Petroleum Science 21 (2024) 597–608

Contents lists available at ScienceDirect

Petroleum Science

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

Original Paper

Dynamic plugging regulating strategy of pipeline robot based on reinforcement learning

Xing-Yuan Miao, Hong Zhao*

College of Mechanical and Transportation Engineering, China University of Petroleum, Beijing, 102249, China

ARTICLE INFO

Article history:
Received 5 December 2022
Received in revised form
23 May 2023
Accepted 16 August 2023
Available online 18 August 2023

Edited by Jia-Jia Fei and Min Li

Keywords:
Pipeline isolation plugging robot
Plugging-induced vibration
Dynamic regulating strategy
Extreme learning machine
Improved sparrow search algorithm
Modified Q-learning algorithm

ABSTRACT

Pipeline isolation plugging robot (PIPR) is an important tool in pipeline maintenance operation. During the plugging process, the violent vibration will occur by the flow field, which can cause serious damage to the pipeline and PIPR. In this paper, we propose a dynamic regulating strategy to reduce the plugging-induced vibration by regulating the spoiler angle and plugging velocity. Firstly, the dynamic plugging simulation and experiment are performed to study the flow field changes during dynamic plugging. And the pressure difference is proposed to evaluate the degree of flow field vibration. Secondly, the mathematical models of pressure difference with plugging states and spoiler angles are established based on the extreme learning machine (ELM) optimized by improved sparrow search algorithm (ISSA). Finally, a modified Q-learning algorithm based on simulated annealing is applied to determine the optimal strategy for the spoiler angle and plugging velocity in real time. The results show that the proposed method can reduce the plugging-induced vibration by 19.9% and 32.7% on average, compared with single-regulating methods. This study can effectively ensure the stability of the plugging process.

1. Introduction

Pipeline transportation plays an important role in oil and gas transportation. However, due to the long usage time in service, pipeline leakage, corrosion and wax build-up may occur, which can cause environmental damage and economic losses (Rai and Kim, 2021). Safe pipeline maintenance and repair technology are vital for pipeline operation. Pipeline isolation plugging robot (PIPR) is crucial to pipeline maintenance work. It can plug the pipeline without stopping the transportation, and avoid opening holes on the pipe wall (Lie and Muangsuankwan, 2015). At the same time, it can work in the high-pressure environment, improving work efficiency and the safety of plugging operations.

A considerable amount of researches have been performed on pipeline isolation plugging techniques over the past few years, and major contributions are in the field of mechanical design. The mechanical structure of PIPR is shown in Fig. 1. Tveit and Aleksandersen designed a smart PIPR (Tveit and Aleksandersen, 2000), which had been successfully applied to land and submarine pipelines. The PIPR developed by T.D. Williamson (TDW) in American ranks among the top of the world. It has successfully completed more than 120 times for high-pressure plugging operations. Yan designed a kind of PIPR consisted of a robot drive unit, connection unit and blocking unit, and verified the passing ability and climbing performance by experimental platform (Yan et al., 2020).

The performance of the PIPR is affected by many factors, such as the structure, sealing material and motion velocity. Studies on structural optimization and motion control of PIPR have been carried out. Zhang used the numerical simulation method for analyzing pressure fluctuation phenomenon during the plugging process (Zhang et al., 2018). Response surface method was used to investigate the factors of deceleration time, flow velocity and PIPR's aspect ratio, which can provide a reference for reducing the pressure fluctuation. Zhao and Hu conducted optimal design using response surface method for reducing unsteady force effects on the x and y directions (Zhao and Hu, 2017). And a modified genetic algorithm was proposed to optimize the design parameter ratios of PIPR. Wang proposed the PID synchronous control method for marine spherical PIPR (Wang et al., 2020), reducing the rotation error of the plug head. Miao proposed the active disturbance rejection control method based on the whale optimization algorithm (WOA-ADRC) to control the PIPR's motion velocity when

* Corresponding author.
E-mail address: hzhao@cup.edu.cn (H. Zhao).

https://doi.org/10.1016/j.petsci.2023.08.016

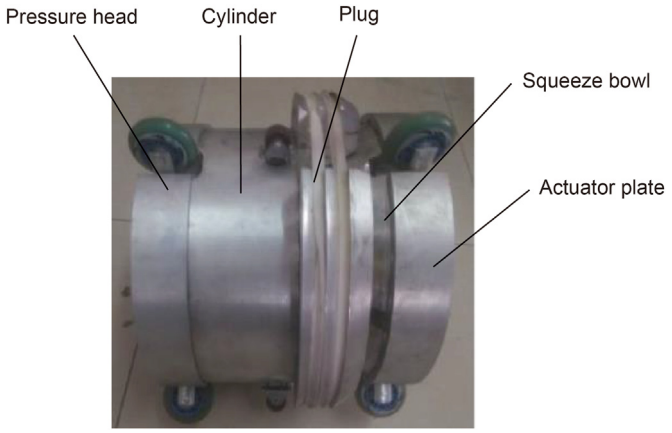

Fig. 1. The mechanical structure of PIPR.

passing through girth weld (Miao et al., 2022a). However, there are few researches on dynamic regulating control during the plugging process, and no studies have shown the combined regulating strategy for PIPR.

Reinforcement learning has been widely used in robot control due to the self-learning ability. Goharimanesh designed a fuzzy reinforcement learning controller for continuous robot (Goharimanesh et al., 2020), and the genetic algorithm was used to optimize control parameters to improve the robot's stability and trajectory tracking ability. Ignacio applied the deep reinforcement learning control strategy to estimate adaptive multi-PID controller parameters for mobile robots (Ignacio et al., 2020). Miao proposed an Actor-Critic controller for PIPR to solve the speed excursions caused by the pipeline large-scale complex deformation (Miao et al., 2022b), which can adapt to the pipeline deformation through self-learning. Wu developed an energy-saving control system for PIPR (Wu et al., 2021a), in which the Q-learning algorithm was used to adjust the opening of the hydraulic pump and accumulator, the energy-saving efficiency of the plugging process had been improved. Therefore, reinforcement learning can perform well in dynamic regulating control for PIPR.

In our previous research (Miao and Zhao, 2022), the PIPR with active spoiler device was proposed, and the vibration reduction controller was established. It proved that the flow field vibration could be reduced by adjusting the spoiler angles. However, the plugging velocity can also affect the stability of the plugging process (Wu and Zhao, 2019). The dynamic regulating strategy combining spoilers and plugging velocity has not been studied. Most previous researches conducted the single-regulating methods, it could not fully reduce the plugging-induced vibration during the entire plugging process.

In this paper, we propose a dynamic regulating strategy for spoiler angle and plugging velocity to reduce the plugging-induced vibration. First of all, the complex structure of the PIPR is simplified, three foldable spoilers are designed on the pressure head. Numerical simulations and experiments are performed to observe the flow field vibration of plugging operation. And the influence of spoiler angle and plugging velocity is studied. On the basis of pressure difference measured by experiments, the mathematical model of plugging-induced vibration and plugging states and spoiler angles is established based on ELM, in which the parameters of model are optimized by ISSA. The modified Q-learning algorithm is designed based on simulated annealing to determine the optimal strategy, regulating the spoiler angle and plugging velocity in real time. Through the visual experiments, the proposed optimal strategy is verified, which can greatly reduce the flow field vibration during the plugging process, compared with single-regulating methods (only regulating the spoiler angle or plugging velocity).

2. Simplified PIPR model with active spoiler device

The plugging-induced vibration of the PIPR is a result of multi-physics. During the plugging operation, the fluid will decrease rapidly as the plugging process increases. The sudden contraction of the in-pipe flow field induces intermittent de-flow of the fluid near the plugging area. It can cause resonance damage and wake vortex of PIPR, resulting in a large vibration phenomenon. However, the internal structure of PIPR model is complicated, it is not convenient to study the changes of external flow field during the plugging process. Therefore, a simplified model with active spoiler device is designed, which basically restores the external form of the traditional model, and the size is reduced based on the geometrical similarity principle and experimental set-up.

The schematic diagram of the plugging process is shown in Fig. 2. The simplified PIPR model is developed on the basis of previous research (Wu et al., 2021b). After the PIPR enters the pipeline, it is driven by the pressure difference between upstream and downstream in the pipeline (Mirshamsi and Rafeeyan, 2015). The extremely low frequency (ELF) signal is sent by remote control center through the signal modem. When the PIPR reached the plugging position, the inner pneumatic cylinder drives the actuator to move. The sliders are pushed by the push tube to slip along the slide until they contact with the left side of the pressure head to achieve self-locking. During this process, the sealing ring is squeezed and expands radially. An interference fit is formed between the sealing ring and the pipe wall. When the plugging operation is completed, the actuator releases pressure, and the slips move to the original position until the sealing ring returns to its original state. Then, the PIPR moves towards the downstream region until it reaches the receiving end.

During the process of plugging operation, the water hammer phenomenon may occur, which can affect the plugging process. Therefore, the control system of PIPR is important. As shown in Fig. 3(a), the PIPR simplified model includes push tube, sliders, sealing ring, and cone tube. The pneumatic control system is designed for plugging operation and active spoiler device, as shown in Fig. 3(b) and (c). Compared to hydraulic system, the pneumatic system can reduce the environmental pollution. The plugging velocity tracking can be achieved by controlling the pneumatic cylinder's motion velocity. The chassis moves to the left under the action of the inner pneumatic cylinder to compress the spring, then the spoilers can open outwards to the designated angles.

3. Numerical simulation and experimental study

3.1. Numerical simulation

During the process of plugging operation, the external structure of the PIPR will change significantly. At the same time, the annular flow area between the PIPR and the pipeline inner wall will gradually decrease, which can inevitably lead to drastic changes in the pressure, velocity and other characteristic parameters of the flow field around the PIPR. Therefore, numerical simulation is performed for different plugging states using FLUENT software. The flow field model around the PIPR is shown in Fig. 4. The structure size of PIPR is shown in Table 1.

According to the simplified PIPR model, the axial stroke of the plugging process is 25 mm. In order to compare with the experiment, the flow medium was water. The parameters of numerical simulation are shown in Table 2. The centerline of the PIPR model is

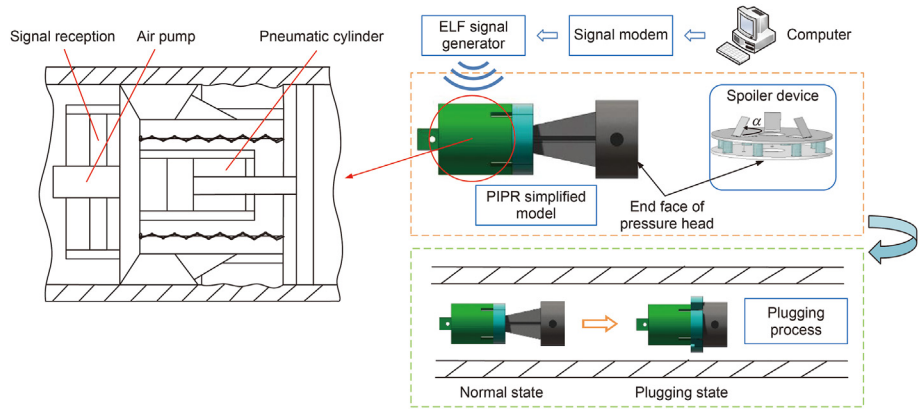

Fig. 2. The schematic diagram of the plugging process.

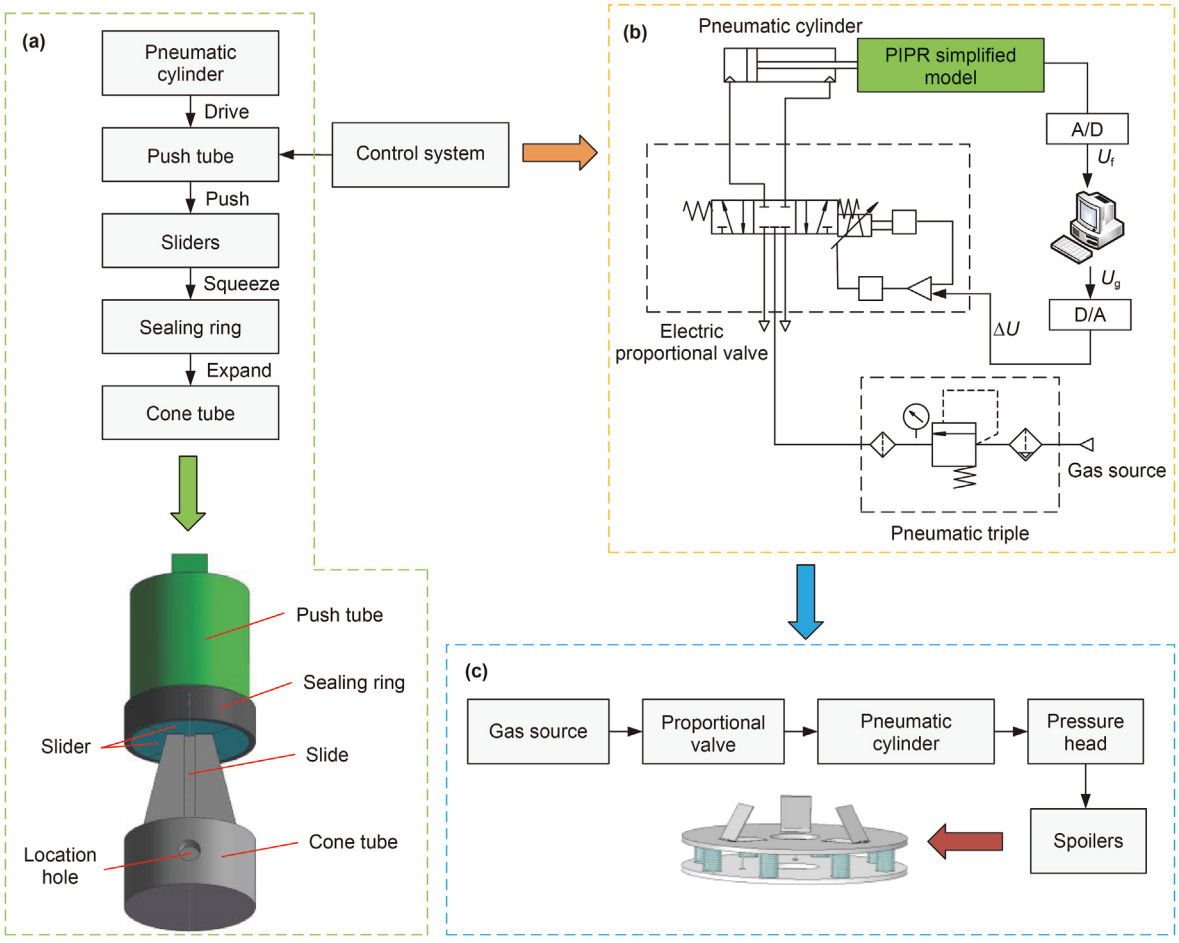

Fig. 3. The schematic diagram of control system: (**a**) PIPR simplified model; (**b**) Pneumatic control system; (**c**) The flow of active spoiler device.

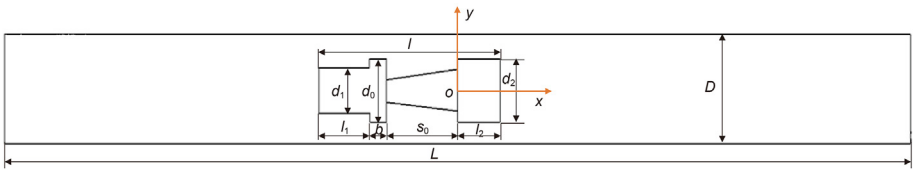

Fig. 4. The flow field model around the PIPR.

Table 1
The parameters of PIPR's structure.

Parameter	Symbol	Value, mm	Parameter	Symbol	Value, mm
The length of pipeline	L	1000	The diameter of sealing ring	d_0	37
The length of PIPR	l	90	The diameter of push tube	d_1	33.5
The length of push tube	l_1	35	The diameter of cone tube	d_2	37
The length of cone tube	l_2	20	The width of sealing ring	b	10
The diameter of pipeline	D	50	The length of slide	s_0	25

Table 2
The parameters of numerical simulation.

Parameter	Symbol	Value	Unit
Density of medium	ρ	998.2	kg/m^3
Dynamic viscosity of medium	μ	1.01×10^{-3}	Pa·s
In-pipe reference pressure	P_0	5	MPa
Inlet velocity	v_0	2.68	m/s
Turbulence intensity	T_i	3.66	%
Hydraulic diameter	H_d	50	mm

coincided with the centerline of the pipeline. The wall of the pipeline and the PIPR is defined as the non-slip wall. The direction of gravity is the negative y-axis. The method of combining structured grid with unstructured grid is used, and the grid near the pipe wall and the PIPR is refined. When the number of grids exceeds 10^6, the results of the numerical simulation are not affected by the number of grids.

The velocity distribution of the symmetry plane of the pipeline ($z = 0$) is shown in Fig. 5. The plugging states of 0%, 40%, and 95% are selected to observe the flow field. It can be clearly observed that the flow field near the front and rear faces of the PIPR will form a stagnation region. When the upstream uniform fluid flows through the front end face, due to its blocking, the fluid kinetic energy loss is serious, and the flow velocity is sharply reduced, thus forming a stagnation flow area on the front end face. With the increase of plugging state, the in-pipe flow area decreases and the upstream fluid velocity decreases, leading to the further increase of the area of the stagnation region. This process will cause impact on the PIPR and affect the stability of the plugging operation. Under the action of high pressure difference, the fluid accelerates to pass through the narrow annular space between the PIPR and the pipe wall. And the flow velocity rises sharply to the peak, a narrow high-speed flow region is formed near the downstream pipe wall. The high-speed flow region merges in the downstream, and causes fluctuation in the middle part of the pipeline, the flow velocity tends to be stable. During the plugging process, the peak flow velocity increases continuously, so pressure pulsation can occur at the pipe wall near the plugging point.

For different plugging states, the flow velocity of downstream centerline of the PIPR is shown in Fig. 6. It can be clearly observed that the flow velocity in the downstream local region near the PIPR is basically negative, which indicates that there is backflow in this region. Before the plugging process reaches 80%, the length of the backflow zone and the plugging state are in the direct ratio. With the increase of the plugging state, the downstream velocity fluctuation becomes more intense.

The total pressure distribution of the symmetry plane of the pipeline ($z = 0$) is shown in Fig. 7. Total pressure is the sum of static pressure and dynamic pressure. Due to the blocking of the PIPR, the surrounding flow field is divided into three parts: the upstream is the high-pressure and low-velocity region, the middle is the plugging jet region, and the downstream is the low-pressure backflow region. With the increase of plugging state, the area of high-pressure and low-pressure zones gradually expands, the area of jet zone decreases. And the pressure gradient in the pipeline increases, leading to the increasing axial force of the PIPR. There are many scattered low-pressure centers in the downstream backflow region, and the location of the low-pressure center is changing with the plugging operation, and the affected zone is further diffusing. It indicates that there is alternative formation and separation of vortices at the tail of the PIPR during the plugging process, which can cause pressure pulsation.

The drag coefficient can reflect the force state of the PIPR in the flow field, it can be expressed as Eq. (1). Fig. 8 shows the axial drag coefficient and radial drag coefficient during the plugging process. The axial drag coefficient has been increasing since the plugging operation. It changes steadily in 0%–80% of the plugging process. However, when the plugging state reaches 80%, the resistance increases by an order of magnitude. Therefore, the PIPR will be subjected to the high axial impact force at the end of the plugging operation, posing a potential threat to the plugging operation. The radial drag coefficient fluctuates with the increase of plugging state, and the fluctuation amplitude increases in the later plugging

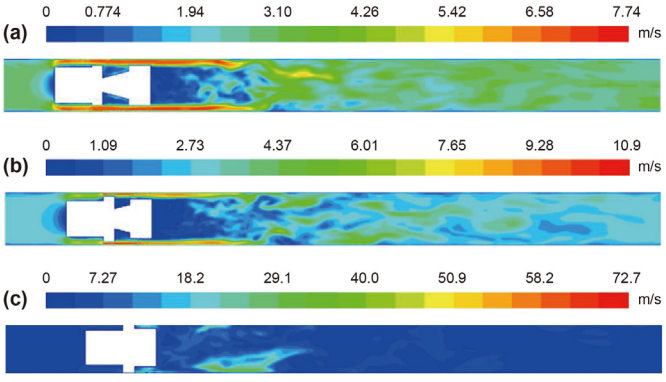

Fig. 5. The velocity distribution of different plugging states: (**a**) 0%; (**b**) 40%; (**c**) 95%.

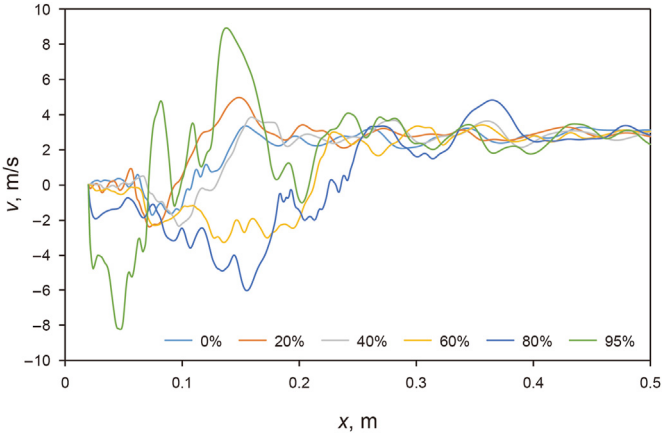

Fig. 6. The flow velocity of downstream centerline.

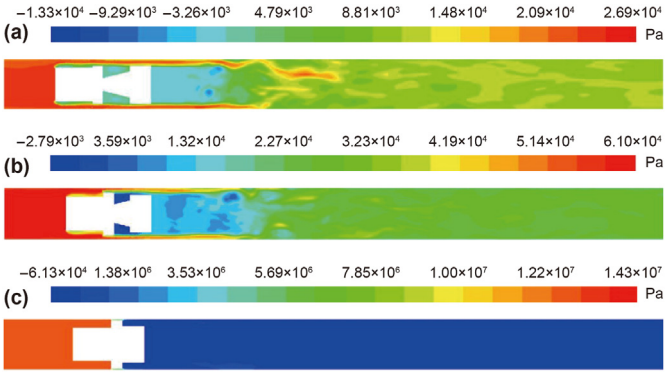

Fig. 7. The total pressure distribution of different plugging states: (**a**) 0%; (**b**) 40%; (**c**) 95%.

period. The results show that there is periodic vortex shedding on the surface of the PIPR, which will cause the destructive vibration of the PIPR.

$$C_d = \frac{F_d}{\frac{1}{2}\rho u^2 A} \tag{1}$$

where F_d is the resistance of the object; ρ is the medium density; A is the cross-sectional area of the object perpendicular to the direction of flow motion; u is the mainstream velocity.

3.2. Dynamic plugging experiment

3.2.1. Experimental set-up

Through the results of numerical simulation, plugging-induced vibration is gradually intense with the plugging process. In order to observe the flow field vibration for different spoiler models, an experiment of dynamic plugging is designed for the simplified PIPR model with spoiler device, as shown in Fig. 9. During the experiment, the hydraulic pump inputs the medium into the pipe. In-pipe flow is adjusted by the throttle valve to make it reach the preset value. In order to facilitate adjustment, stepper motor is used to control the plugging process. When the medium in the pipe flows stably, the stepper motor is started to drive the ball screw, thereby making the PIPR perform the plugging operation. The pulse signal of the controller is adjusted to change the stepper motor's speed to adjust the plugging velocity of the PIPR. The pressure head with the spoilers of the PIPR is changed to repeat the above experimental steps. In order to observe the flow field conveniently, the transparent pipeline of organic glass is selected. Considering the pressure bearing capacity of the pipeline, the in-pipe reference pressure is 5 kPa.

3.2.2. Experimental results

Three pressure monitoring points are set on the pipe around the PIPR. Point A is the upstream of the plugging, point B is the plugging position, and point C is the downstream of the plugging. And the interval between the monitoring points is 100 mm. Experiments are carried out on seven different spoiler models with spoiler angles of 0°, 30°, 60°, 90°, 120°, 150°, and 180°, respectively. And the axial plugging velocity is 2.2 mm/s. In order to observe the fluctuation of the flow field more intuitively, a high-speed camera is used to record the fluctuation of the PIPR's tail region. The flow field state of the spoiler model with the spoiler angle of 30° is shown in Fig. 10. The fluctuation which is greater than 2 mm is defined as the effective fluctuation, and the sum of the effective fluctuations is used to measure the flow field oscillation. When the plugging process is 10%, 30%, 50% and 70%, the sum of effective fluid fluctuations is 7.25, 8.65, 11.45 and 15.25 mm respectively. It can be seen that with the increase of the plugging process, the fluctuation of the flow field gradually increases. When the plugging operation is completed with 70%, the flow field has already appeared a relatively large oscillation.

The pressure curves of monitoring points of 0° and 30° spoiler model are shown in Fig. 11. It can be seen that, during the plugging process, the upstream pressure continues to increase, while the downstream pressure gradually decreases, and finally negative pressure occurs. The midstream region is the concentrated zone of plugging operation. This region changes from upstream to downstream, and the pressure first decreases and then increases. During the plugging process, the pressure difference between upstream and downstream increases gradually, causing pressure pulsation around the PIPR, leading to destructive vibration to the pipeline and PIPR. And the plugging-induced vibration of different spoiler angles is different.

Fig. 12 shows the pressure difference between upstream and downstream ($\Delta P = P_A - P_C$) under different axial plugging velocities. It can be seen that the variation trend of pressure difference is basically same. The final value reaches almost the same level. However, under different plugging velocities, the changing rate of pressure difference is quite different, which will cause severe pressure impact in the pipeline.

4. Dynamic plugging regulating strategy

4.1. Regulating strategy based on modified Q-learning algorithm

Through numerical simulation and experiment, flow field vibration is serious during the plugging process. And the spoiler angle and plugging velocity can affect the flow field. Therefore, we propose a dynamic plugging regulating strategy based on modified Q-learning algorithm. The Q-learning algorithm is one of the commonly used algorithms in reinforcement learning. It introduces

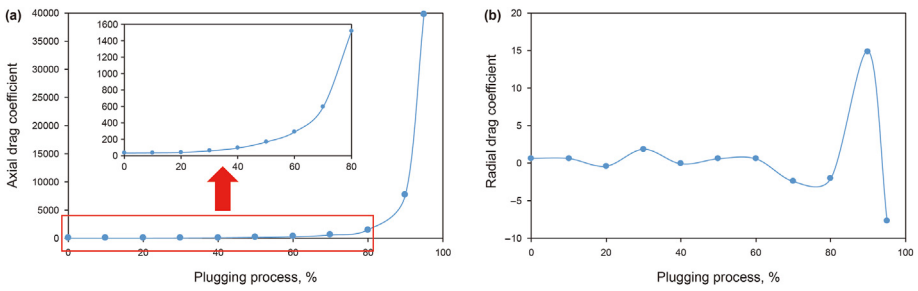

Fig. 8. The drag coefficient during the plugging process: (**a**) Axial drag coefficient; (**b**) Radial drag coefficient.

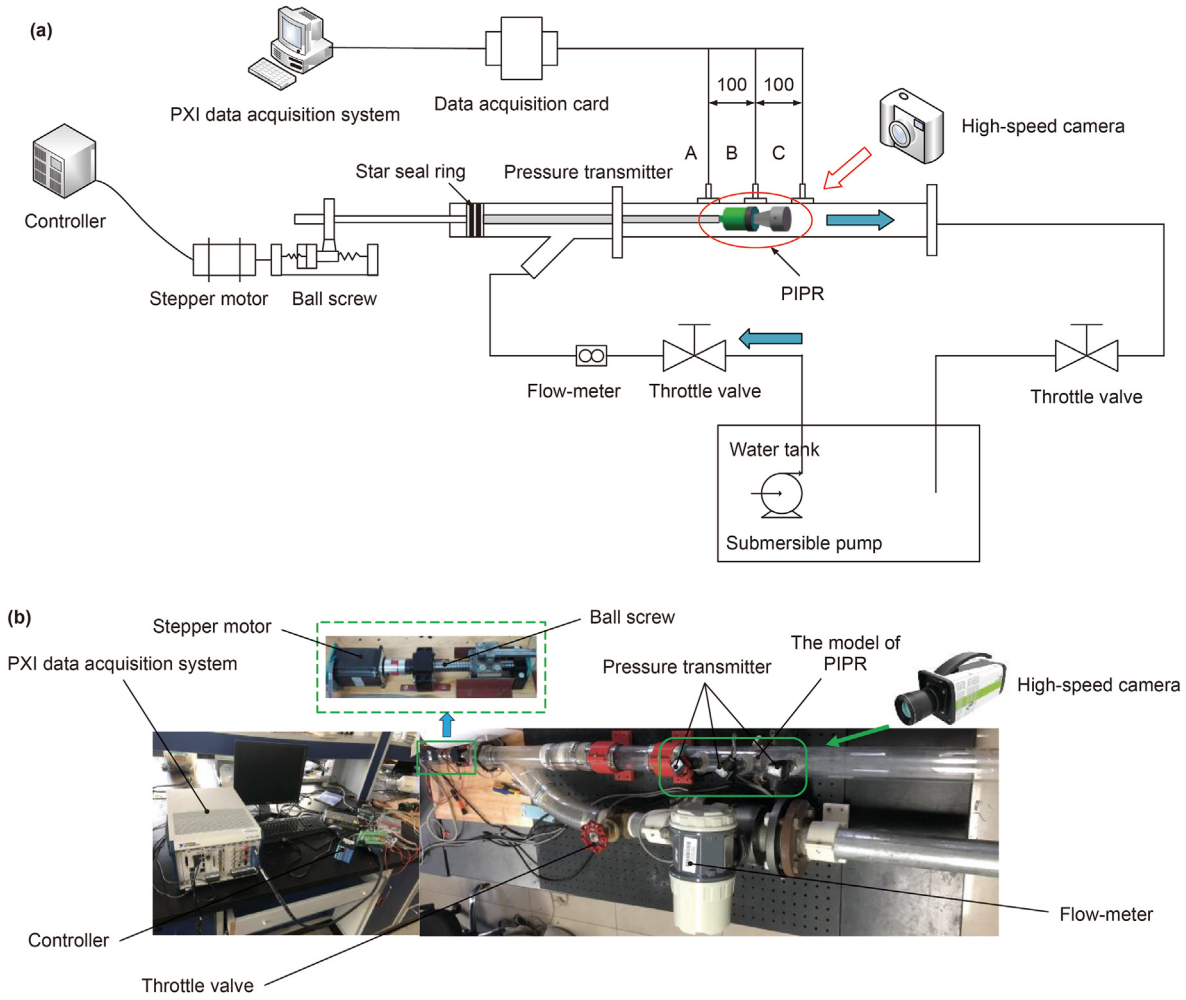

Fig. 9. Experimental set-up of dynamic plugging: (**a**) Schematic diagram; (**b**) Prototype of the experimental device.

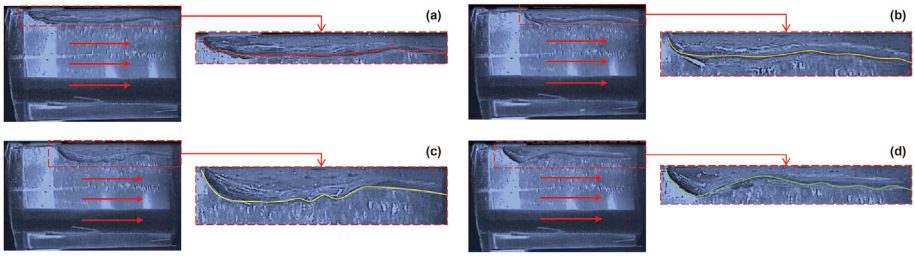

Fig. 10. The flow field state of the spoiler model with the spoiler angle of 30°: (**a**) 10%; (**b**) 30%; (**c**) 50%; (**d**) 70%.

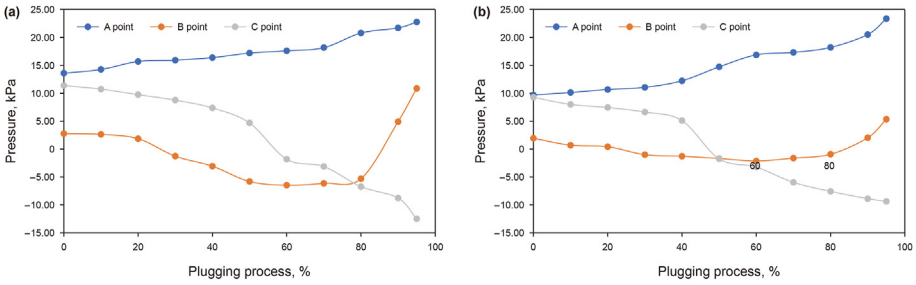

Fig. 11. The pressure curves of monitoring points: (**a**) 0° spoiler model; (**b**) 30° spoiler model.

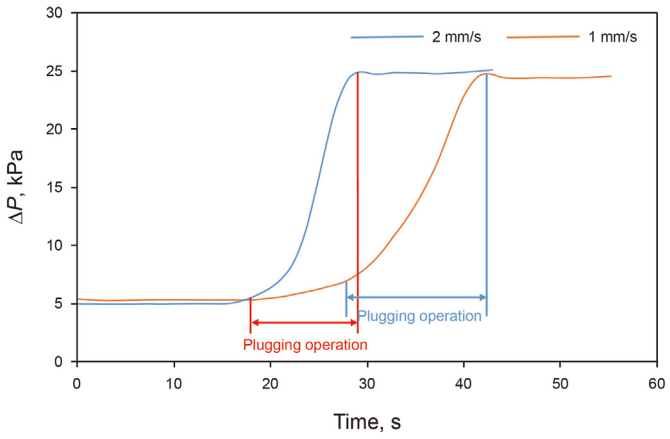

Fig. 12. The pressure difference of different axial plugging velocities.

the state-action value function to evaluate the current executed action, and the maximum state-action value function at the next moment is used as the basis (Rahman et al., 2018). The value function is constantly updated during the iterative process, and its form is shown in Eq. (2).

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha(r_{t+1} + \gamma \max Q(s_{t+1}, a_t) - Q(s_t, a_t)) \tag{2}$$

where $Q(s_t, a_t)$ is the state-action value function of the agent performing the action a_t in the s_t state; α is the learning rate, which determines the size of the value function, and it is set to 0.01; γ is the discount factor, which indicates that the value function's attention degree to the future, it is set to 0.9; r_{t+1} is the instantaneous reward value at the next time; $Q(s_{t+1}, a_t)$ is the state-action value function of the action at the next time.

Traditional Q-learning algorithm is difficult to balance the exploration and utilization for information. Too much exploration will reduce the convergence speed, but too much utilization cannot estimate the optimal reward well, and it is easy to fall into local optimum (Andrea et al., 2020). In order to balance the "exploration-utilization" process, the simulated annealing algorithm is introduced to improve the Q-learning algorithm. The simulated annealing algorithm is based on the Metropolis criterion. This method accepts the optimal solution with a certain probability. At the same time, it also accepts the non-optimal solution with a certain probability, which is of great significance for jumping out of the local optimum. In this study, we used the modified Q-learning algorithm to regulate the spoiler angle and plugging velocity during the plugging process, as shown in Fig. 13. The PIPR is used as an agent, the flow field is set as the algorithm environment. The current state and next action are obtained by observing the flow field environment, and the decision is made according to the reward, thereby updating the Q-table. Based on the experimental results in Section 3.2.2, the spoiler angle can affect the pressure difference (ΔP), and the plugging velocity mainly affects the changing rate of the pressure difference ($\Delta P/\Delta t$). Therefore, the ΔP and $\Delta P/\Delta t$ are used as the feedback of flow field state.

4.1.1. State space

In the modified Q-learning algorithm, the plugging state x, spoiler angle α and plugging velocity v are used as status inputs. The plugging state x is 0–95%, the spoiler angle α is 0–180°, and the plugging velocity v is 0.02–2 mm/s. The three inputs are discretized into multiple values, as shown in Eqs. (3)–(5). The plugging state is selected with the interval of 1%, which is a sequential state. Considering the safety of experiment, the numerical simulation and experiment both stop at the 95% of plugging process. The spoiler angle is set as the same as the experiment. And the plugging velocity is divided into ten values. According to the plugging velocity and spoiler angle in each plugging process, the state can be uniquely determined.

$$x = \{0\%, 1\%, 2\%, \cdots, 94\%, 95\%\} \tag{3}$$

$$\alpha = \{0°, 30°, 60°, 90°, 120°, 150°, 180°\} \tag{4}$$

$$v = \{0.02, 0.04, 0.08, 0.12, 0.16, 0.2, 0.4, 0.8, 1.6, 2\} \tag{5}$$

4.1.2. Action space

The action space includes the change of the spoiler angle and the change of the plugging velocity during the plugging process. The above variables are continuous quantities, which are discretized into a series of fixed values. Due to the uneven distribution of state quantities, they are encoded. The spoiler angle is encoded as 0–6 to represent the seven angles in Eq. (4). The plugging velocity is encoded as 0–9 to represent the ten velocity values in Eq. (5). So the action space can be defined as the changing quantities, as shown in Eq. (6). And the constraint is set to prevent the variables out of range. Any combination of the changing values corresponds to an action strategy.

$$a = \{\Delta\alpha, \Delta v\} \tag{6}$$

where $\Delta\alpha$ is the change of the spoiler angle, which is the range of −6~6; Δv is the change of the plugging velocity, which is the range of −9~9.

4.1.3. Reward function

Reasonable reward function is crucial for the reinforcement learning algorithm, which can affect the choice of action strategy. The ΔP and $\Delta P/\Delta t$ can show the pressure fluctuation and plugging-

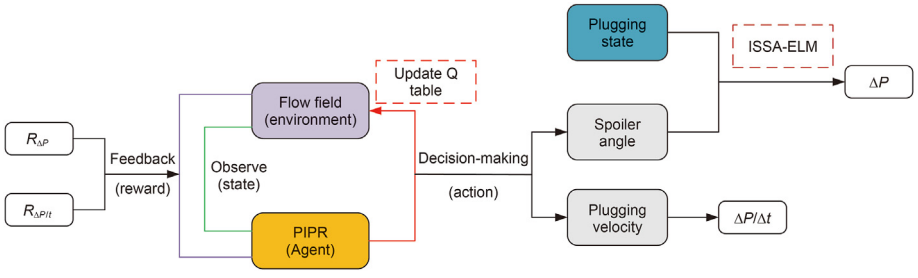

Fig. 13. The flowchart of dynamic plugging regulating strategy.

induced vibration of the plugging process. To ensure the astringency of algorithm, the coefficients are introduced, as shown in Eq. (7). The agent is trained to obtain the largest accumulated reward, that is, to minimize the weighted value of ΔP and $\Delta P/\Delta t$.

$$r = -\left(\lambda_1 \Delta P + \lambda_2 \frac{\Delta P}{\Delta t}\right) \tag{7}$$

where λ_1 and λ_2 are the coefficients, which are 0.01 and 1.

4.1.4. Modified Q-learning algorithm

Traditional Q-learning algorithm used the ϵ-greedy policy to make decisions (Konda et al., 2020), as shown in Eq. (8). It can accept the optimal solution with fixed probability, which can balance the exploration and utilization with a certain degree. The simulated annealing algorithm is based on Metropolis criterion (Huo et al., 2020). It introduces cooling strategy into the algorithm instead of ϵ-greedy policy. In the initial stage, the temperature value T is high, and the probability of accepting the non-optimal solution is very high. As the temperature gradually decreases, the probability of accepting a better solution becomes higher and higher. When the temperature approaches 0, it can only accept the optimal solution. Therefore, the simulated annealing algorithm is likely to converge to the optimal solution. And it can ensure the convergence of the iteration better than the ϵ-greedy policy. The flowchart of modified Q-learning algorithm is shown in Table 3.

$$\text{prob}(a_t) = \begin{cases} 1-\varepsilon \text{ if } a = \text{argmax}Q(s_t, a_t) \\ \varepsilon \text{ others} \end{cases} \tag{8}$$

where prob is the probability of action; ϵ is the greedy, which is set to 0.1.

In the simulated annealing algorithm, the cooling strategy has a great impact on the final results. The cooling strategies commonly used are: logarithmic cooling strategy, rapid cooling strategy, straight-line cooling strategy, and isotropic cooling strategy, as shown in Eqs. (9)–(12).

$$T_k = \frac{\lambda}{\log(1+k)} T_0 \tag{9}$$

$$T_k = \frac{\lambda}{1+k} T_0 \tag{10}$$

$$T_k = \left(1 - \frac{k}{K}\right) \times \lambda T_0 \tag{11}$$

$$T_k = \lambda^k T_0 \tag{12}$$

where T_0 is the initial temperature value; k is the number of iterations; T_k is the temperature value at the k-th iteration; λ is the cooling parameter, which is the range of 0–1, it is usually set as a constant value close to 1.

4.2. Pressure difference model based on ISSA-ELM

4.2.1. The structure of ELM model

Based on the previous research (Miao et al., 2022c), the pressure difference can be used as an index for measuring flow field vibration. ELM is a feedforward neural network algorithm of single hidden layer (Huang et al., 2005), the structure is shown in Fig. 14. Its initial weight w and threshold b are randomly generated, and only the output matrix H and output weight β are calculated in the learning process. This method has strong nonlinear decoupling ability and fewer parameters. And it has faster training speed and better generalization performance than artificial neural network (ANN). The mathematical model of ELM is shown in Eq. (13).

$$t_i = \sum_{i=1}^{L} \beta_i g(w_i \cdot x_i + b_i), x_i \in R^n, w_i \in R^n, \beta_i \in R^m \tag{13}$$

where t_i is the output result; β_i is the weight of the hidden layer and the output layer; g is the activation function; w_i is the weight vector between input and output; b_i is the bias vector; x_i is sample data.

Based on the experimental data, the inputs of the ELM model are

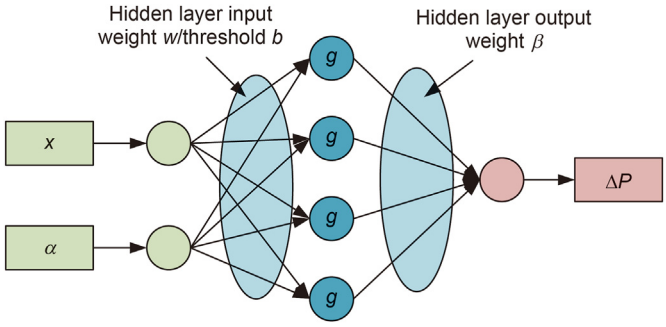

Fig. 14. The structure of ELM model.

Table 3
The flowchart of modified Q-learning algorithm.

Modified Q-learning algorithm:
1 Initialize starting temperature T_0, current temperature T, strategy π
2 For each episode, do:
3 The initial state is s_0, which is set as the current state s_t of the agent
4 Let $T_0 = T$
5 Execute each step in each cycle:
6 According to the current state s_t of the agent, randomly select an action a_r from the action set A
7 According to the current state s_t of the agent, select an action a_p according to the strategy π, and use it as the current action a_t, let $a_p = a_t$
8 Randomly generate a number δ uniformly distributed between (0, 1)
9 Calculate $p = \exp[(Q(s, a_r) - Q(s, a_p))/T]$, compare p with the random number δ according to the Metropolis criterion. If $p > \delta$, action a_r is selected as the current action a_t, $a_r = a_t$; otherwise, keep the current action unchanged
10 The agent performs action a_t, the state changes from s_t to s_{t+1}, and the agent receives the immediate reward value r_t returned by the environment
11 According to Eq. (8), the current state-action value function $Q(s, a)$ of the agent is updated using the reward value r_t obtained in the previous step
12 Determine whether the state s_{t+1} is the target state of the agent. If not, end the learning of this step, let $t = t+1$, and go to step 5; otherwise, go to the next step
13 According to the cooling strategy, the temperature T is cooled down, the learning of this cycle is over, let episode = episode+1, and step 3 is executed again
14 Execute until all desired number of cycles have been learned

the plugging state x and spoiler angle α, and the output is the pressure difference ΔP. This model is used to calculate the reward function and measure the plugging-induced vibration. Due to the randomness of initial weight w and threshold b, the ELM is difficult to achieve optimal solution. Therefore, ISSA is used to optimize the input weight matrix and the hidden layer threshold.

4.2.2. Optimization process of ISSA

SSA is a swarm intelligence optimization algorithm based on the behavior of sparrows foraging and evading predators (Xue and Shen, 2020). Spotters with high fitness values are given preferential access to food and guide the flow of the entire population. The spotters' location update formula is as follows:

$$X_{i,j}^{t+1}=\begin{cases} X_{i,j}^{t}\cdot\exp\left(\dfrac{-i}{\alpha\cdot \mathrm{Iter}_{\max}}\right), R_2<ST \\ X_{i,j}^{t}+Q\cdot L, R_2\geq ST \end{cases} \tag{14}$$

where $X_{i,j}^{t}$ is the i-th sparrow in the j dimension under the current iteration t; α is the random number, $\alpha\in(0,1]$; $\mathrm{Iter}_{\max}$ is maximum number of iterations; R_2、ST is the alert and safe value, respectively; Q is a random number, subject to standard normal distribution; L is a d-dimensional matrix with one row and all elements are 1.

The followers' position is updated as follows:

$$X_{i,j}^{t+1}=\begin{cases} Q\cdot\exp\left(\dfrac{X_{\mathrm{worst}}^{t}-X_{i,j}^{t}}{i^2}\right), i>\dfrac{n}{2} \\ X_{p}^{t+1}+\left|X_{i,j}^{t}-X_{p}^{t+1}\right|\cdot A^{+}\cdot L, i\leq\dfrac{n}{2} \end{cases} \tag{15}$$

where X_{worst} is currently the worst position overall; n is the total number of sparrows, when $i>n/2$, the i-th sparrow is very hungry; X_p is the best place for spotters; A is a one-row d-dimensional matrix with random elements of 1 or -1, $A^{+}=A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}$.

Considering its own safety and the ability to successfully obtain food, sparrows will select 10%–20% individuals from the population for reconnaissance and alert, and the location update is as follows:

$$X_{i,j}^{t+1}=\begin{cases} X_{\mathrm{best}}^{t}+\beta\left|X_{i,j}^{t}-X_{\mathrm{best}}^{t}\right|, f_i>f_g \\ X_{i,j}^{t}+k\left(\dfrac{\left|X_{i,j}^{t}-X_{\mathrm{worst}}^{t}\right|}{(f_i-f_w)+\varepsilon}\right), f_i=f_g \end{cases} \tag{16}$$

where X_{best} is currently the best position overall; β is the step correction factor, subject to standard normal distribution; f_i is the fitness of the sparrow at this time, f_w and f_g respectively represent the overall worst fitness and optimal fitness at this time. K is the random number, $k\in(0,1)$; ε is the extremely small constant, $\varepsilon=10^{-50}$.

However, SSA algorithm has some problems, such as insufficient search ability and increased probability of falling into the extreme value space (Shi et al., 2021). Therefore, the sine-cosine strategy is introduced to optimize the updating method of spotters' location, as shown in Eqs. (17) and (18). A nonlinear weight factor ω is used to adjust the dependence of the individual position update of the population on the current individual information.

$$w=\frac{e^{\frac{t}{\mathrm{Iter}_{\max}}}-1}{e-1} \tag{17}$$

$$X_{i,j}^{t+1}=\begin{cases} w\cdot X_{i,j}^{t}+r_1'\cdot\sin r_2\cdot\left|r_3\cdot X_{\mathrm{best}}-X_{i,j}^{t}\right|, R_2<ST \\ w\cdot X_{i,j}^{t}+r_1'\cdot\cos r_2\cdot\left|r_3\cdot X_{\mathrm{best}}-X_{i,j}^{t}\right|, R_2\geq ST \end{cases} \tag{18}$$

The flowchart of ISSA-ELM model is shown in Fig. 15. Firstly, input samples are normalized, and the parameters of ELM and ISSA are initialized. Secondly, the fitness of each sparrow is calculated and sorted. The optimal fitness value is the optimal position. Then, the positions of spotters and followers are updated, and the fitness after updating of each sparrow is calculated. Finally, the input weight matrix w_j and the hidden layer threshold b_j are obtained. And the prediction results are output.

5. Results and discussion

5.1. Results of cooling strategy

The four cooling strategies described in Section 4.1.4 are used to improve the Q-learning algorithm, respectively. The initial temperature is set to 1000 °C, the cooling parameter λ is set to 0.9, and the number of cycles is 2000. The number of iterations under the four cooling strategies are shown in Table 4. Through comparison, it can be seen that the rapid cooling strategy has the fastest calculation speed. And it has been verified that the algorithm can be converged through this strategy. The curve of rapid cooling strategy is shown in Fig. 16.

5.2. Prediction results of pressure difference model

The pressure difference data of different spoiler models under different plugging states is obtained from dynamic plugging experiment. A nonlinear model is established based on ISSA-ELM. 90% of the dataset is selected as the training data, and 10% of the dataset is selected as the testing data. The population size is 30. The maximum number of iterations is 200, and warning value is 0.6. The ratios of spotters and followers are 0.7 and 0.3, respectively. The proportion of the vigilantes of random distribution is 0.2. The fitness function is the mean squared error (MSE) of testing data.

The fitness value during the iterative process of algorithm is shown in Fig. 17. It can be seen that the fitness value shows a downward trend. At the iterations of 124, the fitness value reaches 3×10^{-6}, which is converged. This shows that ISSA algorithm has good convergence.

The prediction value and actual value of testing data are shown in Fig. 18. It can be seen that the prediction value is very close to the actual value, the relative error is within 0.24%. In order to evaluate the prediction accuracy of the model, the mean absolute error (MAE), the mean relative error (MRE) and the root mean square error (RMSE) are used as the evaluation indicators, as shown in Eqs. (19)–(21). The comparison of different methods is shown in Table 5. It can be seen that all the indicators of ISSA-ELM are the smallest. It indicates that the ISSA-ELM can achieve the accurate prediction for pressure difference.

$$L_{\mathrm{MAE}}=\frac{1}{N}\sum_{i=1}^{N}\left|\hat{y}_i-y_i\right| \tag{19}$$

$$L_{\mathrm{MRE}}=\frac{1}{N}\sum_{i=1}^{N}\frac{\left|\hat{y}_i-y_i\right|}{y_i} \tag{20}$$

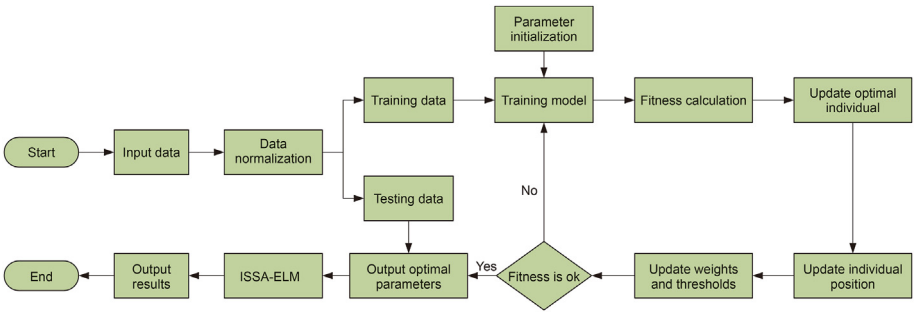

Fig. 15. The flowchart of ISSA-ELM model.

Table 4
The comparison of four cooling strategies.

Cooling strategies	Logarithmic cooling	Rapid cooling	Straight-line cooling	Isotropic cooling
Iterations	16206000	18000	44000	36000

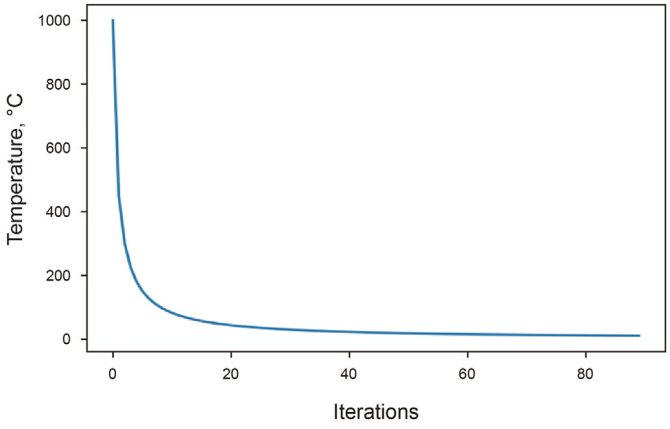

Fig. 16. The curve of rapid cooling strategy.

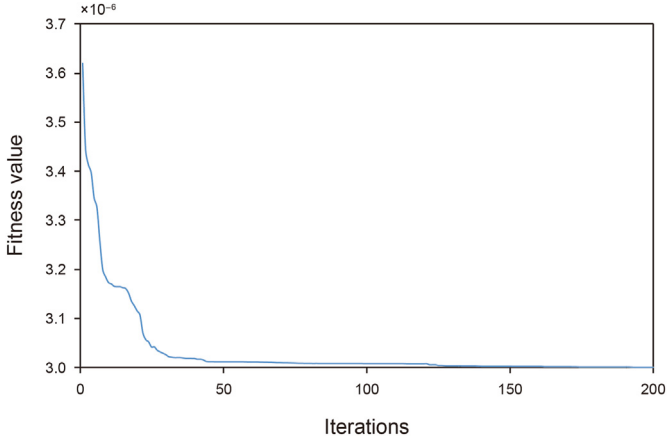

Fig. 17. The fitness value of ISSA algorithm.

$$L_{\mathrm{RMSE}} = \sqrt{\frac{\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2}{N}} \tag{21}$$

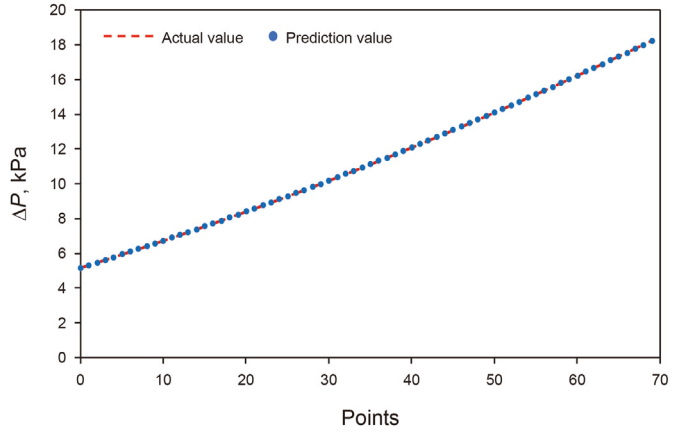

Fig. 18. The prediction result of ISSA-ELM model.

Table 5
The prediction effect of different methods.

Method	MAE	MRE	RMSE
ISSA-ELM	0.016	0.002	0.016
SSA-ELM	0.046	0.003	0.059
ELM	0.072	0.005	0.110
BP	0.422	0.060	0.738

where $\hat{y}_i$ and y_i are the predicted and actual values of pressure difference of the i-th point respectively; N is the number of samples.

5.3. Regulating strategy and experimental validation

After a series of iterations of the modified Q-learning algorithm, the optimal regulating strategy is obtained. The spoiler angle strategy and plugging velocity strategy during the plugging process are shown in Fig. 19. As described in Section 4.1.1, the dynamic plugging process is discretized with the interval of 1%. For the spoiler angle, in the early stage of plugging process, the angle value is 30°. And in the medium and later stage, the angle value is 0° and 120°, respectively. For the plugging velocity, in the plugging process of 0–79%, the plugging velocity is 0.8 mm/s. In the plugging state of 80%, the plugging velocity is the largest as 1 mm/s. In the later stage

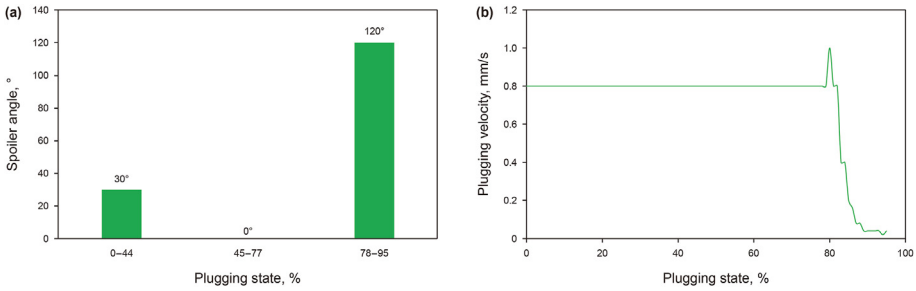

Fig. 19. The optimal regulating strategy: (**a**) The spoiler angle strategy; (**b**) The plugging velocity strategy.

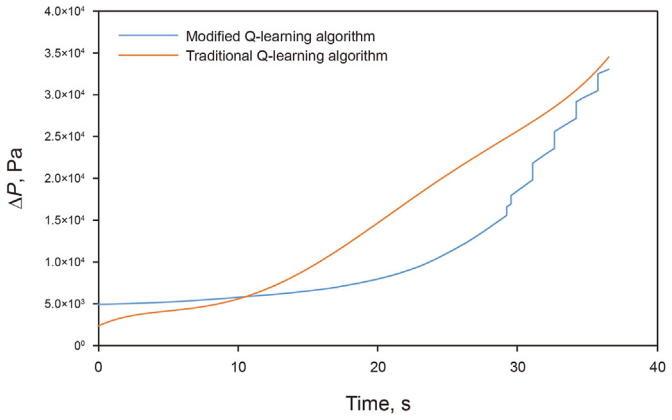

Fig. 20. The pressure difference of the optimal regulating strategy.

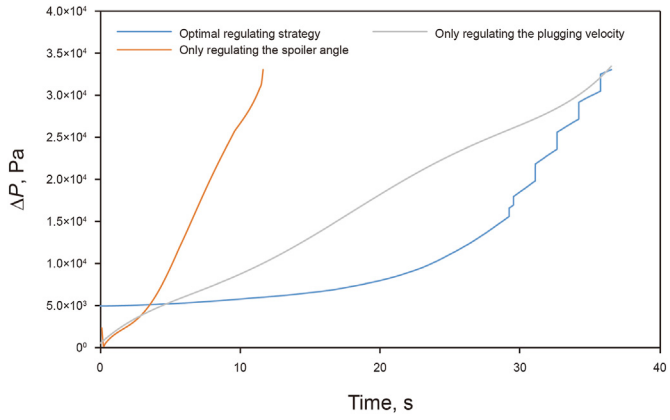

Fig. 21. The pressure difference of different methods.

of plugging operation, the plugging-induced vibration becomes more intense, so the plugging velocity selects the smaller value, and it reduces to 0.02 mm/s when the plugging is nearly completed.

The optimal regulating strategy obtained from the modified Q-learning algorithm is validated by the experiment, and it is compared with the strategy of traditional Q-learning algorithm, as shown in Fig. 20. It can be seen that in the early stage, the pressure difference of modified method is larger than traditional method. But after 20 s, the pressure difference becomes smaller than traditional method. On the whole, the pressure difference of modified method is smaller, the average value is reduced by 24.2% of traditional Q-learning algorithm.

In order to verify the advantage of the optimal regulating strategy, the proposed method is compared with single-regulating methods: only regulating the spoiler angle (the plugging velocity is 2 mm/s) and only regulating the plugging velocity (the spoiler angle is 0°), as shown in Fig. 21 and Table 6. It can be seen from the experimental results of the three methods, through regulating the plugging velocity, the changing rate of pressure difference can be reduced. And through regulating the spoiler angle, the pressure difference can be reduced. Through comparison, the optimal regulating strategy performs better in reducing the plugging-induced vibration than other two methods. For the maximum of ΔP, the three methods have a small gap. But for the average value of ΔP, the proposed method has reduced by 19.9% and 32.7%. For the changing rate of pressure difference, if the plugging process is a uniform movement, the maximum and average value are much higher. And the proposed method is the smallest. Therefore, the dynamic plugging regulating strategy can reduce the plugging-induced vibration, ensuring the plugging process more stable and safer.

6. Conclusions

This study proposes a dynamic regulating strategy for the plugging process based on reinforcement learning. Through numerical analysis and experiments, the flow field vibration caused by plugging operation is gradually serious during the plugging process. And from the results of the simulations and experiments, the PIPR's spoiler angle and plugging velocity can affect the pressure difference and its changing rate, which can influence the plugging-induced vibration. Therefore, a dynamic regulating strategy based on the modified Q-learning algorithm is developed to regulate the spoiler angle and plugging velocity in real time. According to the experimental results, the pressure difference model based on ISSA-ELM is established to obtain the relationship between spoiler angle, plugging state and pressure difference. The prediction results show that the relative error is within 0.24%, which performs better than other methods. The optimal regulating strategy is validated by the experiments. The results indicate that the average pressure difference is reduced by 24.2% compared with traditional Q-learning algorithm. And the proposed method has reduced by 19.9% and 32.7% of single-regulating methods, the changing rate of pressure difference is also greatly reduced. So the regulating strategy consisted of the spoiler angle and plugging velocity is better than single-regulating methods. This study provides a novel approach for reducing the vibration of PIPR, which can ensure sufficient safety during pipeline plugging operation. In addition, the proposed method can be used to guide the pipeline maintenance, which is of great significance for preventing environmental pollution caused by pipeline leakage.

Table 6
The changing rate of pressure difference of different methods.

Method	Maximum of ΔP, Pa	Average value of ΔP, Pa	Maximum of $\Delta P/\Delta t$, Pa/s	Average value of $\Delta P/\Delta t$, Pa/s
Optimal regulating strategy	33077	11122	796	384
Only regulating the spoiler angle	33073	13877	6892	1331
Only regulating the plugging velocity	33498	16530	886	451

Declaration of competing interest

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

This work was financially supported by the National Natural Science Foundation of China (Grant No. 51575528), the Science Foundation of China University of Petroleum, Beijing (No.2462022QEDX011).

References

Andrea, G., Juan, G., Bartek, L., et al., 2020. Design of an active vision system for high-level isolation units through Q-Learning. Applied Sciences-Basel 10 (17), 5927. https://doi.org/10.3390/app10175927.

Goharimanesh, M., Mehrkish, A., Janabi-Sharifi, F., 2020. A fuzzy reinforcement learning approach for continuum robot control. J. Intell. Rob. Syst. 100 (3–4), 809–826. https://doi.org/10.1007/s10846-020-01237-6.

Huang, G., Zhu, Q., Siew, C., 2005. Extreme learning machine: theory and applications. Neurocomputing 70, 489–501. https://doi.org/10.1016/j.neucom.2005.12.126.

Huo, L., Zhu, J., Wu, G., et al., 2020. A novel simulated annealing based strategy for balanced UAV task assignment and path planning. Sensors 20 (17), 4769. https://doi.org/10.3390/s20174769.

Ignacio, C., Mariano, D., Gerardo, G., 2020. An adaptive deep reinforcement learning approach for MIMO PID control of mobile robots. ISA (Instrum. Soc. Am.) Trans. 102, 280–294. https://doi.org/10.1016/j.isatra.2020.02.017.

Konda, R., La, H., Zhang, J., 2020. Decentralized function approximated Q-Learning in multi-robot systems for predator avoidance. IEEE Rob. Autom. Lett. 5 (4), 6342–6349. https://doi.org/10.1109/lra.2020.3013920.

Lie, R.G., Muangsuankwan, N., 2015. Remote-controlled plugging technology minimizes platform downtime: valve replacement through SmartPlug® isolation. In: ASME 2015 India International Oil and Gas Pipeline Conference. American Society of Mechanical Engineers. https://doi.org/10.1115/IOGPC2015-7910.

Miao, X., Zhao, H., 2022. Regulating control of in-pipe intelligent isolation plugging tool based on adaptive dynamic programming. J. Pipeline Syst. Eng. Pract. 13 (2), 04022003. https://doi.org/10.1061/(ASCE)PS.1949-1204.0000635.

Miao, X., Zhao, H., Gao, B., et al., 2022a. Motion analysis and control of the pipeline robot passing through girth weld and inclination in natural gas pipeline. J. Nat. Gas Sci. Eng. 104, 104662. https://doi.org/10.1016/j.jngse.2022.104662.

Miao, X., Zhao, H., Song, F., et al., 2022b. Dynamic characteristics and motion control of pipeline robot under deformation excitation in subsea pipeline. Ocean Eng. 266, 112790. https://doi.org/10.1016/j.oceaneng.2022.112790.

Miao, X., Zhao, H., Gao, B., et al., 2022c. Vibration reduction control of in-pipe intelligent isolation plugging tool based on deep reinforcement learning. Int. J. Precision Eng. Manufactur. Green Technol. 9 (1), 1477–1491. https://doi.org/10.1007/s40684-021-00405-9.

Mirshamsi, M., Rafeeyan, M., 2015. Dynamic analysis and simulation of long pig in gas pipeline. J. Nat. Gas Sci. Eng. 23, 294–303. https://doi.org/10.1016/j.jngse.2015.02.004.

Rahman, M., Rashid, S., Hossain, M., 2018. Implementation of Q learning and deep Q network for controlling a self-balancing robot model. Robotics and biomimetics 5, 1–6. https://doi.org/10.1186/s40638-018-0091-9.

Rai, A., Kim, J.M., 2021. A novel pipeline leak detection approach independent of prior failure information. Measurement 167, 108284. https://doi.org/10.1016/j.measurement.2020.108284.

Shi, M., Liang, Y., Qin, L., et al., 2021. Prediction method of ball valve internal leakage rate based on acoustic emission technology. Flow Meas. Instrum. 81, 102036. https://doi.org/10.1016/J.FLOWMEASINST.2021.102036.

Tveit, E., Aleksandersen, J., 2000. Remote controlled (Tether-Less) high pressure isolation system. In: SPE Asia Pacific Oil and Gas Conference and Exhibition, 8. Society of Petroleum Engineers, Brisbane, Australia. https://doi.org/10.2118/64513-MS.

Wang, W., Guo, J., Fan, J., et al., 2020. Research on the proportional-integral-derivative synchronous control method of the marine spherical isolation plug in the rotation process. Proc. Inst. Mech. Eng., Part M: J. Eng. Marit. Environ. 234 (4), 810–819. https://doi.org/10.1177/1475090220913704.

Wu, T., Zhao, H., 2019. An energy-saving and velocity-tracking control design for the pipe isolation tool. Adv. Mech. Eng. 11 (4), 1–16. https://doi.org/10.1177/1687814019845949.

Wu, T., Zhao, H., Gao, B., et al., 2021a. Energy-saving for a velocity control system of a pipe isolation tool based on a reinforcement learning method. Int. J. Precision Eng. Manufactur. Green Technol. 9 (1), 225–240. https://doi.org/10.1007/s40684-021-00309-8.

Wu, T., Zhao, H., Gao, B., et al., 2021b. Structural optimization strategy of pipe isolation tool by dynamic plugging process analysis. Petrol. Sci. 18 (6), 225–240. https://doi.org/10.1016/j.petsci.2021.09.010.

Xue, J., Shen, B., 2020. A novel swarm intelligence optimization approach: sparrow search algorithm. Systems Science & Control Engineering 8 (1), 22–34. https://doi.org/10.1080/21642583.2019.1708830.

Yan, H., Wang, L., Li, P., et al., 2020. Research on passing ability and climbing performance of pipeline plugging robots in curved pipelines. IEEE Access 8, 173666–173680. https://doi.org/10.1109/ACCESS.2020.3025560.

Zhao, H., Hu, H.R., 2017. Optimal design of a pipe isolation plugging tool using a computational fluid dynamics simulation with response surface methodology and a modified genetic algorithm. Adv. Mech. Eng. 9 (10), 1–12. https://doi.org/10.1177/1687814017715563.

Zhang, K., Ding, Q.X., Liu, S.H., et al., 2018. Research on pressure fluctuation phenomenon using the smart isolation tool in subsea pipeline maintenance operation. Proc. Inst. Mech. Eng., Part M: J. Eng. Marit. Environ. 233 (2), 643–652. https://doi.org/10.1177/1475090218787918.