



## Original Paper

# Optimization of plunger lift working systems using reinforcement learning for coupled wellbore/reservoir

Zhi-Sheng Xing<sup>a</sup>, Guo-Qing Han<sup>a,\*</sup>, You-Liang Jia<sup>b</sup>, Wei Tian<sup>b</sup>, Hang-Fei Gong<sup>b</sup>,  
Wen-Bo Jiang<sup>a</sup>, Pei-Dong Mai<sup>a</sup>, Xing-Yuan Liang<sup>a</sup>

<sup>a</sup> College of Petroleum Engineering, China University of Petroleum (Beijing), Beijing, 102249, China

<sup>b</sup> Oil and Gas Technology Institute, Changqing Oilfield CNPC, Xi'an, 710018, Shaanxi, China



## ARTICLE INFO

## Article history:

Received 15 August 2024

Received in revised form

9 March 2025

Accepted 10 March 2025

Available online 10 March 2025

Edited by Yan-Hua Sun

## Keywords:

Plunger lift

Liquid loading

Deliquification

Reinforcement learning

Deep deterministic policy gradient (DDPG)

Artificial intelligence

## ABSTRACT

In the mid-to-late stages of gas reservoir development, liquid loading in gas wells becomes a common challenge. Plunger lift, as an intermittent production technique, is widely used for deliquification in gas wells. With the advancement of big data and artificial intelligence, the future of oil and gas field development is trending towards intelligent, unmanned, and automated operations. Currently, the optimization of plunger lift working systems is primarily based on expert experience and manual control, focusing mainly on the success of the plunger lift without adequately considering the impact of different working systems on gas production. Additionally, liquid loading in gas wells is a dynamic process, and the intermittent nature of plunger lift requires accurate modeling; using constant inflow dynamics to describe reservoir flow introduces significant errors. To address these challenges, this study establishes a coupled wellbore–reservoir model for plunger lift wells and validates the computational wellhead pressure results against field measurements. Building on this model, a novel optimization control algorithm based on the deep deterministic policy gradient (DDPG) framework is proposed. The algorithm aims to optimize plunger lift working systems to balance overall reservoir pressure, stabilize gas–water ratios, and maximize gas production. Through simulation experiments in three different production optimization scenarios, the effectiveness of reinforcement learning algorithms (including RL, PPO, DQN, and the proposed DDPG) and traditional optimization algorithms (including GA, PSO, and Bayesian optimization) in enhancing production efficiency is compared. The results demonstrate that the coupled model provides highly accurate calculations and can precisely describe the transient production of wellbore and gas reservoir systems. The proposed DDPG algorithm achieves the highest reward value during training with minimal error, leading to a potential increase in cumulative gas production by up to 5% and cumulative liquid production by 252%. The DDPG algorithm exhibits robustness across different optimization scenarios, showcasing excellent adaptability and generalization capabilities.

© 2025 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

As gas reservoir development enters its mid-to-late stages, liquid loading becomes a primary challenge for most natural gas wells (Tan B. et al., 2023; Zhu et al., 2019). Plunger lift is one of the most effective techniques for deliquification, utilizing the well's produced fluids to lift the plunger and the liquid above it to the surface (Tan X. et al., 2023; Zhao et al., 2018). The plunger acts as a barrier between gas and liquid, minimizing gas slippage and liquid fallback. With the advancement of artificial intelligence and digital

plunger control systems, gas well production is increasingly moving towards intelligent operations and digital transformation (Chen et al., 2024; Xie et al., 2023; Zhao and Bai, 2018). Unconventional gas reservoirs, primarily consisting of shale gas and tight gas, exhibit characteristics such as rapid energy depletion and significant variations in gas and liquid production capacity (Gupta et al., 2017; Sayman et al., 2022a). In most cases, liquid loading in gas wells during the later stages of production is a dynamic process. Therefore, deliquification techniques, represented by plunger lift, must achieve self-optimization and self-regulation. Currently, the optimization of plunger lift working systems is gradually evolving towards intelligence and automation. Deep exploration of data collected by digital plunger control systems can significantly

\* Corresponding author.

E-mail address: [hanguoqing@163.com](mailto:hanguoqing@163.com) (G.-Q. Han).

| Nomenclature |                                              | $\rho$                             | Density, kg/m <sup>3</sup>          |
|--------------|----------------------------------------------|------------------------------------|-------------------------------------|
| <i>A</i>     | Area, m <sup>2</sup>                         | <i>Subscripts and superscripts</i> |                                     |
| <i>a</i>     | Acceleration, m/s <sup>2</sup>               | <i>c</i>                           | Casing                              |
| <i>C</i>     | Drag coefficient                             | <i>cb</i>                          | The bottom of casing                |
| <i>d</i>     | Diameter, m                                  | <i>fric</i>                        | Friction                            |
| <i>f</i>     | Friction factor                              | <i>g</i>                           | Gas                                 |
| <i>g</i>     | Gravitational acceleration, m/s <sup>2</sup> | <i>gc</i>                          | Casing gas                          |
| <i>H</i>     | Height, m                                    | <i>gt</i>                          | Tubing gas                          |
| <i>M</i>     | Molecular weight, kg/mol                     | <i>l</i>                           | Liquid                              |
| <i>m</i>     | Mass, kg                                     | <i>lc</i>                          | Casing liquid column                |
| <i>P</i>     | Pressure, Pa                                 | <i>lt</i>                          | Tubing liquid column, tubing liquid |
| <i>q</i>     | Production, m <sup>3</sup> /s                | <i>p</i>                           | Plunger                             |
| <i>r</i>     | Reward                                       | <i>pb</i>                          | Lower part of the plunger           |
| <i>R</i>     | Gas constant, J/mol/K                        | <i>pt</i>                          | Upper part of the plunger           |
| <i>T</i>     | Temperature, K                               | <i>r</i>                           | Reservoir                           |
| <i>v</i>     | Velocity, m/s                                | <i>t</i>                           | Tubing                              |
| <i>Z</i>     | Z factor                                     | <i>tb</i>                          | The bottom of tubing                |
| $\gamma$     | Relative density                             | <i>wf</i>                          | Bottomhole                          |

enhance lifting efficiency and reduce operational costs (Barros et al., 2018; Zhao K. et al., 2021). The ability to automatically adjust parameters and working systems in real-time reduces the laboratory intensity of field management personnel, aligns with the development trends of modern digital oilfields, and significantly improves the rationality of working systems and lifting efficiency (Hashmi et al., 2018; Sayman et al., 2022b; Tong et al., 2017).

The effectiveness of plunger lift application largely depends on the rationality of the working systems. A well-designed working system ensures that more liquid is expelled during a short shut-in period while maintaining high production rates during the after-flow period. The optimization of plunger lift working systems primarily focuses on adjusting the timing of well opening and closing cycles. The most commonly used method for plunger lift optimization is based on the empirical model proposed by Foss and Gaul (1965), which was developed using field data from the Ventura Avenue gas field in the United States. This model predicts key operating parameters, such as maximum casing pressure, minimum casing pressure, tubing pressure, lifted liquid volume, cycle frequency, and required gas injection volume, based on high gas–liquid ratio plunger lift well production data. However, since plunger lift is an intermittent production technique, significant variations in reservoir pressure occur during different production stages (after-flow and shut-in periods). Using steady-state inflow performance relationship (IPR) curves, which describes the relationship between bottomhole flowing pressure and production rate, to describe gas well production dynamics and reservoir inflow during these stages introduces substantial calculation errors by neglecting the transient gas–liquid flow near the wellbore. Consequently, many researchers have developed coupled reservoir–wellbore models to describe the transient production behavior of gas wells. For instance, Ozkan et al. (2003) proposed a coupled reservoir–wellbore optimization algorithm for plunger lift based on Laplace transforms and the Duhamel principle, which established an analytical model of wellbore and reservoir flow. Hu et al. (2007) developed an implicit coupling model between wellbore flow and near-wellbore reservoir flow, which simulates various transient flow characteristics under the mutual influence of wellbore and reservoir dynamics. Parsa et al. (2013) employed the optimization algorithm proposed by Ozkan et al. (2003), which focuses on maximizing gas production and ensuring that the pressure buildup during shut-in period is sufficient to lift the plunger to the surface. Hashmi et al. (2018) established a

minimum casing pressure calculation model based on energy balance during the plunger lift operation, ensuring continuous production from the well. Xiang and Kabir (2019) addressed the errors in conventional steady-state IPR models by solving the diffusion equation under different production cycles and using Bessel functions to reduce the discrepancies caused by low reservoir permeability and sinusoidal oscillations in bottomhole flow pressure. They developed a transient IPR inflow dynamic relationship model for plunger lift wells. Peng et al. (2024) developed a highly coupled wellbore–reservoir simulator to determine liquid loading in vertical and inclined gas wells. The model uses drift-flux analysis to predict flow pattern transitions, with wellhead pressure or flow rate constraints as boundary conditions. It accurately describes the gas–liquid dynamic flow within the wellbore and reservoir, revealing significant changes in gas production, water production, and flow pattern during liquid loading, including liquid film reversal. In typical gas well reservoir descriptions, the reservoir is often considered infinite. For intermittent production techniques, the pressure drop funnel within the well’s control area exhibits significant inner and outer zone characteristics. These characteristics are influenced by periodic liquid loading and deliquification operations. The pressure distribution in the outer zone only changes significantly when long-term equivalent bottomhole flowing pressure changes, while the pressure distribution in the inner zone exhibits cyclic changes between positive and reverse funnels.

Currently, adjustments of plunger lift working systems are primarily based on dynamic analysis by field experts, with manual parameter adjustments made intermittently. Most methods are predominantly experience-based, resulting in suboptimal optimization performance and high trial-and-error costs. With the development of big data and artificial intelligence, traditional manual optimization methods are gradually being replaced by data-driven models. Machine learning (ML) techniques are increasingly applied in various fields to address complex engineering problems. Many researchers have used AI technologies to process large datasets and predict dynamic production changes at different stages of gas well production, which serves as a basis for optimizing production. Kamari et al. (2016) utilized least squares support vector machines (LSSVM) to predict the maximum liquid lifting capacity of plunger lift systems. They proposed an optimal parameter adjustment mechanism based on coupled simulation annealing (CSA) and integrated CSA-LSSVM with outlier detection

models like Hat matrix and Williams plot. Singh (2017) developed a fault diagnosis tool for plunger lift based on the classification and regression tree (CART) algorithm, considering 8–10 parameters influencing plunger lift operations. Nandola et al. (2018) introduced a binary decision-making process for plunger lift cycles, converting surface measurement time series data into cycle-wise outputs and state variables into continuous thresholds, which were then used in a reduced-order cycle-to-cycle model for time-domain optimization. Aires et al. (2020) used a fuzzy controller and a linear model predictive controller to simulate plunger operations. The fuzzy controller adjusted plunger velocity by controlling the wellhead secondary valve, maintaining optimal plunger velocity. Akhiartdinov et al. (2020) used machine learning to simulate the plunger lift process, using field measurement parameters as inputs and instantaneous gas production as outputs to train a feedforward artificial neural network (ANN) model, which was developed as a virtual flow meter. The model predicted optimal plunger lift timings, optimizing the plunger lift production regime. Romero et al. (2020) introduced software that classifies plunger lift faults using machine learning, specifically focusing on production loss detection through a neural network model. Xie et al. (2023) trained a transformer-based encoder to identify cycle points in continuous data and proposed an unsupervised clustering method using a deep neural network autoencoder for optimizing clustering loss and reconstruction loss, applied to fault condition identification in plunger lift operations. Zhong et al. (2024) analyzed production data using long short-term memory (LSTM) models and convolutional neural networks (CNNs) to identify time-series features in plunger lift production data. They compared the regression accuracy and computational precision of different models and identified the ABILSTM model as the best for high-precision minute-level predictions. Wu et al. (2024) developed a wellbore–reservoir coupled intermittent production well model using a transient simulator and optimized the intermittent production regime using genetic algorithm (GA) and proximal policy optimization (PPO) methods, which effectively enhanced gas production and liquid unloading. Zhu et al. (2024) proposed a data-driven gas well classification method based on the linear discriminant analysis-discriminant analysis (LDA-DA) model, evaluating gas wells from liquid removal capacity (LDC) and liquid removal intensity (LPI) perspectives. Chen et al. (2024) developed a semi-supervised learning model for classifying the severity of liquid loading in gas wells, combining the strengths of supervised and unsupervised learning to fully utilize the entire dataset with minimal labeling. They also introduced a dynamic time relation self-attention (DTRSA) module, enhancing the model's resistance to interference from other anomalies and improving the accuracy of liquid loading severity classification.

Machine learning can predict gas and liquid production under different plunger lift working systems and optimize operating parameters. Typical ML algorithms require millions of data points to successfully identify and predict plunger lift optimization patterns. However, if reservoir conditions or surface production parameters change, the inability to adjust rapidly can cause reservoir damage. For example, if produced water is not promptly removed, it can accumulate at the well bottom, increasing bottomhole flowing pressure and reducing gas production, potentially leading to well shutdown due to water encroachment. Furthermore, due to the changing and complex environmental conditions, the control model may require frequent adjustments to adapt to new conditions, which is not ideal for intelligent control. Plunger lift is an intermittent production technique, where reservoir inflow characteristics differ significantly between the production and shut-in stages. Using a constant inflow performance relationship to describe reservoir flow dynamics introduces significant errors. To achieve efficient and

sustained gas well production, while ensuring accurate transient production calculations under changing environmental conditions, this study adopts an integrated wellbore–reservoir coupling model, combining reservoir simulation results with the DDPG reinforcement learning model. The optimization algorithm for plunger lift working systems is established with the goals of balancing overall reservoir pressure, stabilizing gas–water ratios in individual wells, and maximizing gas production. By intelligently controlling and optimizing the plunger lift working systems, this method can enhance deliquification efficiency, allowing gas wells to operate continuously, efficiently, and in accordance with production rules, thereby improving overall development effectiveness and economic benefits.

## 2. Wellbore–reservoir coupling model development

### 2.1. Wellbore model

#### 2.1.1. Transient plunger motion

The plunger lift system, as a specialized intermittent production process, operates by utilizing the produced fluids within the well to lift the plunger and the overlying liquid to the surface. The production process primarily involves four stages: plunger upstroke, after-flow, plunger downstroke, and shut-in. The plunger upstroke stage is further divided into the gas discharge stage from the wellbore and the liquid discharge stage at the wellhead. Upon opening the wellhead valve, the gas above the plunger is expelled, leading to a rapid pressure drop above the plunger, initiating its ascent. The significant pressure difference across the plunger at the moment of well opening causes a sudden increase in plunger velocity. As the plunger velocity increases, the friction between the plunger, the liquid slug above it, and the tubing wall also rises. With continued ascent, the pressure beneath the plunger decreases, resulting in a gradual deceleration. The plunger velocity stabilizes when the pressure below it balances the pressure above and the frictional forces. Concurrently, the bottomhole flowing pressure decreases, causing the annular gas to expand and the liquid in the annulus to be displaced into the tubing. Once all the annular liquid has been displaced, the gas begins to enter the tubing, with the plunger ascent driven by both the reservoir gas flow and the expanding annular gas. The liquid column above the plunger is expelled at the wellhead, reducing the pressure above the plunger and causing a brief increase in both tubing pressure and plunger velocity. The plunger upstroke stage concludes when the plunger reaches the catcher at the wellhead. The plunger acceleration during ascent, especially when far from the wellhead, can be calculated using momentum equations based on the force balance acting on the plunger in the vertical direction.

$$a_p = \frac{(P_{pb} - P_{pt} - P_{fric})A_t}{m_p + m_{lt}} - g \quad (1)$$

The plunger downstroke stage consists of falling through gas and then through liquid. Upon closing the wellhead valve, the plunger descends under gravity. During the falling through gas stage, the plunger is primarily influenced by gravity, the upward thrust of reservoir gas, and frictional forces, with gravity being the dominant factor, resulting in accelerated motion. When the plunger contacts the liquid column in the tubing, it transitions to the falling through liquid stage, where the forces of gravity, buoyancy, and friction come into equilibrium, causing the plunger to descend at a near-constant velocity. The plunger downstroke stage concludes when the plunger impacts the bottomhole seating. The descent velocity of a plunger can be calculated using the formula provided

by Zhao Q. et al. (2021).

$$v_p = \frac{C_d}{\sqrt{\rho_q}} \cdot \frac{A_p}{A_t} \sqrt{\frac{2gM_p}{A_t}} \quad (2)$$

### 2.1.2. Wellbore gas–liquid mass variation

Most plunger lift wells lack packers, leading to issues such as gas–liquid separation and variations in liquid column height within the tubing and annulus over time. The gas and liquid masses within the tubing and casing at any given moment can be calculated using the following equations. During the plunger downstroke and shut-in stages, with the wellhead valve closed, no gas or liquid can exit the wellbore, so both  $m_{gout}$  and  $m_{lout}$  are zero.

$$m_{gt}^{j+1} = m_{gt}^j + m_{gtin} - m_{gtout} \quad (3)$$

$$m_{gc}^{j+1} = m_{gc}^j + m_{gcin} \quad (4)$$

$$m_{lt}^{j+1} = m_{lt}^j + m_{ltin} - m_{ltout} \quad (5)$$

$$m_{lc}^{j+1} = m_{lc}^j + m_{lcin} \quad (6)$$

The total gas and liquid mass entering the tubing and casing from the reservoir at any given time should equal the gas and liquid mass entering the wellbore from the reservoir at that time step.

$$m_{gres} = m_{gtin} + m_{gcin} \quad (7)$$

$$m_{lres} = m_{ltin} + m_{lcin} \quad (8)$$

The tubing pressure and bottomhole flowing pressure at the tubing shoe can be calculated using the following equations.

$$P_t = m_{gt} \frac{Z_t R T_t}{A_t (H - H_{lt}) M_g} \quad (9)$$

$$P_{tb} = P_t \exp\left(\frac{0.03415 \gamma_g (H - H_{lt})}{Z_t T_t}\right) \quad (10)$$

$$P_{wft} = P_{tb} + \rho_l g H_{lt} \quad (11)$$

The casing pressure and bottomhole flowing pressure at the casing shoe can also be calculated similarly.

$$P_c = m_{gc} \frac{Z_c R T_c}{A_c (H - H_{lc}) M_g} \quad (12)$$

$$P_{cb} = P_c \exp\left(\frac{0.03415 \gamma_g (H - H_{lc})}{Z_c T_c}\right) \quad (13)$$

$$P_{wfc} = P_{cb} + \rho_l g H_{lc} \quad (14)$$

It should be noted that the bottomhole flowing pressures at the tubing shoe and casing shoe must be equal at any given moment. Therefore, Eq. (11) should equal Eq. (14), and an iterative method is employed to calculate the dynamic changes in tubing pressure, casing pressure, and wellbore parameters at different time steps.

## 2.2. Reservoir model

In intermittent production systems, the pressure drawdown funnel within the gas well control area exhibits distinct inner and

outer zone characteristics. During the shut-in stage of the plunger lift well, the shut-in process disturbs the reservoir near the wellbore, while the outer zone pressure remains at the formation pressure. The pressure distribution in the inner zone exhibits cyclic changes between positive and negative funnels, with the boundary between the inner and outer zones maintaining a long-term equivalent high pressure. Therefore, it is crucial to determine the range of the reverse pressure funnel near the wellbore during the shut-in phase and incorporate it as a factor in the calculation process. To simulate the inner and outer zone characteristics of the pressure drawdown funnel during intermittent production, a numerical reservoir model was developed using CMG simulation software, as shown in Fig. 1. The optimization area is located in a tight sandstone gas reservoir within the Ordos Basin, characterized by a gentle westward-dipping monocline with simple structural features and a formation dip angle generally not exceeding 1°. The effective porosity of the reservoir ranges from 0.08% to 20.92%, with permeability ranging from  $0.003 \times 10^{-3}$  to  $20.33 \times 10^{-3} \mu m^2$ , indicating a wide distribution but relatively low average porosity and permeability, at 4.57% and  $0.1 \times 10^{-3} \mu m^2$ , respectively. The natural gas is of high quality, containing no hydrogen sulfide, with methane comprising 95% of the gas composition and an average specific gravity of 0.585. The reservoir lacks edge and bottom water, and its development energy primarily relies on the gas's own expansion. The Upper Paleozoic Formation is mainly an elastic-driven tight lithologic gas reservoir. Relative permeability curves for gas and water were normalized using experimental data from three core samples in the target area. The irreducible water saturation of this reservoir is 47.5%, with an average residual oil saturation of 10.5%, and the saturation at the isopoint is approximately 69.5%.

The block contains 15 producing gas wells, labeled T001–T015, with test results indicating low water production despite multiple wells producing water. Water production is high initially but stabilizes over time, with gas production either increasing or stabilizing, and the water–gas ratio (WGR) generally remains below  $0.5 \text{ m}^3/10^4 \text{ m}^3$ . This study focuses primarily on well T001. Given the short duration of the intermittent production cycle in plunger lift systems, with shut-in times typically ranging from 1 to 6 h, it is important to consider the grid block size in numerical simulation. If the grid is too large, the pressure wave may not propagate to the next grid block within a short time step, resulting in significant simulation errors. Therefore, local grid refinement was applied near well T001 to capture the inner and outer zone characteristics of the pressure drawdown funnel.

## 2.3. Coupling model

For the wellbore–reservoir coupling model, the wellbore calculations were performed using a custom model that provides easier control during the coupling process. The wellbore model was integrated as the main function in the simulation, with the reservoir model serving as the boundary condition for pressure and flow rates. The numerical coupling between the two models was implemented using an implicit scheme. The plunger upstroke and after-flow stages correspond to the reservoir model's well production mode, while the plunger downstroke and shut-in stages correspond to the reservoir shut-in mode. As shown in Fig. 2, the wellbore model controls the plunger lift operation based on the specified production schedule (after-flow and shut-in times). The wellbore model calculates the plunger state, gas–liquid distribution, and pressure distribution at each time step, transmitting the bottomhole flowing pressure to the reservoir model. The reservoir model returns the gas and liquid production rates for the wellbore calculations. To ensure consistency in fluid PVT properties under varying temperature and pressure conditions, a PVT TABLE was

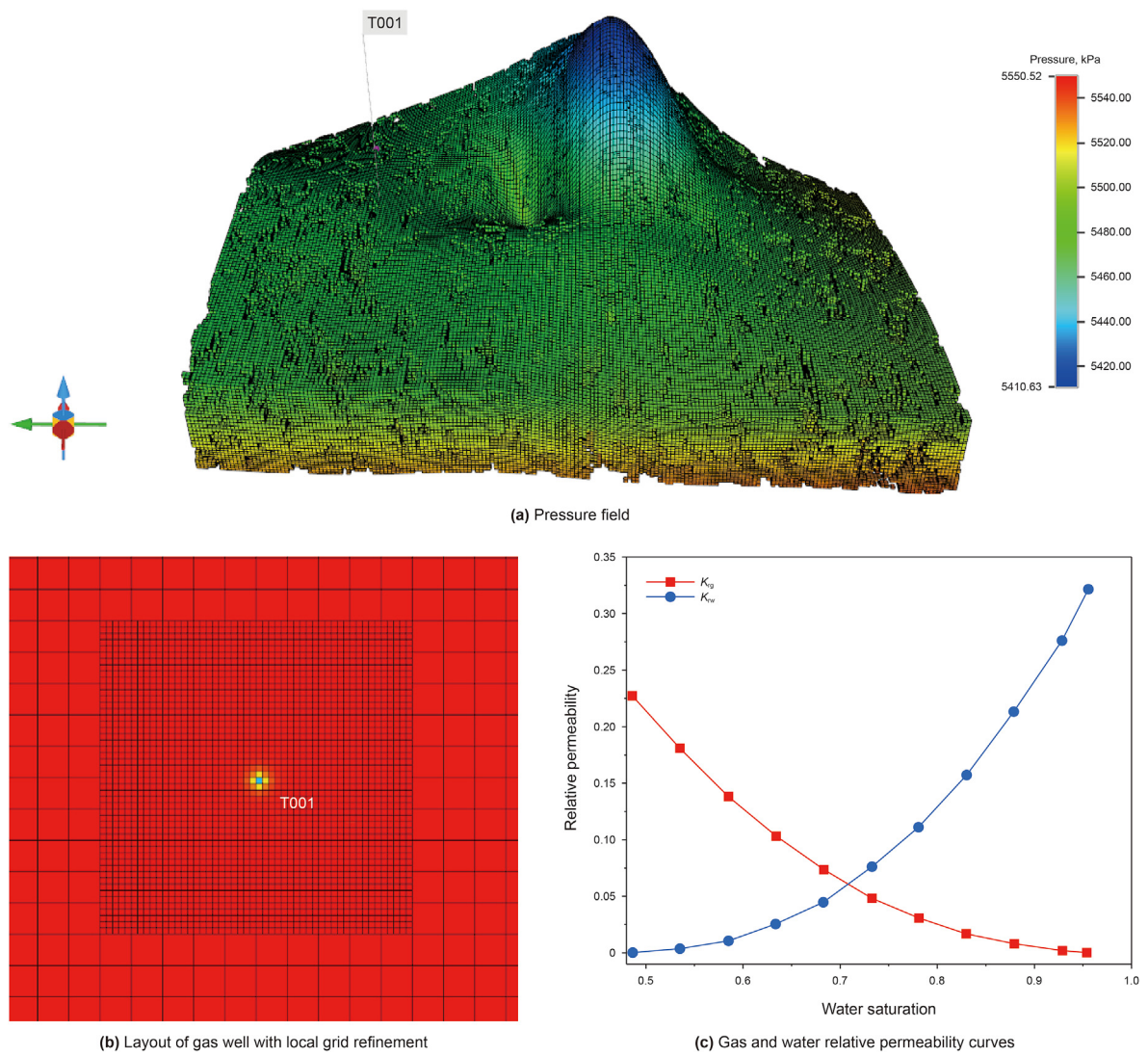


Fig. 1. Reservoir model established using CMG.

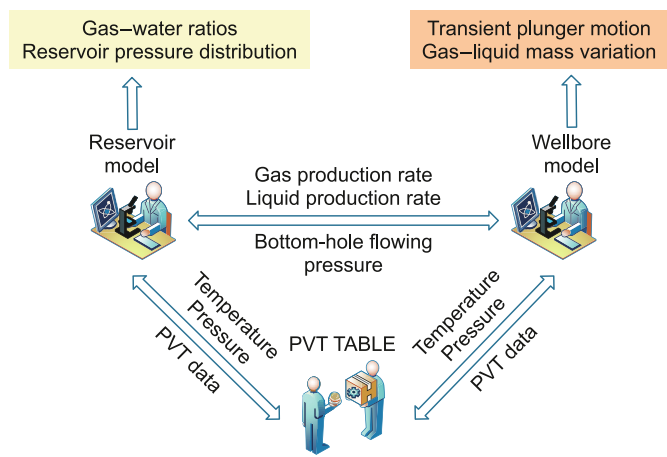


Fig. 2. Coupling model workflow diagram.

used to store fluid properties in advance, allowing the required fluid properties to be retrieved during the coupling calculations.

**3. Optimization of plunger lift working systems using reinforcement learning**

*3.1. Reinforcement learning model for optimizing plunger lift working systems*

*3.1.1. Methodology*

Machine learning can be categorized into three main types based on feedback mechanisms: supervised learning, unsupervised learning, and reinforcement learning (RL). In recent years, deep learning (DL) has significantly advanced, largely due to the superior feature representation capabilities of deep neural networks. These developments have resolved many challenges in both academia and industry, leading to notable research outcomes. Among these, RL stands out as a crucial approach for sequential decision-making, endowing agents with self-supervised learning abilities, enabling them to interact with their environment and continuously refine strategies based on received rewards. In a discrete action space, neural networks are employed to approximate a discrete distribution, which represents the probability of selecting each action. In contrast, in a continuous action space, neural networks approximate the parameters of a probability density function, allowing

policy gradient algorithms to efficiently handle tasks in high-dimensional or continuous action spaces. By optimizing these parameters, policies are directly updated to maximize the expected cumulative reward. Compared to value function-based algorithms, policy gradient methods are simpler and exhibit better convergence properties, although they are often characterized by higher variance, slower convergence rates, and challenges in determining appropriate learning rates. To accelerate the learning rate of policy gradient methods, deterministic policy gradient (DPG) algorithms have been proposed. Unlike stochastic policy gradient methods, DPG associates a single action with each state, which significantly speeds up the computation during gradient updates, reduces data requirements, and diminishes reliance on sampling. This algorithm, while efficient, does not require extensive data sampling in the action space. The earliest DPG algorithms, however, employed linear function approximators, resulting in suboptimal performance. The deep deterministic policy gradient (DDPG) algorithm is an enhancement of the deep Q-network (DQN) algorithm, effectively extending DQN to multidimensional continuous action spaces by integrating the DPG algorithm. The DDPG algorithm is based on the actor-critic (AC) framework, combining deep neural networks with the actor-critic structure. It enhances the stability of the training process by introducing noise to the agent's actions and employing an experience replay mechanism. The actor network observes the state of the environment through artificial neural networks, which are used to determine the selection probabilities of the agent's actions, thereby facilitating interaction with the environment. The network's parameters are then updated based on the received rewards to maintain and refine the agent's action-selection strategy. The critic network evaluates the value of each state-action pair by observing the state of the environment and the agent's actions, updating the neural network based on the discrepancy between the actual and predicted values.

One of the core tasks of the DDPG algorithm in deep reinforcement learning is the design of the actor-critic network structure. The DDPG algorithm consists of two networks: the actor network and the critic network, with parameters denoted as  $\theta^\mu$  and  $\theta^Q$ , respectively. Each network is further divided into a training network and a target network. The critic network updates its parameters,  $\theta^Q$ , by minimizing a loss function, which evaluates the value of actions, while the actor network updates  $\theta^\mu$  based on the value function, outputting the expected action for the current state.

$$L(\theta^Q) = \frac{1}{N} \sum_t [y_t - Q(s_t, a_t | \theta^Q)]^2 \quad (15)$$

To reduce the Markovian nature of data sampling in the state space, the method inherits DQN's approach of fitting value functions with neural networks and updating policies with DPG, incorporating DQN's experience replay mechanism. This mechanism stores the training trajectory as  $(s_t, a_t, r_t, s_{t+1})$  and replays samples according to a specific rule to eliminate correlations between data, thus accelerating convergence. Independent target networks are set to correct biases, further reducing correlations and increasing accuracy. Here,  $s_t$ ,  $a_t$ , and  $r_t$  denote the system's state, action, and reward at time  $t$ , respectively, while  $s_{t+1}$  represents the new state acquired at time  $t + 1$  after performing action  $a_t$  in state  $s_t$ .

$$y_t = r_t + \gamma Q'(s_{t+1}, \mu'(s_{t+1} | \theta^{\mu'}) | \theta^{Q'}) \quad (16)$$

The actor network updates independently, with the current

actor and critic networks responsible for iterative updates of the policy network parameter  $\theta^Q$  and the value network parameter  $\theta^\mu$ . The deterministic policy gradient is used to maximize the reward generated by actions, and the parameter  $\theta^\mu$  is updated accordingly.

$$\nabla_{\theta^\mu} J(\theta) = \frac{1}{N} \sum_t [\nabla_a Q(s_t, \mu(s_t | \theta^\mu) | \theta^Q) \nabla_{\theta^\mu} \mu(s_t | \theta^\mu)] \quad (17)$$

The TD-error is employed to train the critic network parameters by minimizing the loss function:

$$\text{TD-error} = \text{reward}(s_t, a_t) + \gamma v'(s_{t+1}, a_{t+1}; \theta^{Q'}) - v(s_t, a_t; \theta^Q) \quad (18)$$

$$\text{Loss} = (\text{TD-error})^2 \quad (19)$$

In the training process, the neural network in the actor module is trained by maximizing the value corresponding to the state-action pair  $\langle s_t, a_t \rangle$ , using the average value of state and action evaluations as the loss function:

$$\text{Loss} = -\text{mean}[v(s_t, a_t; \theta^Q)] \quad (20)$$

In the DDPG algorithm, simply relying on the behavior selection strategy output by the algorithm may lead to insufficient exploration of the task environment. To enhance exploration, Ornstein–Uhlenbeck (OU) random noise is added to the DDPG policy. The OU process, defined as

$$dx_t = \theta(\mu - x_t)dt + \sigma dW_t \quad (21)$$

OU random noise is time-correlated, making it suitable for generating exploration noise in the optimization process of plunger lift operations over successive cycles. To prevent oscillations and divergence in network gradient calculations, and to avoid large fluctuations in network parameters, a "soft update" method is used to update the parameters of the two target networks, with  $\tau$  being the update coefficient:

$$\theta^{Q'} \leftarrow \tau \theta^Q + (1 - \tau) \theta^{Q'} \quad (22)$$

$$\theta^{\mu'} \leftarrow \tau \theta^\mu + (1 - \tau) \theta^{\mu'} \quad (23)$$

### 3.1.2. Action space

The optimization of the plunger lift working systems primarily involves controlling the after-flow and shut-in times of the well. These two variables are crucial for efficiently removing liquid accumulation from the bottom of the well and ensuring high productivity. Based on the coupled wellbore–reservoir model established for a production well T001, it is assumed that the control period spans  $n$  days. The intermittent production of plunger lift is divided into four stages, including both opening and closing actions. The adjustment of the after-flow and shut-in times is regarded as one cycle, which needs to be adjusted  $m$  times within  $n$  days. The action space is defined as

$$a_t = [t_1, t_2, t_3, \dots, t_m] \quad (24)$$

According to the control mode of the intelligent system, the action feature of the wellhead electromagnetic valve is as follows:

$$v_i = \begin{cases} 0 & \text{Increase the after-flow time } \Delta t_1 \text{ and increase the shut-in time } \Delta t_2 \\ 1 & \text{Increase the after-flow time } \Delta t_1 \text{ and decrease the shut-in time } \Delta t_2 \\ 2 & \text{Decrease the after-flow time } \Delta t_1 \text{ and increase the shut-in time } \Delta t_2 \\ 3 & \text{Decrease the after-flow time } \Delta t_1 \text{ and decrease the shut-in time } \Delta t_2 \\ 4 & \text{Increase the after-flow time } \Delta t_1 \text{ and maintain the current shut-in time} \\ 5 & \text{Decrease the after-flow time } \Delta t_1 \text{ and maintain the current shut-in time} \\ 6 & \text{Increase the shut-in time } \Delta t_2 \text{ and maintain the current after-flow time} \\ 7 & \text{Decrease the shut-in time } \Delta t_2 \text{ and maintain the current after-flow time} \\ 8 & \text{Maintain the current after-flow and shut-in time} \end{cases} \quad (25)$$

Whenever the agent gives feedback (ranging from 0 to 8) to the plunger controller, the controller adjusts the wellhead valve to increase, decrease, or maintain the current after-flow or shut-in time by  $\Delta t$ , where  $\Delta t$  should be determined based on actual operating conditions. The after-flow time should not be too short or too long, as a short after-flow time results in insufficient production, while an excessively long after-flow time can lead to severe liquid loading, reducing gas production and wasting time. Similarly, the shut-in time should be carefully managed, as a short shut-in time may not provide enough energy to lift the plunger to the wellhead, while an overly long shut-in time could cause the plunger to rise too quickly, damaging wellhead equipment and reducing daily gas production. Therefore, a rational working system ensures maximum liquid unloading in the shortest shut-in time, while also maintaining high productivity during the after-flow stage. In regulating the working systems, upper and lower limits for the after-flow and shut-in times must be set to ensure that the plunger operates normally within the wellbore.

### 3.1.3. State space

Currently, the optimization of plunger lift working systems only considers whether the plunger can be successfully lifted, with the primary goal of maximizing gas production. However, the impact of different working systems on gas production, the dynamic changes in the near-wellbore reservoir during intermittent production, and the effects of long-term intermittent production on reservoir pressure and water saturation are not fully considered. To achieve efficient and sustainable gas production, this study employs a coupled wellbore–reservoir model, integrating reservoir simulation results. The optimization aims to balance reservoir pressure changes and stabilize the gas–water ratio across wells, optimizing the entire reservoir pressure level through intelligent control of the plunger lift operation. Therefore, attributes directly related to gas production, reservoir pressure, and gas–water ratio are selected as the state space, including wellhead production rate ( $q$ ), gas–water ratio (GWR), and reservoir pressure ( $P_r$ ). The wellhead production rate and gas–water ratio are obtained from the wellbore model, while reservoir pressure is derived from the reservoir model. To make comprehensive optimization decisions, both current and future factors are considered, similar to human reasoning. The current wellhead production rate ( $q^i$ ), gas–water ratio ( $\text{GWR}^i$ ),

which is the ratio of wellhead gas production rate to water production rate, and reservoir pressure ( $P_r^i$ ) under the current working system are selected to describe the current system state, while the future wellhead production rate ( $q^{i+1}$ ), gas–water ratio ( $\text{GWR}^{i+1}$ ), and reservoir pressure ( $P_r^{i+1}$ ) under the adjusted working system are selected to describe the future system state. Hence, the system state is described by a six-dimensional state space.

$$s_t = [q^i, \text{GWR}^i, P_r^i, q^{i+1}, \text{GWR}^{i+1}, P_r^{i+1}] \quad (26)$$

### 3.1.4. Reward

Before deploying the controller based on the DDPG algorithm in an online environment, extensive offline pre-training is required. During this pre-training, the agent optimizes its strategy by maximizing the reward function value through a trial-and-error exploration mechanism. Consequently, the reward function ultimately determines the control performance of the trained controller. The plunger lift system, under certain operational conditions, provides feedback on different actions, which is reflected in the system's rewards or penalties. The ultimate goal of the system is to optimize action decisions to maximize the reward, achieving optimal control. The reward rules should be designed so that their maximum value is equivalent to the optimization of the primary objective. To achieve efficient and long-term stable production in gas wells, this study minimizes  $\Delta\text{GWR}$  and  $\Delta P_r$  while maximizing  $\Delta q$  as the optimization objectives. To ensure that the three rewards are of the same order of magnitude, the reward values are formulated as the relative changes, specifically the ratio of the incremental or decremental change to the value at the previous time step. The reward function designed accordingly.

$$r_q = \frac{\Delta q}{q^{i-1}} = \frac{q^{i+1} - q^i}{q^i} \quad (27)$$

$$r_{\text{GWR}} = \frac{\Delta\text{GWR}}{\text{GWR}^{i-1}} = \frac{\text{GWR}^{i+1} - \text{GWR}^i}{\text{GWR}^i} \quad (28)$$

$$r_{P_r} = \frac{\Delta P_r}{P_r^{i-1}} = \frac{P_r^{i+1} - P_r^i}{P_r^i} \quad (29)$$

To ensure that the adjustment of the working systems does not cause reservoir damage or wellhead equipment failure, a cumulative reward approach is adopted, considering all rewards from the current time to the final state. This ensures a smooth adjustment to the target working systems during the optimization process.

$$r = r_1 + r_2 + r_3 + \cdots + r_m \quad (30)$$

The reward function at different time points is designed as follows. During actual optimization, different weight coefficients can be set according to various production needs to meet the requirements for high and stable production. To further enhance learning speed and eliminate unreasonable training outcomes, the reward function includes not only basic items describing the main optimization objectives but also a penalty for premature termination of the simulation and a positive reward for meeting control requirements. To ensure safe and stable operation of the gas well, the opening and closing times must remain within a reasonable range. Thus, the adjustment constraints on these times serve as a condition for premature simulation termination. If the opening or closing times exceed the limits, the simulation terminates early, and a large negative value is assigned to the reward function. At this point, it is necessary to consider whether a different production process is needed to maintain normal production under the current reservoir and production conditions. On the other hand, if  $\Delta GWR$  and  $\Delta P_r$  remain within a small range, a positive reward is given.

$$r_i = \mu_1 |r_q|^2 - \mu_2 |r_{GWR}|^2 - \mu_3 |r_{Pr}|^2 + F + M \quad (31)$$

Because the environment is random, or the environmental changes are uncertain, the occurrence of the next state is also random. The action taken in the current state may not necessarily be repeated in the next state, and the same reward may not be guaranteed. The future state is uncertain, and during the accumulation of rewards, the impact of future reward changes on the reward value must be considered. Therefore, the future reward state must be adjusted using discounted future rewards  $G_t$  to replace cumulative rewards.

$$G_t = \gamma_1 r_1 + \gamma_2 r_2 + \gamma_3 r_3 + \cdots + \gamma_m r_m \quad (32)$$

### 3.2. Reinforcement learning framework for optimizing plunger lift working systems

Determining reasonable after-flow and shut-in times is one of the critical factors in controlling plunger lift operations. A well-designed working system can ensure maximum liquid removal within a short shut-in while maintaining high production after the well is opened. Different working systems result in varying wellhead gas production rates, gas–water ratios, and reservoir pressure distributions. By setting different opening and closing cycles through the wellhead electromagnetic valve, the plunger is controlled to upstroke and downstroke within the wellbore, discharging accumulated liquid, reducing bottomhole flowing pressure, and thus increasing gas production. The optimization control principle of plunger lift operations is illustrated in Fig. 3, which mainly includes environment, state, reward, action, and agent. The plunger controller executes the plunger lift working systems control algorithm while receiving remote open-close commands. It manages the movement and operation of the plunger. A digital pressure gauge is used to monitor the tubing and casing pressures in real-time, providing critical downhole and wellhead data to ensure the stable operation of the plunger lift process. The tubing and casing pressures are important parameters for judging the plunger lift operation and evaluating the effectiveness of the working systems. The state refers to the production parameters of the plunger lift system, including wellhead gas production rate, gas–water ratio, and reservoir pressure distribution. The agent can be understood as the "neural center," capable of sensing the current state, evaluating the effects of actions chosen in previous states, and interacting with the environment. The agent is responsible for making decisions, such as increasing the after-flow time, maintaining the current working systems, or decreasing the shut-in time, which corresponds to the action in the model. After receiving the agent's action signal, the plunger controller adjusts the working systems, reflected in the plunger lift operation by increasing, decreasing, or maintaining the well after-flow and shut-in times. After the action is executed, feedback from the environment—encompassing the entire production system including the wellbore and reservoir—is obtained. The environment's function is to output the system's state and the effects of the action after it is executed. For example, extending the after-flow time might increase daily gas production, while reducing the after-flow time might slow the decline in reservoir energy. These feedback values,

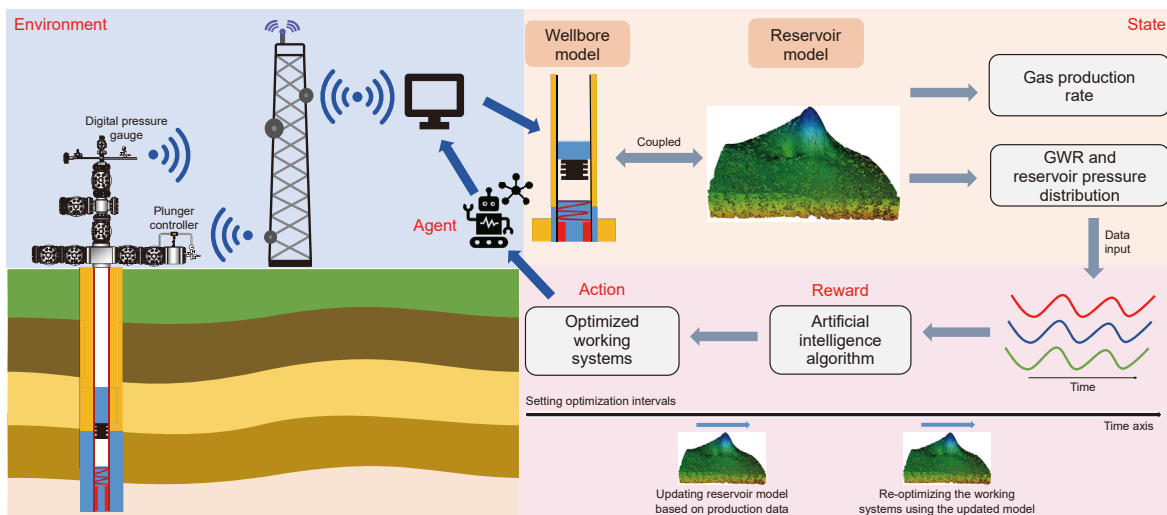


Fig. 3. Optimization control principle of plunger lift operations.

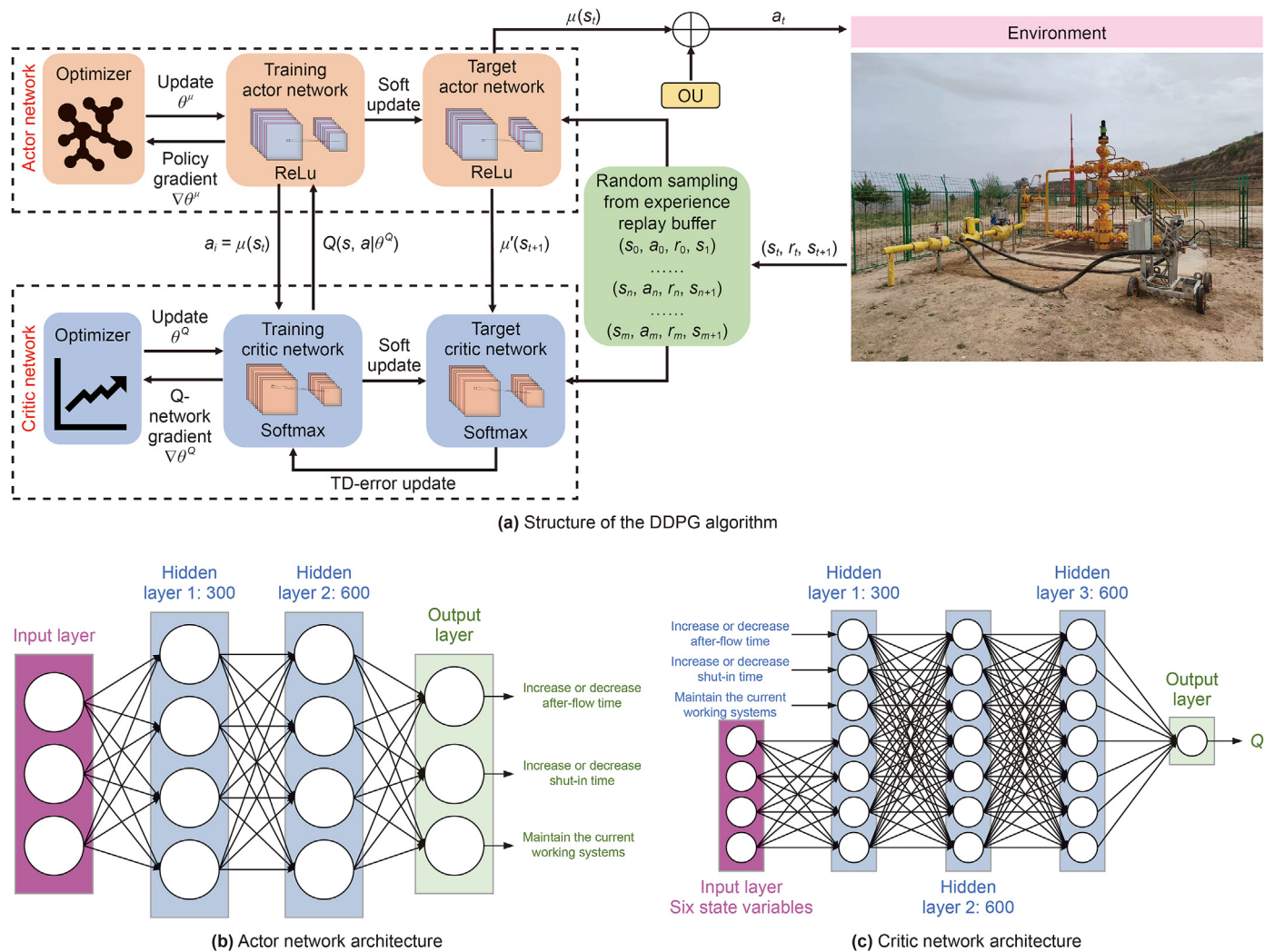


Fig. 4. Framework of the plunger lift optimization algorithm based on the DDPG algorithm.

once quantified, form the rewards in the model. These four elements—environment, state, reward, and action—constitute the basic framework of the plunger lift system.

The specific optimization process is as follows: the optimization time interval is manually set, and calculations are performed using a coupled wellbore–reservoir model. The wellbore model provides the wellhead gas production rate, and the reservoir model provides the gas–water ratio and reservoir pressure distribution. The working system optimization algorithm is established with the goal of balancing overall reservoir pressure, stabilizing the gas–water ratio for each well, and maximizing gas production. The optimized working systems is then transmitted to the plunger controller, which executes the control algorithm while the digital pressure gauge records various data during the operation. At the next optimization time point, the reservoir model is updated based on the monitoring data, and the optimization algorithm is re-executed to optimize the working systems for the subsequent time step. This process is repeated to achieve the plunger lift working system optimization.

The reinforcement learning model architecture for plunger lift working system optimization based on the coupled wellbore–reservoir model is shown in Fig. 4. This architecture is based on the DDPG algorithm and establishes an actor-critic network. The actor-critic network consists of a training network

and a target network, resulting in a total of four networks. The training network is used to update the network parameters, while the target network follows a periodic soft update strategy to assist in training the neural network.

Each set of data is normalized within the neural network, and random noise is added to the output of the actor network to achieve signal input. An exploration strategy is constructed by adding noise to the original actor policy using an OU random process with parameters  $\theta = 0.2$  and  $\sigma = 0.2$ . This process generates time-correlated values centered around zero, allowing for effective exploration in the physical environment. The actor network implements the current deterministic action policy, enabling the plunger controller to execute the current working systems. The actor network receives state variables collected from the wellbore and reservoir models as input and has two hidden layers with 300 and 600 nodes, respectively. In these hidden layers, each neuron uses a rectified linear unit (ReLU) activation function to convert input signals into output signals, which represent the probability distribution of actions. This is used to approximate the policy model  $\pi(a_t | s_t)$  and output the corresponding action. The critic network estimates the Q value of the current state-action pair using a set of parameters  $\theta^Q$ , which is crucial for network convergence. The critic network uses state variables  $s_t$  and action variables  $a_t$  as input, with three hidden layers of 600 nodes each to approximate the Q

function value. Each neuron in the hidden layers uses a softmax activation function to convert input signals into output signals, resulting in the current network  $Q$  value. The final output layer of the actor network uses a  $\tanh(x)$  activation function, mapping real numbers to the  $[0, 1]$  range, after which the actor network output is scaled to the range of well opening and closing adjustment times, mapping network outputs to the control actions of the plunger

parameters of the target actor and critic networks, thereby enhancing the stability and convergence of the training process.

The algorithm designed in this study based on the above framework is shown in Algorithm 1.

**Algorithm 1.** Plunger lift optimization algorithm based on the DDPG algorithm

---

**Algorithm 1** Plunger lift optimization algorithm based on the DDPG algorithm

---

```

1  Initialize the DDPG model;
2  Initialize the weights of the actor and critic networks, and copy the parameters of the actor and critic networks to the
   corresponding target networks  $Q'$  and  $\mu'$ :  $\theta^{Q'} \leftarrow \theta^Q, \theta^{\mu'} \leftarrow \theta^\mu$ ;
3  Initialize the experience replay buffer;
4  for episode = 1, 2, ...,  $N$ :
5      Initialize the environment model and obtain the initial plunger lift well working systems;
6      for  $t = 1, 2, \dots, T$ :
7          Obtain action  $a_t = \mu(s_t|\theta^\mu) + \phi$  based on the OU random noise and the current policy;
8          The plunger controller executes action  $a_t$ , obtaining reward feedback and the next state;
9          if the after-flow and shut-in times exceed the constraints:
10             Reinitialize the environment model and re-optimize;
11         end if
12         if the experience replay buffer is not full:
13             Store the interaction data  $(s_t, a_t, r_t, s_{t+1})$  between the agent and the environment in the experience replay
13             buffer;
14         else:
15             Randomly replace a set of data in the experience replay buffer with the obtained interaction data;
16             Randomly sample a set of interaction data from the replay buffer as training data for the actor and critic
               networks;
17         end if
18         The actor network provides  $a_{t+1}$  to the critic network based on  $s_{t+1}$ ;
19         The critic network calculates  $Q(s_t, a_t)$  and  $Q(s_{t+1}, a_{t+1})$ , and computes TD-error;
20         The critic network updates network parameters  $\theta^Q$  based on TD-error;
21         The actor network updates network parameters  $\theta^\mu$  based on the  $Q(s_t, a_t)$  provided by the critic network;
22         Soft update the target network parameters  $\theta^{Q'}, \theta^{\mu'}$ :
               
$$\theta^{Q'} \leftarrow \tau\theta^Q + (1 - \tau)\theta^{Q'}$$

               
$$\theta^{\mu'} \leftarrow \tau\theta^\mu + (1 - \tau)\theta^{\mu'}$$

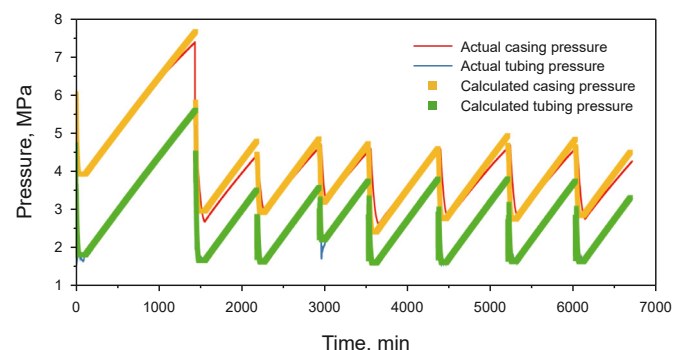
23     end for
24 end for

```

---

controller. The critic's final output layer directly outputs  $Q$  without using the ReLU activation function.

During training, the actor network is responsible for selecting actions under different environmental states, while the critic network evaluates the value of these actions. They learn from each other in a mutually supervised relationship. The training neural network is used to evaluate the value of the current state and behavior, while the target neural network receives the state at the next time step and the corresponding behavior output by the target actor network, and makes a value judgment. The DQN experience replay mechanism is introduced to store interaction data between the agent and the environment  $(s_t, a_t, r_t, s_{t+1})$ . The DDPG algorithm enhances training stability and effectiveness by randomly sampling batches from the experience replay buffer. A soft update mechanism for the target network is applied to gradually update the



**Fig. 5.** Comparison of calculated wellhead tubing and casing pressures with field data.

**Table 1**  
Hyperparameter settings.

| Parameter                    | Value  |
|------------------------------|--------|
| Episode                      | 500    |
| Actor network learning rate  | 0.0003 |
| Critic network learning rate | 0.001  |
| Discount factor              | 0.99   |
| Update coefficient           | 0.001  |
| Mini batch                   | 64     |
| Max step                     | 120    |

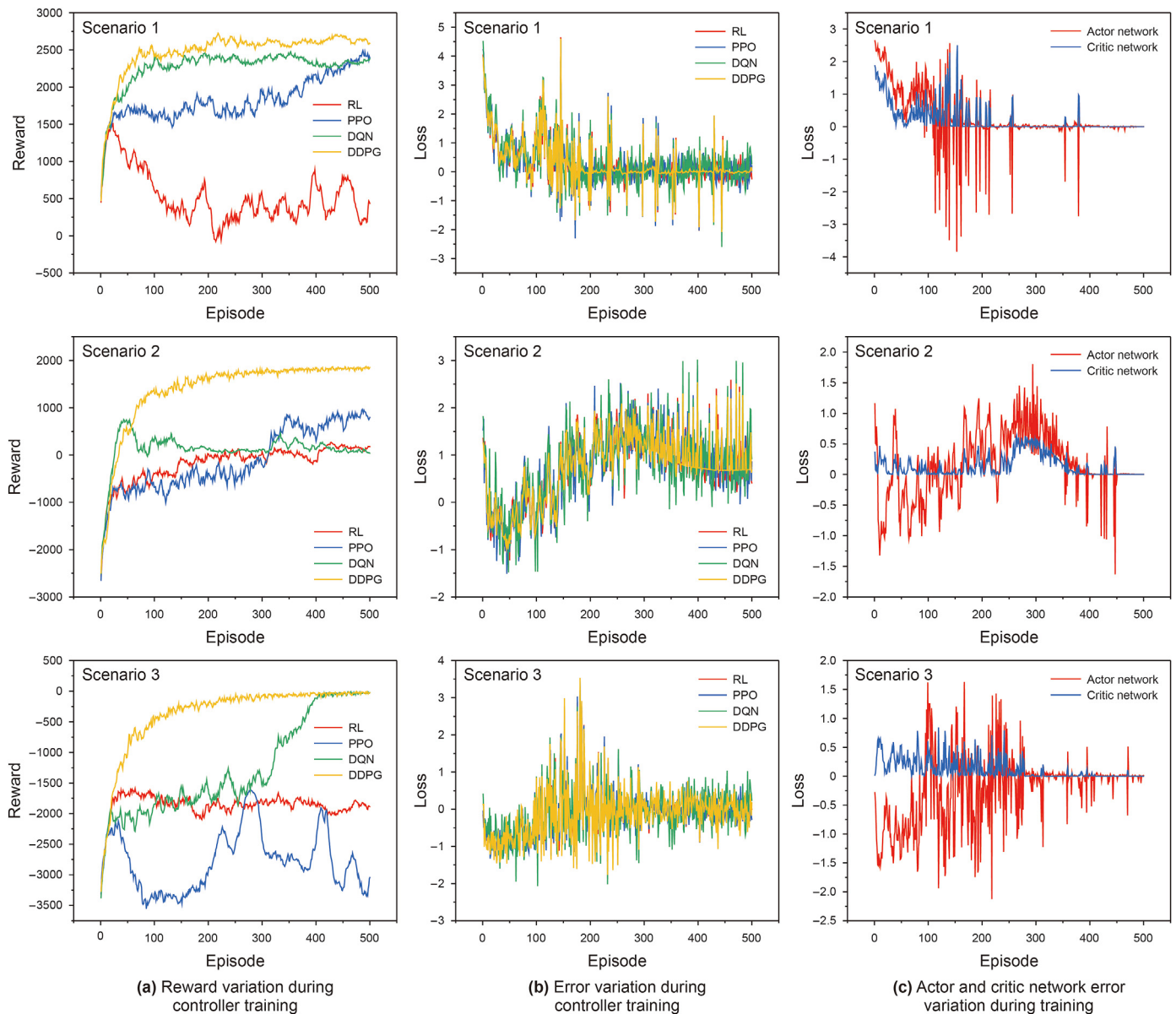
**Table 2**  
Weight coefficients for different scenarios.

| Scenario   | $\mu_1$ | $\mu_2$ | $\mu_3$ |
|------------|---------|---------|---------|
| Scenario 1 | 0.8     | 0.1     | 0.1     |
| Scenario 2 | 0.5     | 0.4     | 0.1     |
| Scenario 3 | 0.5     | 0.1     | 0.4     |

4. Simulation experiment

4.1. Experimental setup

To validate the operability of the optimization algorithm for plunger lift working systems, this study established a reservoir and wellbore model based on field data to conduct simulation experiments and optimization analyses. The reservoir model was constructed using the commercial numerical simulation software CMG (Fig. 1), with the model establishment process detailed in Section 2.2, so it will not be elaborated upon here. The production well T001 in the development block was selected for this study. This well is a vertical well with a depth of 3020 m, a bottomhole seating set at 3009 m, and uses a conventional rod-type plunger with an outer diameter of 48 mm. The tubing and casing have outer diameters of 64 and 177.8 mm, respectively. The wellhead receives remote control signals from a plunger controller to remotely



**Fig. 6.** Training results.

control the solenoid valve and, thus, the plunger's up-and-down movement. The current working systems involve a cycle of 3 h of shut-in followed by 3 h of production. Real-time data on wellhead tubing pressure, casing pressure, and gas production rate are measured, with data acquisition intervals set to 1 min. The calculated wellhead tubing and casing pressures from the established wellbore–reservoir coupling model were compared with field measurements (as shown in Fig. 5), indicating a good fit. This suggests that the model is highly accurate and can precisely describe the transient production conditions in the gas reservoir and wellbore, providing a solid foundation for subsequent intelligent optimization and control.

The hyperparameters used in the DDPG algorithm for the experiment are shown in Table 1, while the hyperparameter settings for other optimization algorithms are provided in Appendix A.

To validate the performance and robustness of the proposed control strategy for optimizing the plunger lift working systems, different weight coefficients were set in the reward function according to various production requirements. Specifically,  $\mu_1$  represents the weight of gas production rate,  $\mu_2$  represents the weight of formation pressure distribution, and  $\mu_3$  represents the weight of gas–water ratio. Scenario 1 aims for maximum gas production while considering stable production and water control; Scenario 2 aims for long-term stable production, balancing gas production and water control; and Scenario 3 focuses primarily on water control. The specific weight parameter settings for each scenario are detailed in Table 2.

#### 4.2. Experimental results and discussion

The variations in training reward values for the plunger lift optimization control model based on the DDPG algorithm, under different scenarios, are compared with the reward value variations from other reinforcement learning models. These comparisons are depicted in Fig. 6(a). The horizontal axis represents the total number of training episodes, with a total of 500 episodes in the experiment. The vertical axis shows the cumulative reward value for each episode. From the figures, it is evident that in all three scenarios, the RL and PPO algorithms exhibit relatively low reward values and poor stability. Although the DQN algorithm is more stable, it requires a higher number of episodes to reach optimal performance, resulting in a slower convergence speed. In contrast, the DDPG algorithm demonstrates stable performance once it

converges to the optimal solution, outperforming the DQN algorithm. Around the 100th episode, the DDPG algorithm's agent has already learned a series of control actions that maximize the target value. This finding confirms that the DDPG-based reinforcement learning model can converge effectively, even in a relatively small state space. The oscillations in the reward value curves are attributed to the sustained exploration/exploitation probability, which prevents the model from becoming trapped in a local optimum. The DDPG algorithm tends to overestimate  $Q$  values, leading to significant reward fluctuations during convergence. However, in solving complex high-dimensional problems, this instability in DDPG convergence often helps avoid premature convergence to suboptimal reward values. The comparison with other reinforcement learning models shows that traditional RL, PPO, and DQN algorithms start with lower initial rewards, have lower exploration efficiency, and slower or even nonexistent convergence rates. In contrast, the DDPG algorithm in this study selects more reasonable actions early in training, leading to higher initial rewards, greater exploration efficiency, faster reward convergence, and significantly higher rewards than other models. To verify the effectiveness of the optimization control, the actual performance was evaluated by observing the errors during training. The errors for different reinforcement learning models are shown in Fig. 6(b). During optimization, the RL model exhibits the largest errors and the poorest control stability. Similarly, the PPO and DQN models show large errors and perform poorly during the optimization process. However, the DDPG algorithm's error oscillations gradually decrease, significantly enhancing the stability of the optimization control. The error variations during the DDPG algorithm's controller training process in different scenarios are depicted in Fig. 6(c). It can be seen that the actor network's error oscillates between  $-4$  and  $3$  across different scenarios, gradually converging to near-zero as training progressed. The critic network's error exhibits an overall downward trend, with significant initial oscillations, indicating high instability in the controller's strategy at the beginning of training. As training continues, the control strategy stabilizes, with error oscillations decreasing and ultimately converging to near-zero.

In the optimization simulation process, upper and lower limits were set for the after-flow and shut-in times. To ensure production efficiency, the lower limit for the after-flow time was set at 1 h, and for the shut-in time, it was set at 6 h. The upper limit for the after-flow time was determined by considering the liquid loading issue in

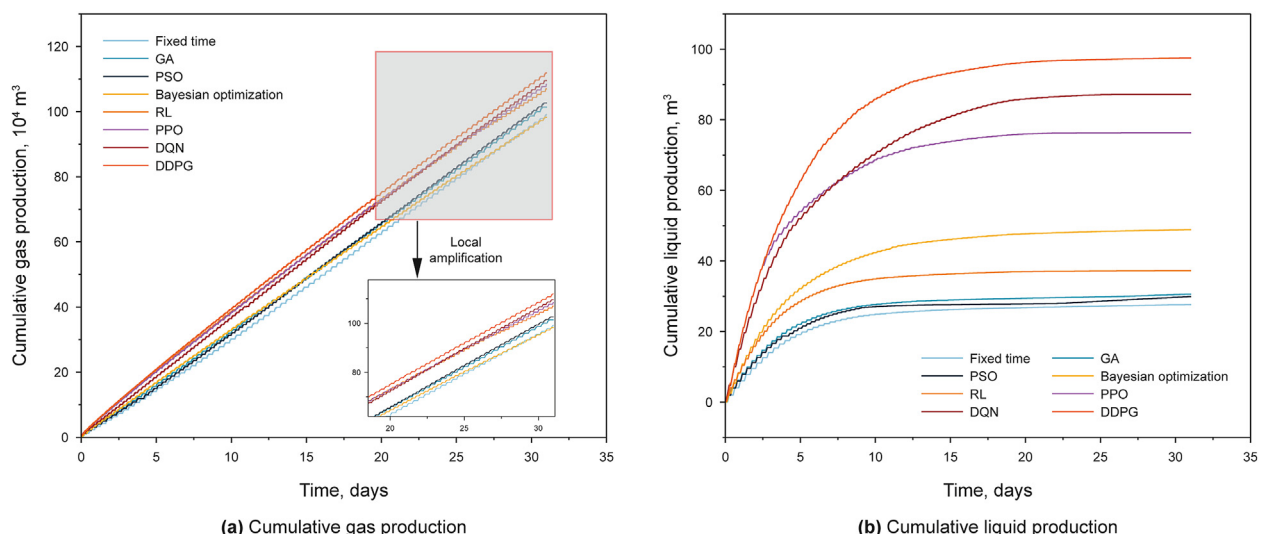


Fig. 7. Coupled model calculation results.

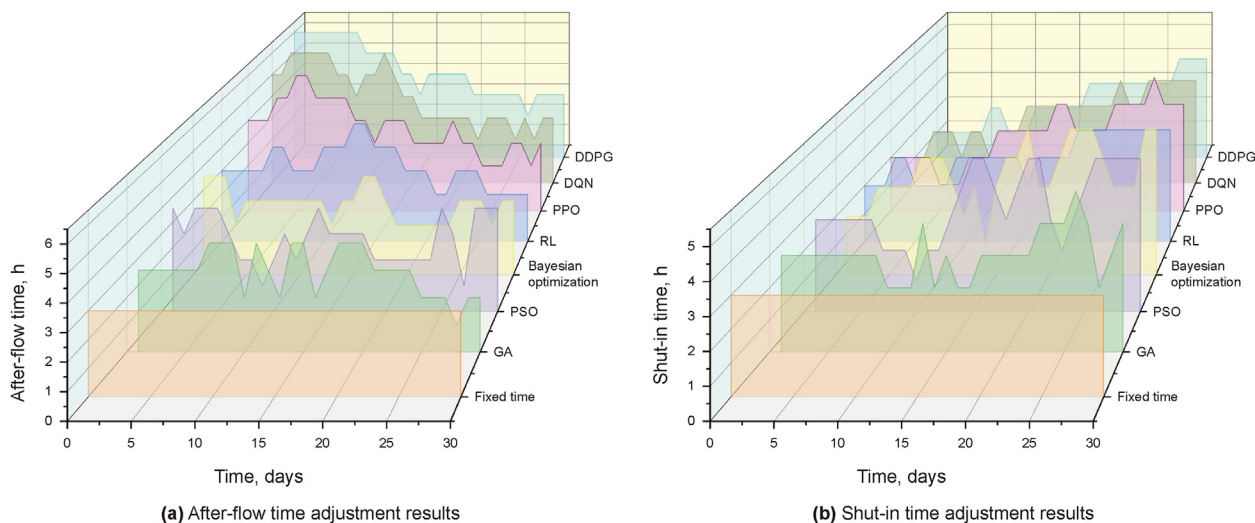


Fig. 8. After-flow and shut-in time adjustments for different optimization models.

the gas well. If the wellbore is in a liquid loading state, continued production would be uneconomical. Therefore, by calculating the critical liquid loading gas flowrate for the block, the after-flow time was capped at 6 h. The lower limit for the shut-in time was set at 1 h to ensure that sufficient energy could be accumulated in the wellbore during the shut-in period, allowing the plunger to be lifted to the wellhead. To prevent plunger lift failure, the lower limit was set to 1 h. These upper and lower limits for after-flow and shut-in times can be adjusted at different production stages.

Taking Scenario 1 as an example, the fixed after-flow and shut-in time mode used the existing working system with 3 h opening and 3 h closing periods. The cumulative gas and liquid production simulated results using different optimization algorithms compared with the fixed time system are shown in Fig. 7. From Fig. 7, it can be observed that the cumulative gas production using the DDPG algorithm increased by approximately 5%, and cumulative liquid production increased by 252% compared to the fixed time system. Compared to traditional optimization algorithms, such as genetic algorithm (GA), particle swarm optimization (PSO), and Bayesian optimization, reinforcement learning algorithms demonstrate superior performance in terms of cumulative gas and liquid production. This advantage arises because traditional stochastic gradient methods typically optimize objectives at the current moment (single-step optimal), making them less effective for problems involving long-term rewards. In the plunger lift production process studied, short-term production increases may lead to long-term damage, such as water breakthrough in the wellbore.

DDPG, leveraging the cumulative reward mechanism of reinforcement learning (discounted cumulative returns), can balance short-term gains with long-term objectives. By optimizing long-term rewards, it identifies globally optimal production strategies. Furthermore, DDPG efficiently handles high-dimensional state parameters, maintaining performance and robustness under multivariate conditions. With the representational power of deep neural networks, it effectively processes high-dimensional states and approximates complex nonlinear relationships.

The adjustment results for after-flow and shut-in times under different optimization models are shown in Fig. 8. The RL optimization algorithm, as shown in Fig. 6, indicates that the agent is still in the learning phase, with many action-state values yet to be explored or updated, leading to increasing reward values. Therefore, the RL model performs worse than other models. During the early production stage, when reservoir energy is abundant, the after-flow time should be longer and the shut-in time shorter. As production continues, with decreasing reservoir pressure and gas–water ratio, the optimization models begin adjusting the after-flow and shut-in times. The three models show similar adjustment trends, with a continuous decrease in after-flow time and an increase in shut-in time. However, the PPO and DQN algorithms have large optimization errors, with frequent fluctuations in after-flow and shut-in times during the optimization process, which is unfavorable for production. The DDPG algorithm's optimization process is smoother and better suited to this scenario, providing the most optimal optimization path. In summary, compared to other

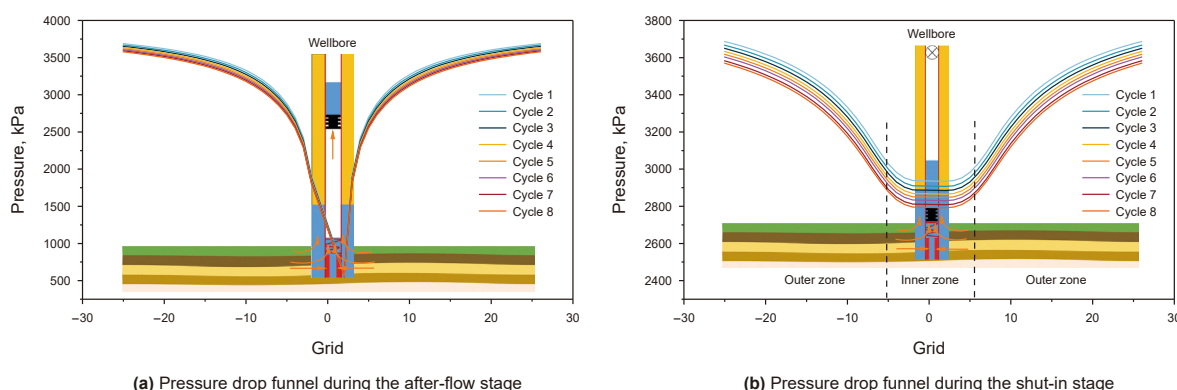


Fig. 9. Pressure drop funnels during different cycles simulated by the coupled model.

algorithms, the DDPG algorithm demonstrates robustness across different optimization scenarios, with good adaptability and generalization capabilities, enabling the system to achieve efficient production and long-term stable output under various optimization objectives.

In the plunger lift gas production process, the liquid accumulated at the bottom of the well is continuously discharged, resulting in a continuous reduction in bottomhole flowing pressure. As production continues, the pressure decline varies from the near-wellbore region to the far-well end, forming a pressure drop funnel centered around the wellbore. For the intermittent production process of plunger lift, during the after-flow stage, fluids are continuously produced from the reservoir and flow out of the wellhead. During the shut-in stage, since a production pressure differential still exists, fluids continue to flow into the wellbore, storing energy. Therefore, whether during the after-flow or shut-in stages, as shown in Fig. 9, the reservoir continuously produces fluids, and the reservoir pressure distribution, i.e., the pressure drop funnel, is constantly changing. Fig. 9(a) illustrates the pressure drop funnel during the plunger lift upstroke and after-flow production stages simulated using the coupled model. At the beginning of after-flow, the elastic release of compressed gas in the gas wellbore causes a rapid drop in wellhead pressure, resulting in a high instantaneous flow rate at the wellhead for a short period, followed by a slower and more stable pressure decline. During after-flow production, as the wellbore pressure decreases, the bottomhole flowing pressure drops, increasing the production pressure differential and reservoir productivity. As the pressure release in the near-well region concludes, reservoir productivity gradually declines. Fig. 9(b) illustrates the pressure drop funnel during the shut-in stage simulated using the coupled model. Due to the short shut-in time in intermittent plunger lift production and the slow reservoir seepage rate, the pressure propagation speed is slow, leading to a pressure drop funnel with distinct inner and outer zones within the control range of the gas well. The pressure distribution in the outer zone is similar to the pressure drop funnel during the after-flow stage, unaffected by the intermittent production system. The pressure distribution in the inner zone exhibits a cyclical variation between a positive and negative funnel, with its shape and range influenced by the periodic liquid accumulation and discharge system. As the production time increases, the inner zone gradually expands.

5. Conclusions

The optimization of the plunger lift working systems is influenced by numerous factors, making it a multifactor optimization challenge. Traditional optimization methods often yield only locally optimal solutions and cannot guarantee global optimality, nor can they quickly and comprehensively address parameter optimization issues. In this study, a coupled wellbore–reservoir model for plunger lift wells was developed, enabling transient production simulation of the wellbore and reservoir. An optimization control algorithm for plunger lift operation based on the DDPG framework was proposed, targeting overall pressure equilibrium in the gas reservoir, stable gas–water ratio across wells, and maximizing gas production. The following conclusions were drawn.

- (1) A coupled wellbore–reservoir model for plunger lift wells was established using self-developed software and the commercial numerical simulation software CMG. The model's results were compared with field-measured data, showing high accuracy. The established coupled model can accurately describe transient production within the gas

- reservoir and wellbore, providing a foundation for subsequent intelligent control.
- (2) Based on the coupled model, an optimization control algorithm for plunger lift operation was developed using the DDPG framework, aiming for overall pressure equilibrium in the gas reservoir, stable gas–water ratio across wells, and maximizing gas production.
- (3) Comparative results of different optimization algorithms indicate that the DDPG algorithm achieves the highest reward values and smallest errors during training, with a maximum increase in cumulative gas production of 5% and a 252% increase in cumulative liquid production. The results demonstrate that the DDPG algorithm exhibits robustness, adaptability, and generalization capabilities across different optimization scenarios.
- (4) The pressure drop funnel within the control range of an intermittent gas well exhibits distinct inner and outer zone characteristics compared to conventional gas wells. The coupled model developed in this study simulated the production of an intermittent gas well, revealing that the pressure distribution in the outer zone remains consistent with the after-flow stage pressure drop funnel, unaffected by the intermittent production regime. The pressure distribution in the inner zone shows a cyclical variation between a positive and negative funnel, with its shape and range influenced by the periodic liquid accumulation and discharge regime. The inner zone expands as the production time increases.

CRediT authorship contribution statement

**Zhi-Sheng Xing:** Writing – original draft, Methodology, Investigation, Conceptualization. **Guo-Qing Han:** Supervision, Methodology, Conceptualization. **You-Liang Jia:** Investigation, Formal analysis. **Wei Tian:** Validation, Investigation, Data curation. **Hang-Fei Gong:** Methodology, Investigation. **Wen-Bo Jiang:** Validation. **Pei-Dong Mai:** Formal analysis. **Xing-Yuan Liang:** Writing – review & editing.

Declaration of competing interest

All authors disclosed no relevant relationships.

Acknowledgement

This work was financial support from Science Foundation of China University of Petroleum, Beijing (No. 2462023YJRC019), National Natural Science Foundation of China (No. 52204059) and Key Core Technology Research Project Foundation of PetroChina Group (No. 2023ZG18).

Appendix. A Hyperparameter settings for different optimization algorithms

RL algorithms:

Table A.1  
Hyperparameter settings for RL algorithms

| Parameter       | Value |
|-----------------|-------|
| Episode         | 500   |
| Discount factor | 0.99  |
| Max step        | 120   |

PPO algorithms:

**Table A.2**  
Hyperparameter settings for PPO algorithms

| Parameter                    | Value  |
|------------------------------|--------|
| Episode                      | 500    |
| Actor network learning rate  | 0.0003 |
| Critic network learning rate | 0.001  |
| Discount factor              | 0.99   |
| Mini batch                   | 64     |
| Max step                     | 120    |

DQN algorithms:

**Table A.3**  
Hyperparameter settings for DQN algorithms

| Parameter                       | Value |
|---------------------------------|-------|
| Episode                         | 500   |
| Learning rate                   | 0.001 |
| Discount factor                 | 0.99  |
| Target update frequency         | 2     |
| Experience replay buffer length | 100   |
| Mini batch                      | 64    |
| Max step                        | 120   |

DDPG algorithms:

**Table A.4**  
Hyperparameter settings for DDPG algorithms

| Parameter                       | Value  |
|---------------------------------|--------|
| Episode                         | 500    |
| Actor network learning rate     | 0.0003 |
| Critic network learning rate    | 0.001  |
| Discount factor                 | 0.99   |
| Update coefficient              | 0.001  |
| Experience replay buffer length | 100    |
| Mini batch                      | 64     |
| Max step                        | 120    |

References

Aires, J.D.M., Gadelha, J.R.T., de A Dantas, A.F.O., Bolonhini, É.H., Dórea, C.E.T., Maitelli, A.L., 2020. Plunger lift optimization: a comparison of fuzzy and predictive controllers with manipulation of a secondary gas valve. *J. Petrol. Sci. Eng.* 192, 107261. <https://doi.org/10.1016/j.petrol.2020.107261>.

Akhiartdinov, A., Pereyra, E., Sarica, C., Severino, J., 2020. Data analytics application for conventional plunger lift modeling and optimization. In: SPE Artificial Lift Conference and Exhibition. <https://doi.org/10.2118/201144-MS>.

Barros, J.L., Claramunt, J.L., Ferrigno, E., 2018. Novel approach in digital diagnostic for plunger lift in unconventional wells. In: SPE Artificial Lift Conference and Exhibition. <https://doi.org/10.2118/190930-MS>.

Chen, P., Chen, Y., Yang, C., Xu, Y., Feng, G., 2024. Gas well production optimization: classifying liquid loading severity in shale gas wells using semi-supervised learning. *Gas Science and Engineering* 128, 205394. <https://doi.org/10.1016/j.jgsce.2024.205394>.

Foss, D.L., Gaul, R.B., 1965. Plunger-life performance criteria with operating experience-Ventura Avenue Field. *Drill. Drilling and Production Practices*. API-65-124.

Gupta, A., Kaisare, N.S., Nandola, N.N., 2017. Dynamic plunger lift model for deliquification of shale gas wells. *Comput. Chem. Eng.* 103, 81–90. <https://doi.org/10.1016/j.compchemeng.2017.03.005>.

Hashmi, G.M., Hasan, A.R., Kabir, C.S., 2018. Simplified modeling of plunger-lift

assisted production in gas wells. *J. Nat. Gas Sci. Eng.* 52, 454–460. <https://doi.org/10.1016/j.jngse.2018.02.009>.

Hu, B., Sagen, J., Chupin, G., Haugset, T., Ek, A., Sommersel, T., Xu, Z.G., Mantecon, J.C., 2007. Integrated wellbore/reservoir dynamic simulation. In: Asia Pacific Oil and Gas Conference and Exhibition. <https://doi.org/10.2118/109162-MS>.

Kamari, A., Bahadori, A., Mohammadi, A.H., 2016. Prediction of maximum possible liquid rates produced from plunger lift by use of a rigorous modeling approach. *SPE Prod. Oper.* 32, 7–11. <https://doi.org/10.2118/180912-PA>.

Nandola, N.N., Kaisare, N.S., Gupta, A., 2018. Online optimization for a plunger lift process in shale gas wells. *Comput. Chem. Eng.* 108, 89–97. <https://doi.org/10.1016/j.compchemeng.2017.09.001>.

Ozkan, E., Keefer, B., Miller, M.G., 2003. Optimization of plunger-lift performance in liquid loading gas wells. In: SPE Annual Technical Conference and Exhibition. <https://doi.org/10.2118/84135-MS>.

Parsa, E., Ozkan, E., Lowery, B., Cathcart, D., Lohmann, M., 2013. Enhanced plunger lift performance utilizing reservoir modeling. In: SPE Production and Operations Symposium. <https://doi.org/10.2118/164473-MS>.

Peng, L., Pagou, A.L., Tian, L., Chai, X., Han, G., Yin, D., Zhang, K., 2024. A fully coupled compositional wellbore/reservoir model for predicting liquid loading in vertical and inclined gas wells. *Geoenery Science and Engineering* 239, 212874. <https://doi.org/10.1016/j.geoen.2024.212874>.

Romero, A., Feldmann, C., Alonso, K.S., Martinez, G., Barros, J., Montero, M., Martinez, G., Claramunt, J.L.A., Martinez, G., Ferrigno, E., 2020. Prescriptive model for automatic online plunger lift unconventional wells optimization. In: Latin America Unconventional Resources Technology Conference, Unconventional Resources Technology Conference. <https://doi.org/10.15530/urtec-2020-1427>.

Sayman, O., Jones, K., Hale, R., Pereyra, E., Sarica, C., 2022a. A field case study of plunger lift related tubing deformation. *J. Nat. Gas Sci. Eng.* 97, 104342. <https://doi.org/10.1016/j.jngse.2021.104342>.

Sayman, O., Tang, Y., Chow, J., Farrell, K., Pereyra, E., Sarica, C., 2022b. Experimental and field investigations of bypass plunger lift. In: SPE Annual Technical Conference and Exhibition. <https://doi.org/10.2118/210463-MS>.

Singh, A., 2017. Application of data mining for quick root-cause identification and automated production diagnostic of gas wells with plunger lift. *SPE Prod. Oper.* 32, 279–293. <https://doi.org/10.2118/175564-PA>.

Tan, B., Liu, X., Liu, Y., Chang, Y., Tian, W., Jia, Y., Han, G., Liang, X., 2023. Mechanism of liquid unloading by single flowing plunger lift in gas wells. *Int. J. Hydrogen Energy* 48 (7), 2571–2582. <https://doi.org/10.1016/j.ijhydene.2022.10.118>.

Tan, X., Deng, Y., Luo, A., Li, Xiaoping, Tian, W., Li, Xuri, Liu, Y., Zhou, C., 2023. Experimental and numerical simulation research on sealing performance and drainage efficiency of different types of plungers. *Geoenery Science and Engineering* 224, 211517. <https://doi.org/10.1016/j.geoen.2023.211517>.

Tong, Z., Zhao, G., Wei, S., 2017. A novel intermittent gas lifting and monitoring system toward liquid unloading for deviated wells in mature gas field. *J. Energy Resour. Technol.* 140 (5), 052906. <https://doi.org/10.1115/1.4038623>.

Wu, Q., Han, G., Yang, D., Zhu, Z., Peng, L., Ma, H., Liang, X., 2024. Reliable and efficient algorithms to optimize gas well intermittent production through automatically intelligent control. In: SPE Gas & Oil Technology Showcase and Conference. <https://doi.org/10.2118/219130-MS>.

Xiang, Z., Kabir, C.S., 2019. Simplified transient-IPR modeling in intermittent gas-lift and plunger-lift systems. *J. Petrol. Sci. Eng.* 179, 31–43. <https://doi.org/10.1016/j.petrol.2019.04.040>.

Xie, Y., Ma, S., Wang, H., Li, N., Zhu, J., Wang, J., 2023. Unsupervised clustering for the anomaly diagnosis of plunger lift operations. *Geoenery Science and Engineering* 231, 212305. <https://doi.org/10.1016/j.geoen.2023.212305>.

Zhao, K., Bai, B., 2018. Transient liquid leakage during plunger lifting process in gas wells. *J. Nat. Gas Sci. Eng.* 59, 250–261. <https://doi.org/10.1016/j.jngse.2018.09.009>.

Zhao, K., Tian, W., Li, X., Bai, B., 2018. A physical model for liquid leakage flow rate during plunger lifting process in gas wells. *J. Nat. Gas Sci. Eng.* 49, 32–40. <https://doi.org/10.1016/j.jngse.2017.10.008>.

Zhao, K., Mu, L., Tian, W., Bai, B., 2021. Gas-liquid flow seal in the smooth annulus during plunger lifting process in gas wells. *J. Nat. Gas Sci. Eng.* 95, 104195. <https://doi.org/10.1016/j.jngse.2021.104195>.

Zhao, Q., Zhu, J., Cao, G., Zhu, H., Zhang, H.-Q., 2021. Transient modeling of plunger lift for gas well deliquification. *SPE J.* 26, 2928–2947. <https://doi.org/10.2118/205386-PA>.

Zhong, Z., Wang, H., Li, N., Zhu, H., Zhu, J., Wang, J., 2024. Prediction of plunger lift dynamics using a bidirectional long short-term memory neural network with an innovative forecasting strategy. In: 2024 IEEE 3rd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA). <https://doi.org/10.1109/EEBDA60612.2024.10485764>.

Zhu, J., Cao, G., Tian, W., Zhao, Q., Zhu, H., Song, J., Peng, J., Lin, Z., Zhang, H.-Q., 2019. Improved data mining for production diagnosis of gas wells with plunger lift through dynamic simulations. In: SPE Annual Technical Conference and Exhibition. <https://doi.org/10.2118/196201-MS>.

Zhu, Z., Han, G., Liang, X., Chang, S., Yang, B., Yang, D., 2024. Rapid classification and diagnosis of gas wells driven by production data. *Processes* 12, 1254. <https://doi.org/10.3390/pr12061254>.