eq. (2) from difference matrix Table 3. Next, entropy values, diversity coefficients, and final entropy weights for each feature are obtained by Eq. (3), shown in Table 6.

So the final weights of each feature to select offset wells is $w = \{0.24, 0.16, 0.27, 0.33\}$. Take well #2 as instance, the final score calculated by Eq. (4) is $A_2 = 1 * 0.24 + 0.32 * 0.16 + 0.99 * 0.27 + 1 * 0.33 \approx 0.89$, the scores of other offset wells are given in Table 7. Then we can select optimal offset wells according to the scores.

5.3 Intelligent Design Results and Discussion

According to Table 7, the three offset wells with the highest scores (well #2, #3, #5) are selected as the basis for parameter determination, then we refer to the well trajectory parameters of the 10 wells listed in Table 2. The type of well profile to be designed is determined as "Vertical–Build–Hold–Drop–Hold" by voting, we also expanded the voting scope to all 9 offset wells, the target profile still obtained the majority of votes, which preliminarily proves the reliability of the selection result; the kickoff point ranges from 2300 to 3100 m, and the dogleg severity is 3.6/1.5 (°/30 m). Now we obtain the constraint conditions of kickoff point and dogleg severity, and calculate the well trajectory segment length based on the mathematical model and take it as the optimization objective.

Table 5: Probability matrix of post-processed difference values of offset wells.

Well Name	Wellhead Distance (m)	Target Distance (m)	Time Distance (m)	Formation Difference
#1	0.00	0.39	0.00	0.00
#2	0.25	0.13	0.46	0.33
#3	0.25	0.08	0.17	0.33
#4	0.25	0.22	0.22	0.00
#5	0.25	0.11	0.15	0.33
#6	0.00	0.00	0.00	0.00
#7	0.00	0.01	0.00	0.00
#8	0.00	0.01	0.00	0.00
#9	0.00	0.06	0.00	0.00

Table 6: Values of entropy weight algorithm.

Features	Entropy Values	Diversity Coefficients	Entropy Weights
Wellhead distance	0.63	0.37	0.24
Target distance	0.75	0.25	0.16
Time difference	0.58	0.42	0.27
Formation difference	0.50	0.50	0.33

Table 7: Offset well selection scoring.

Well Name	#1	#2	#3	#4
Scores	0.16	0.89	0.69	0.45
#5	**#6**	**#7**	**#8**	**#9**
0.69	0	0.003	0.003	0.02

The two main contributions are optimal offset well range extraction and RL-based PSO tuning, the key purpose of selecting optimal offset wells in our work is to improve computational efficiency. To quantify the performance improvement brought by RL-PSO, we tested three configurations on the target well: (a) Standard GA with the optimal offset well range ({[2300, 3100] m, 3.6/1.5°/30 m}); (b) Standard PSO with the optimal offset well range ({[2300, 3100] m, 3.6/1.5°/30 m}); and (c) Full RL-PSO with the optimal offset well range. Each configuration was run 10 times independently to guarantee result stability. The designed result of experts is also given for comparison. Results are presented in Table 8.

It can be seen from the experimental results that compared with GA with Offset and PSO with Offset, RL-PSO with Offset achieves the shortest trajectory length and the minimum standard deviation, demonstrating superior optimization performance and stronger result stability. In terms of convergence speed, RL-PSO converges within an average of 48 iterations, representing a 32.4% reduction compared with PSO (71 iterations). This verifies that the reinforcement learning-based parameter tuning strategy can effectively improve the optimization accuracy and anti-interference capability of the PSO algorithm. For computational cost, although RL-PSO consumes slightly more computation per iteration, the substantial decline in iteration count yields an overall time advantage. Furthermore, compared with the expert-designed

scheme, intelligent optimization algorithms drastically cut down the design time, RL-PSO shortens the trajectory length by 54.59 m and improves computational efficiency by approximately 17 times. The results fully demonstrate that the proposed improved algorithm possesses practical value and superior performance.

Table 8: Comparison among different optimization algorithms.

Method	Average Length (m)	Std	Average Iterations	Average Time (s)
GA + Offset	4308.04	1.59	77	68.4
PSO + Offset	4280.03	1.63	71	41.7
RL-PSO + Offset	4248.71	1.48	48	35.3
Expert	4303.30	–	–	600

Mean convergence curves with standard deviation bands derived from ten independent repeated runs are given in Fig. 5. For every 10 iterations, the best fitness value from all ten runs is collected to calculate the average value and standard deviation. The figure intuitively reflects the relative convergence speed of each algorithm, and all quantitative statistical indicators are detailed in Table 8.

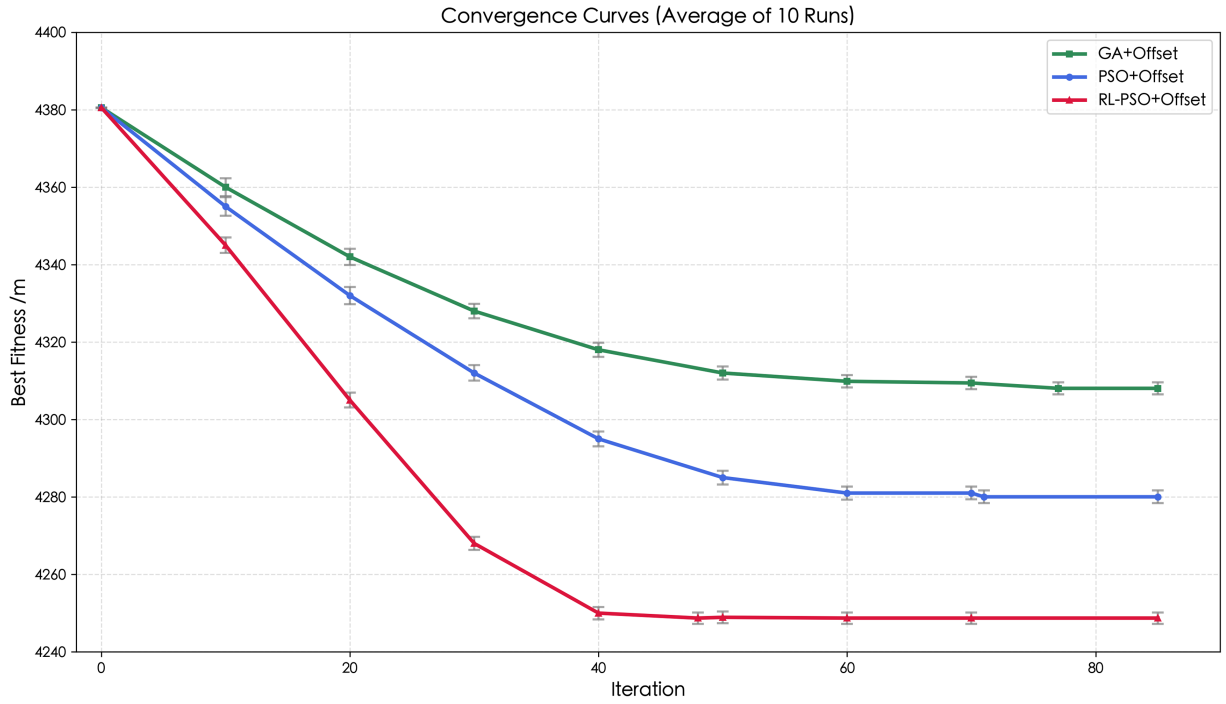


Figure 5: Convergence curves.

5.4 Parameter Sensitivity Analysis

To assess the sensitivity of formation similarity to the coding cost values, we conducted a sensitivity analysis [23] by varying the cost set around the baseline {10, 5, 1} with DTW distance 159. Three alternative cost sets were tested: {8, 4, 1} with DTW distance 127, {12, 6, 1} with DTW distance 229, and {10, 5, 2} with DTW distance 168. After taking the reciprocal operation, variations of this value have little influence on the final offset well score. This confirms that the offset well selection results are robust to moderate

changes in formation coding costs, and the chosen cost values $\{10, 5, 1\}$ are appropriate for the hierarchical formation structure.

To avoid singularity during reciprocal transformation, we assigned 10 to the zero distance features. Theoretically, we can assign any positive constant, as long as this constant is larger than the reciprocals of all non-zero distances, values such as 5, 10 and 20 are all above the critical value and ensure that zero-distance features have the highest weight, thus keeping the ranking table.

The initial parameters of RL-PSO are $\{N = 30, \eta = 0.2, \delta = 0.8, k_1 = 0.6, k_2 = 0.2, k_3 = 0.2\}$ according to the steps of Section 4.2. The number of particles $N = 30$ is a standard choice for PSO on low-dimensional problems, the learning rate $\eta = 0.2$, and discount factor $\delta = 0.8$ are standard Q-learning defaults widely used in the literature. A sensitivity analysis was conducted on $\{k_1 = 0.5, k_2 = 0.25, k_3 = 0.25\}$ with final average trajectory length 4257.36 m, $\{k_1 = 0.7, k_2 = 0.15, k_3 = 0.15\}$ with final average trajectory length 4264.05 m, it showed that the final trajectory length varied by less than 0.3% (<15 m), indicating that the results are not highly sensitive to moderate weight perturbations around the chosen values.

It should be clarified that the proposed RL-PSO optimization framework is equally applicable to 3D well trajectory design. The core of the method—using the RL agent to adaptively adjust PSO parameters and search for optimal design parameters that minimize total trajectory length—is independent of the dimensionality of the trajectory. The distinction between 2D and 3D trajectory design lies solely in the underlying mechanical calculation model.

6 Conclusion

To address the existing drawbacks of traditional well trajectory design, such as excessive reliance on expert experience and low design efficiency, and make full use of offset well data and formation parameters, this paper proposes an intelligent well trajectory design method based on optimal offset wells and reinforcement learning-particle swarm optimization (RL-PSO) algorithm, to obtain the optimal design parameters of the well trajectory automatically that fully meet the engineering constraints and practical operation requirements.

A multi-dimensional comprehensive evaluation for offset well selection is established, which takes wellhead distance, target distance, completion time difference, and formation difference as evaluation features. The difference of each feature is obtained by feature coding and preprocessing; especially for formation data, the feature is taken as a character sequence, and DTW is applied to compute the difference between two wells. Then the entropy weight method is used to get the weight of each feature, and the comprehensive score of each well is calculated. The optimal offset wells are selected to provide reliable data support for intelligent well trajectory design.

Trajectory optimization is formulated as a single-objective constrained optimization problem that minimizes well trajectory length subject to engineering constraints defined by the parameter ranges of offset wells. Additional optimization objectives will be incorporated in future research. The RL-PSO algorithm is used to dynamically tune PSO optimization parameters, balance global exploration and local exploitation, guarantee engineering-compliant trajectories, and improve the accuracy and efficiency of trajectory design.

The research results show that the intelligent design method proposed in this paper can obtain a more economical well trajectory, improve design efficiency, and reduce the dependence on expert experience. It provides technical support for the intelligent and automated development of drilling engineering. The proposed method is currently verified only based on a single oilfield block. Restricted by the available drilling data, multi-block and multi-oilfield verification is not carried out in the present study. In future research, abundant data from multiple different oilfields will be collected to complete cross-field validation and further

improve the generalization performance of the RL-PSO algorithm. Besides, friction torque and drillability will be incorporated to establish a multi-objective optimization formulation, while advanced algorithms such as deep reinforcement learning can be adopted for optimized solutions.

Acknowledgement: Not applicable.

Funding Statement: National Key R&D Program of China: Key Technologies and Equipment for Polar Oil & Gas Drilling, Subproject: Environmental Protection and Waste Treatment Technology for Polar Drilling, Number: 2022YFC2806403.

Author Contributions: The authors confirm contribution to the paper as follows: Jiaqi Zhang contributed to the conceptualization and design of the study, drafting and revising of the manuscript. Guo Wang was responsible for the literature review, theoretical framework development. Song Deng contributed to manuscript review and editing, and discussions on results. Duo Yuan contributed to experimental setup and manuscript refinement. Xiaoli Hu contributed to the mathematical model building, especially for well trajectory. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the corresponding author Jiaqi Zhang, upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Li G, Song X, Tian S, Zhu Z. Intelligent drilling and completion: a review. Engineering. 2022;18(2):33–48. doi:10.1016/j.eng.2022.07.014.
2. Heidari Varnamkhasti MR, Al Hinaai Q, Patil R, Al Lawati AB, Al Ghasani R, Vasquez Bautista RO. Hard rock horizontal drilling—using mechanical specific energy, downhole vibrations analysis, and well design to optimize performance. In: Proceedings of the SPE Conference at Oman Petroleum & Energy Show; 2022 Mar 21–23; Muscat, Oman. Richardson, TX, USA: Society of Petroleum Engineers; 2022. doi:10.2118/200262-ms.
3. Farzay O, Shakhouri A, Gholami R, Al Ajmi AM. Optimum directional well path design considering collapse and fracture pressures. Int J Oil Gas Coal Technol. 2022;30(4):388. doi:10.1504/ijogct.2022.124414.
4. Yu R, Diao B, Gao D. Design method for relief well trajectory to avoid side-track operations and control the relative distance. Geoenergy Sci Eng. 2023;231(3):212325. doi:10.1016/j.geoen.2023.212325.
5. Chen D, Mao K, Ye Z, Li W, Yan W, Wang H. An artificial intelligent well trajectory design method combining both geological and engineering objectives. Geoenergy Sci Eng. 2024;236(1):212736. doi:10.1016/j.geoen.2024.212736.
6. Yousefzadeh R, Ahmadi M. Well trajectory optimization under geological uncertainties assisted by a new deep learning technique. SPE J. 2024;29(9):4709–23. doi:10.2118/221476-pa.
7. Maita MA, Noguera D. Optimization of drilling times by customizing the trajectory design in Vaca muerta. In: Proceedings of the SPE Argentina Region Young Professionals Symposium; 2024 Nov 12; Neuquén, Argentina. Richardson, TX, USA: Society of Petroleum Engineers; 2024. doi:10.2118/223487-ms.
8. Srivastava A, Raizada N. Optimization of time, effort, and costs for well planning using standardization and automation of the well design process on a cloud-based solution. J Phys Conf Ser. 2024;2834(1):012188. doi:10.1088/1742-6596/2834/1/012188.
9. Hashemi A, Todorov D. Integrating automated offset well selection and machine learning for enhanced ROP prediction and optimization in oil and gas drilling. In: Proceedings of the International Petroleum Technology Conference; 2025 Feb 18–20; Kuala Lumpur, Malaysia. doi:10.2523/iptc-24701-ms.
10. Deng S, Pan H, Wang H, Xu S, Yan X, Li C, et al. A hybrid machine learning optimization algorithm for multivariable pore pressure prediction. Pet Sci. 2024;21(1):535–50. doi:10.1016/j.petsci.2023.09.001.

11. Seyyedattar MS, Parvizi H, Malekzadeh E. Optimization of wellhead location selection and well path design. Pet Sci Technol. 2012;30(21):2236–45. doi:10.1080/10916466.2010.514585.
12. Masood S, Qureshi MP, Shah MB, Ashraf S, Halim Z, Abbas G. Dynamic time wrapping based gesture recognition. In: Proceedings of the 2014 International Conference on Robotics and Emerging Allied Technologies in Engineering (iCREATE); 2014 Apr 22–24; Islamabad, Pakistan.
13. Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D, et al. Playing atari with deep reinforcement learning. arXiv:1312.5602. 2013.
14. Andrieu C, de Freitas N, Doucet A, Jordan MI. An introduction to MCMC for machine learning. Mach Learn. 2003;50(1):5–43. doi:10.1023/A:1020281327116.
15. Ming F, Gong W, Wang L, Jin Y. Constrained multi-objective optimization with deep reinforcement learning assisted operator selection. IEEE/CAA J Autom Sin. 2024;11(4):919–31. doi:10.1109/jas.2023.123687.
16. Qi B, Chen P, Xia H, Wei C. Optimizing the method used to design the trajectory of a horizontal well. Chem Technol Fuels Oils. 2014;50(4):337–43. doi:10.1007/s10553-014-0530-4.
17. Xiao J, Zhang Z, Terzi S, Anwer N, Eynard B. Dynamic task allocations with Q-learning based particle swarm optimization for human-robot collaboration disassembly of electric vehicle battery recycling. Comput Ind Eng. 2025;204(5):111133. doi:10.1016/j.cie.2025.111133.
18. Mitchell FR, Robinson JD, Vogiatzis JP, Pehoushek F, Moran JH, Ausburn BE, et al. Using resistivity measurements to determine distance between wells. J Petrol Technol. 1972;24(6):723–40. doi:10.2118/3495-pa.
19. Patro SK, Sahu KK. Normalization: a preprocessing stage. Int Adv Res J Sci Eng Technol. 2015;2(3):20–2. doi:10.17148/iarjset.2015.2305.
20. Paleo BW. Levenshtein distance. In: Dictionary of algorithms & data structures. Gaithersburg, MD, USA: National Institute of Standards and Technology; 2010.
21. Liu ZK, Li Q, Zhang SX. Optimize oil field simulation measures through multi-factor comprehensive evaluation method based on fuzzy mathematics and grey correlation. Adv Mater Res. 2012;524–527:1503–6. doi:10.4028/www.scientific.net/amr.524-527.1503.
22. Ren Y. Research on intelligent design and optimization methods for well trajectory [dissertation]. Wuhan, China: Changjiang University; 2024. (In Chinese).
23. Xiao J, Zhang Z, Terzi S, Tao F, Anwer N, Eynard B. Multi-scenario digital twin-driven human-robot collaboration multi-task disassembly process planning based on dynamic time Petri-net and heterogeneous multi-agent double deep Q-learning network. J Manuf Syst. 2025;83:284–305. doi:10.1016/j.jmsy.2025.09.011.