ORIGINAL RESEARCH **Open access**

CNN-LSTM Surrogate-Driven Deep Reinforcement Learning for Intelligent CBM Drainage Optimization

Yifan Zhao[1*], Liang Chen[1]

Abstract

Coalbed methane (CBM) production forms a substantial share of unconventional natural gas development and is essential for boosting both output and financial returns by refining its full life-cycle drainage approach. Conventional drainage methods largely depend on expert judgment and repeated experimentation, which limits their ability to respond to intricate and fluctuating production conditions. This study presents a comprehensive life-cycle optimization framework for CBM production and drainage using Deep Reinforcement Learning (DRL). Employing the Deep Q-Network (DQN) algorithm, the research enables the system to learn and refine drainage decisions throughout the production lifecycle, supporting fully intelligent operational control. An auto-regressive surrogate model built upon a deep learning architecture (CNN-LSTM) is incorporated to establish the reinforcement learning environment for gas production and drainage simulation. This surrogate replaces the resource-intensive full-physics model, substantially accelerating the agent-environment interaction required for DRL. Two reinforcement learning strategies were developed, one using gas production as the reward and the other using Net Present Value (NPV). Results indicate that the DRL strategy guided by NPV as the reward improved net gain by 5.83% over historical performance. Relative to conventional techniques, this method markedly enhances NPV and delivers a superior drainage policy. Overall, the DRL-based life-cycle drainage optimization for CBM proves to be both practical and highly effective, offering a smart technological pathway for unconventional natural gas resource development. The outcomes emphasize the framework's excellent adaptability and its capability to tackle intricate optimization challenges within dynamic production settings.

Keywords Deep reinforcement learning, Coalbed methane, Drainage optimization strategy, Auto-regressive surrogate

*Correspondence:
Yifan Zhao
yifanzhao@tsinghua.edu.cn
[1] Department of Artificial Intelligence and Energy Informatics, School of Electrical Engineering, Tsinghua University, Beijing, China

Introduction

Natural gas serves as a vital transitional energy resource, acting as a lower-carbon substitute to coal for electricity generation and industrial processes. Nevertheless, its future relevance will hinge on the pace of global energy transition and the degree to which renewables and other low-carbon solutions can replace it. Worldwide, the bp Energy Outlook 2024 underscores that natural gas demand is closely tied to the speed of this transition. Under the Current Trajectory scenario, demand is projected to rise by approximately one-fifth by 2050, increasing its portion of primary energy to just over 25%. This expansion is primarily fueled by growing consumption in developing economies undergoing rapid industrialization.

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

In contrast, the Net Zero scenario sees demand peaking near the end of this decade and declining to roughly half of 2022 levels by 2050, as societies shift toward widespread electrification and cleaner fuel options [1]. As of 2024, natural gas maintains a prominent position in the global energy landscape. Per the Australian Energy Update 2024, it represented 26% of Australia's total energy consumption, while production reached 6,200 × 10^{15} joules — approximately 2.5 times the level from a decade earlier. Even with a 3% drop in consumption during 2022–23, natural gas remains a key energy source, particularly in the mining industry for power generation and chemical feedstock applications [2]. As an unconventional natural gas resource, CBM exhibits a specialized production mechanism, governed by its intricate geological properties and fluid flow behavior [3, 4].

The production life cycle of coalbed methane involves three primary phases [4-6]. The initial phase involves water drainage and reservoir pressure reduction. Here, gas wells extract water to lower reservoir pressure, during which only single-phase water flow exists within the coal seam. This phase confirms the absence of free gas in the coal cleats [5]. The subsequent phase is gas desorption. Once reservoir pressure falls below a critical threshold, coalbed methane begins desorbing from the coal matrix. Combined with ongoing water removal and gas output, this critical desorption behavior confirms that the coalbed methane exists in an under-saturated condition. The final phase is the convergence seepage stage. Desorbed methane migrates from the matrix into the cleat system. Water output from the well steadily declines while gas production steadily rises, potentially transitioning into a single-phase gas flow period [6-9].

Dynamic analysis and numerical simulation of coalbed methane production depend on precise records of gas and water production [10-15]. In the coal seam water drainage process, water must be lifted from the wellbore to the surface via pumps, as illustrated in **Figure 1**. The drainage volume measured at the wellhead represents surface-measured water. In contrast, coal seam water production refers to the actual inflow from the reservoir into the wellbore [8]. Both the wellhead drainage volume and reservoir water inflow directly influence bottom-hole flowing pressure, which governs the extent of gas desorption and overall CBM recovery efficiency. Because of wellbore storage effects, changes in bottom-hole flowing pressure cause the surface drainage volume to differ from the actual reservoir water production [9, 10]. In medium- to high-rank CBM reservoirs, surface drainage rates are typically low, while bottom-hole pressure fluctuates significantly; thus, wellbore storage significantly influences production dynamics [11]. Previous studies reveal that the greatest discrepancy between daily water production and daily drainage volume occurs during the early production period, reaching a maximum deviation of 72% [10]. However, during the middle and later production stages, slower and smaller variations in bottom-hole pressure mean the difference between well drainage volume and reservoir water inflow becomes negligible [13]. Consequently, the scientific optimization of coalbed methane drainage strategy remains a critical challenge requiring immediate attention.

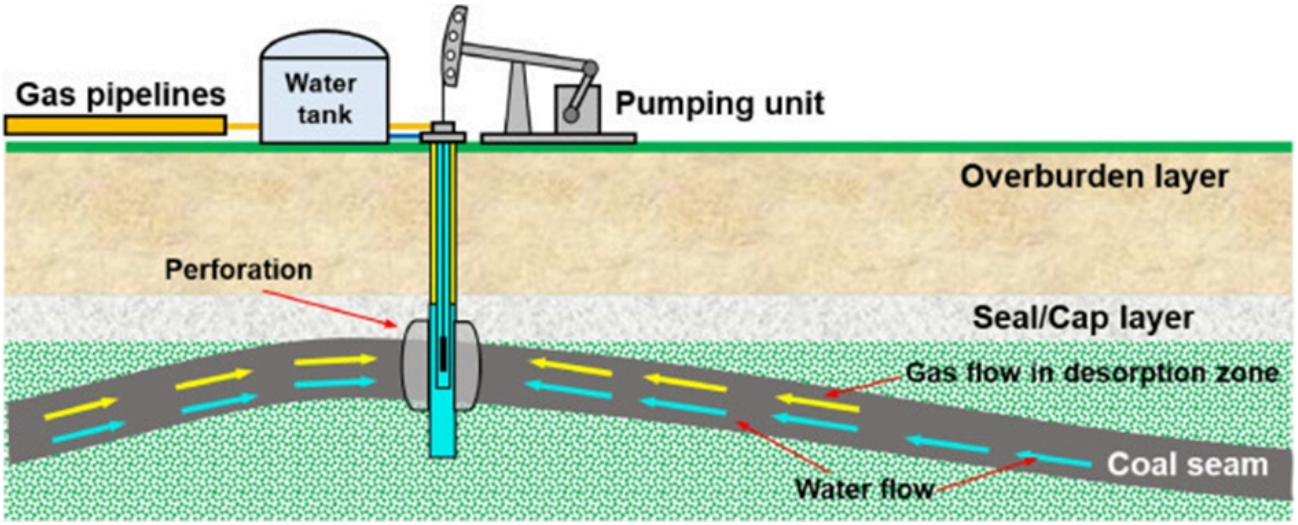


Figure 1. The production process of coalbed methane.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

In recent years, as coalbed methane extraction has been pursued alongside other unconventional energy sources, experts have leveraged artificial intelligence tools to fine-tune drainage protocols in producing wells, thereby securing the highest possible economic gains [16-25]. Optimizing coalbed methane output is intricate given CBM's singular properties and fluctuating nature. Clarkson and McGovern [16] developed an advanced CBM exploration instrument (CPT) by combining single-well reservoir simulators, gridded reservoir models, Monte Carlo simulations, and economic components. Extending integrated modeling principles, Du *et al*. [17] presented a novel technique for estimating CBM production potential by combining extreme gradient boosting (XGBoost) with Bayesian optimization. This investigation illustrated the promise of data-oriented methods for forecasting and enhancing CBM output. Wang *et al*. [18] emphasized forecasting output from deep CBM wells with Long Short-Term Memory (LSTM) combined with Bayesian optimization. Lv *et al*. [19] proposed a refined FOA-GRNN model for predicting coalbed methane output.

Validation via benchmark tests against alternative algorithms confirmed its superior accuracy and overall results. Yang *et al*. [20] suggested a physics-restricted neural network for forecasting production from CBM wells. They designed a GRU-MLP neural network combined with parameters refined using the NSGA-II algorithm. Jin *et al.* [21] focused on upgrading CBM assessment standards in the high-yield enrichment district of the southern Qinshui Basin in China. They investigated key factors influencing CBM well productivity and established boundary values for multiple evaluation metrics in this abundant area. Apart from standard machine learning and deep learning techniques, certain efforts have incorporated reinforcement learning into reservoir oversight and enhancement. Sun [22] proposed a deep reinforcement learning system employing the DQN algorithm for optimal multi-period scheduling in geosystem oversight, with a particular emphasis on carbon storage reservoirs. The system equips agents to pursue maximum rewards and hastens processing through deep multitask learning for swift estimation of complex, high-dimensional state changes [22]. Moreover, Li [23] developed a DRL model for reservoir management optimization. Experimental assessments utilized PSO and SVM algorithms as comparators to establish their performance. Mo *et al.* [24] formulated a deep autoregressive model for modeling state evolution. It leverages the system state at time t to anticipate the state at t + Δt, facilitating deep learning predictions within high-dimensional environments. Geneva and Zabaras [25] proposed a general technique for developing deep autoregressive models with random initial states.

To summarize, AI methods that integrate modeling, adopt data-centric strategies, and incorporate physics-constrained models have shown considerable promise for enhancing CBM production and reservoir oversight. Such techniques effectively navigate the intricacy and variability of CBM reservoirs, furnish reliable production projections, and steer development plans. Furthermore, reinforcement learning (RL) clearly has significant practical value and prospects.

This work advances a life-cycle production and drainage optimization plan for coalbed methane founded on Deep Reinforcement Learning (DRL). Using the Deep Q-Network (DQN) algorithm, the study develops and optimizes drainage approaches across the coalbed methane extraction sequence, enabling smart automation for drainage management. Initially, a CNN-LSTM deep learning model was implemented using data-driven techniques as a proxy for coalbed methane flow modeling. This proxy subsequently serves as the simulation setting within the DRL structure, enabling swift evaluations and decisions. Ultimately, the DQN algorithm within DRL trained drainage policies for 17 wells according to two reinforcement learning modes: 1) using gas production as reward and 2) using Net Present Value (NPV) as reward. Favorable outcomes emerged when compared with historical NPV profiles. This technique permits selecting drainage actions for multiple wells at each time interval — whether raising or lowering drainage rates — to achieve peak economic returns across the full coalbed methane operation.

Materials and Methods

This portion details the deployment of deep learning-based surrogate models as alternatives to classic forward numerical modeling of coalbed methane, together with the entire workflow for embedding this environment into a deep reinforcement learning system to train drainage policies for coalbed methane wells.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Governing equations for coalbed methane flow simulation

Drawing from the flow dynamics observed during coalbed methane extraction under standard adsorption scenarios, a numerical flow model was formulated with these assumptions [26-28]: 1) The coal seam includes an adsorption layer, a matrix pore network, and a fracture network. 2) Adsorbed coalbed methane resides exclusively in the adsorption layer of the matrix pores. The matrix functions as the storage medium for coalbed methane, whereas the fracture network provides the conduit for fluid movement. 3) Adsorption characteristics for liquids and gases in the coal seam comply with the Langmuir isothermal adsorption model. 4) The coal reservoir exhibits homogeneity and anisotropy. 5) Movement of gas and water inside coal fractures adheres to Darcy's law. 6) The coal seam shows compressibility, and water qualifies as a mildly compressible fluid. 7) The gas well links solely to the fracture network, excluding any direct matrix-to-wellbore flow.

Under these modeling premises, the governing equations describing two-phase gas and water flow inside the fracture system appear as follows [26]:

$$\nabla\cdot\left[\alpha\frac{k_a k_{frg}}{\mu_{fg}B_{fg}}\nabla\left(p_{fg}+\rho_{fg}gH\right)\right]+q_{fmg}-q_{vg}=\frac{\partial}{\partial t}\left(\frac{\phi_f s_{fg}}{B_{fg}}\right) \quad (1)$$

$$\nabla\cdot\left[\alpha\frac{k_a k_{frw}}{\mu_{fw}B_{fw}}\nabla\left(p_{fw}+\rho_{fw}gH\right)\right]+q_{fmw}-q_{vw}=\frac{\partial}{\partial t}\left(\frac{\phi_f s_{fw}}{B_{fw}}\right) \quad (2)$$

In **Eqs. 1 and 2**, α— Shape factor of the matrix block, decimal; H— Coal seam depth, m; φ_f — Fracture porosity, decimal; g— Gravitational acceleration, valued at 9.8 kg·m/s^2; p_{fw},p_{fg} — Pressures of the water and gas phases within the fracture-cleavage system, respectively, MPa; μ_{fw},μ_{fg} — Viscosities of the water and gas within the fracture-cleavage system, respectively, mPa·s; k_{frw},k_{frg} — Relative permeabilities of the water and gas within the fracture-cleavage system, respectively, decimal; s_{fw},s_{fg} — Saturations of the water and gas within the fracture-cleavage system, respectively, decimal; B_{fw}, B_{fg} — Volume factors of the water and gas within the fracture-cleavage system, respectively, decimal; ρ_{fw},ρ_{fg} — Densities of the water and gas within the fracture-cleavage system, respectively, kg/m^3; q_{vw},q_{vg} — Sink terms for gas and water in the fracture-cleavage grid, respectively, m^3/d; q_{fmw},q_{fmg} — Crossflow rates of water and gas between the fracture and matrix, respectively; m^3/d.

The Eclipse simulator performs forward modeling of coalbed methane production, supplying the training data for the autoregressive surrogate model.

Data-driven surrogate model in coalbed methane flow simulation

Agent training in reinforcement learning requires frequent interaction with the environment. Such interactions become highly demanding when relying on detailed physics-driven models. Hence, the current study creatively establishes a link between drainage controls and resulting gas output from coalbed methane using a CNN-LSTM deep learning model, thereby speeding up the search for optimal policies in the RL procedure.

For uncertainty analysis, repeated solution of the discretized versions of **Eqs. 1 and 2** across multiple drainage strategies — while holding initial and boundary conditions fixed — commonly necessitate hundreds or thousands of executions.

One individual simulation cycle can be represented as in **Eq. 3:**

$$x=f(m) \quad (3)$$

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Here, f stands for the reservoir simulation procedure and $\boldsymbol{m} \in \mathbb{R}^{n_w n_t}$ indicates the operational strategy applied to n_w wells over *nt* time steps, taken here as daily drainage rates. x captures the resulting daily gas production rates over all time steps for the n_w wells.

Surrogate modeling based on data employs an economical, non-invasive stand-in for the numerical simulator [29]. Progress in deep learning stands out for its capacity to detect pertinent data patterns while estimating mappings from inputs to outputs. Accordingly, the reservoir simulation procedure can be approximated via **Eq. 4:**

$$\boldsymbol{x} \approx \hat{\boldsymbol{x}} = \hat{f}(\boldsymbol{m}; \theta) \quad (4)$$

Here, $\widehat{f}$ signifies the surrogate model, $\widehat{x}$ indicates the approximated state outputs anticipated to mirror the simulated x values closely, and θ denotes the trainable parameters of the deep neural network. Consistent with **Eq. 4,** the aim is to create a surrogate that provides time-varying states corresponding to designated well-drainage strategies.

CNN-LSTM architecture

Convolutional neural networks (CNNs) identify localized patterns in signals by convolving filter kernels with input signal segments [30, 31]. The filter kernel applies uniformly across the input signal, yielding matching output feature maps. Every output map from a filter constitutes a feature map for the following layer, where layer depth corresponds to the count of filter kernels. The mathematical representation of the convolution process is detailed in **Eq. 5.**

$$\mathbf{F}_l = \sigma(\mathbf{W}_l * \mathbf{F}_{l-1} + \mathbf{b}_l) \quad (5)$$

$\mathbf{F}_{l-1}$ and $\mathbf{F}_l$ denote the feature maps from layers l-1 and l, respectively; σ represents a nonlinear activation function; $\mathbf{W}_l$ indicates a kernel matrix; ∗ marks the convolution process; and $\mathbf{b}_l$ is the bias term. Furthermore, feature maps across layers serve as transformations of the input map m (which comprises wells' drainage strategies), with $F_0 = \mathbf{m}$.

Long Short-Term Memory (LSTM) advances conventional Recurrent Neural Networks (RNNs) by mitigating the vanishing and exploding gradient problems [32]. Unlike basic RNNs, LSTM maintain and refresh essential data to prevent loss across extended multi-layer sequences [33]. In the LSTM setup, input signals first pass through the CNN layer before entering the LSTM cell. The forget gate regulates the retention of prior time-step information at the current step, as calculated by **Eq. 6.**

$$\mathbf{C}^t = \mathbf{f}^t \circ \mathbf{C}^{t-1} + \mathbf{i}^t \circ \widetilde{\mathbf{C}}^t \quad (6)$$

Here, $\mathbf{C}^{t-1}$ refers to the cell state carried from the prior time step, $\widetilde{C}^t$ is the candidate cell state, ° symbolizes the Hadamard product, f^t determines elements to remove from $\mathbf{C}^{t-1}$, and indicates elements to incorporate into $\widetilde{C}^t$. The output state H^t receives updates from C^t and $_O{}^t$, per **Eq. 7.**

Figure 2 displays the full data-handling workflow for the surrogate model using the combined CNN-LSTM structure adopted in this research. Numerical modeling of coalbed methane first generates pertinent operational and output data. Raw data is then preprocessed, including augmentation, sorting, noise removal, and splitting into datasets. These measures expand sample volume, clear disturbances, and balance the training, validation, and test portions. Following preprocessing, drainage volumes per well from the training data at the current time step and the preceding nt steps are supplied as input features to the network's initial segment. The network's latter segment yields predicted gas production per well for the current time step.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

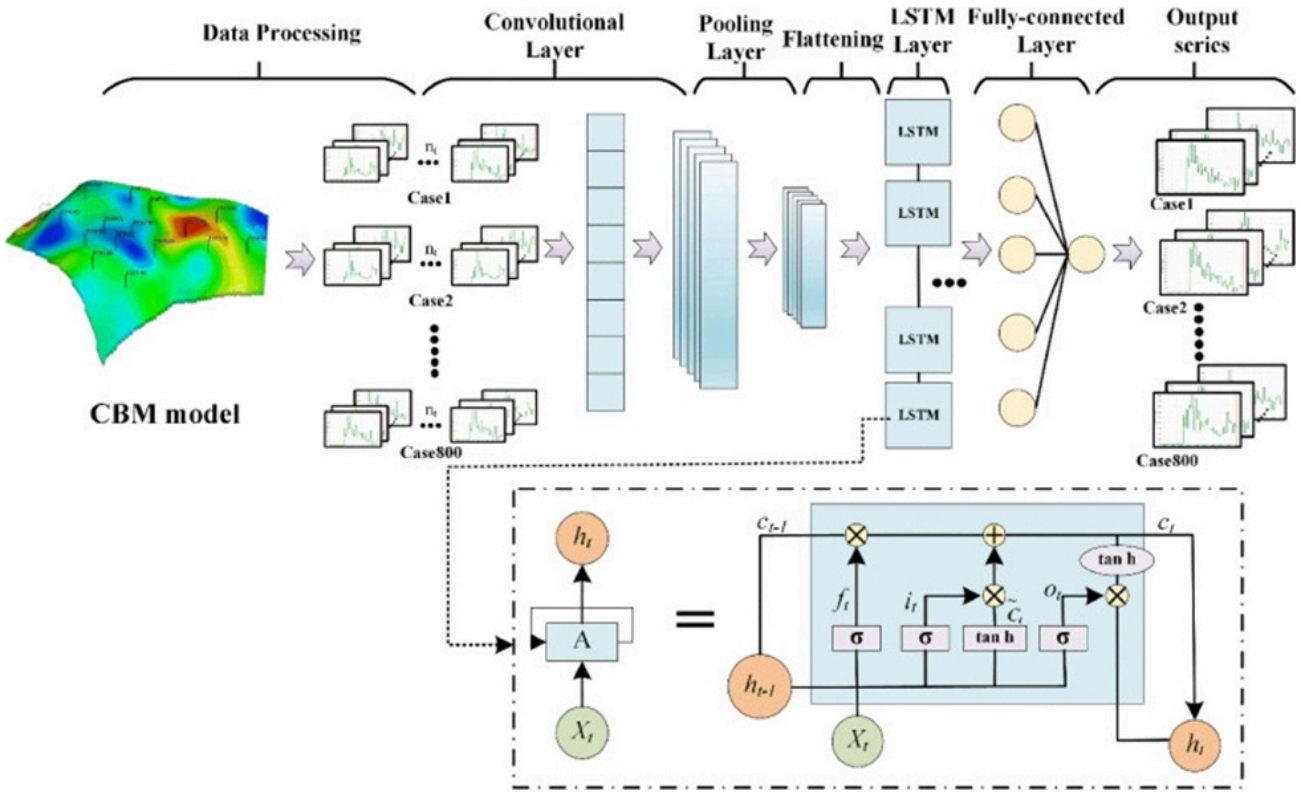


Figure 2. Surrogate model based on CNN-LSTM architecture.

This combined CNN-LSTM structure harnesses CNN's proficiency in extracting spatial features together with LSTM's effectiveness in processing sequential temporal data. As a result, it accurately detects the intertwined spatial and temporal patterns in coalbed methane production datasets, leading to markedly better predictive accuracy and trustworthiness.

The CNN-LSTM surrogate model serves as the rapid-state generator in deep reinforcement learning, efficiently advancing the overall procedure. The upcoming section examines the full research workflow involving deep reinforcement learning for optimizing drainage strategies in coalbed methane wells.

Training procedure: Preparing the data stands as an essential step for successful deep neural network development. Shifting data toward a zero-centered scale frequently improves training effectiveness. In this research, well control variables and response outputs exhibit varied numerical ranges. Min-max normalization was performed across all input variables. The main training goal is to narrow the gap between forecasts from the full-physics simulator and those from the surrogate. To measure the model's general effectiveness, gas loss Eg is introduced as:

$$E_g = \frac{1}{T}\frac{1}{M}\sum_{t=1}^{T}\sum_{i=1}^{M}|q_g^{it} - \widehat{q}_g^{it}| \quad (7)$$

Where T is the total calculating time, M is the total number of wells, is the gas rate from the full-physics model, is the gas rate from the surrogate model. Network parameters θ are continuously updated during training to reduce this loss. Gradients with respect to θ are computed automatically via backpropagation. The ADAM optimizer [34] is employed, serving as an enhanced variant of stochastic gradient descent (SGD) that demonstrates superior efficiency for numerous deep neural network structures. Effective surrogate training was completed, with the overall duration determined by

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

factors such as dataset volume, batch size, optimizer selection, learning rate settings, and GPU capacity. Convergence occurred within 60 minutes on an Nvidia RTX-4070 GPU with a batch size of 8.

Deep reinforcement learning based on DQN

Reinforcement learning represents a key area within machine learning. In essence, it focuses on preparing computational agents to establish precise correspondence between environmental states and ideal actions over a given period, ultimately securing the greatest reward. For example, in coalbed methane operations, such an agent can operate at the skill level of a seasoned operator [35]. **Figure 3** depicts a classic reinforcement learning configuration consisting of two primary elements: the environment and the agent. By observing the environment (state), executing actions, and receiving feedback, the agent progressively learns to maximize the objective (reward) and develops proficient strategies for complex situations.

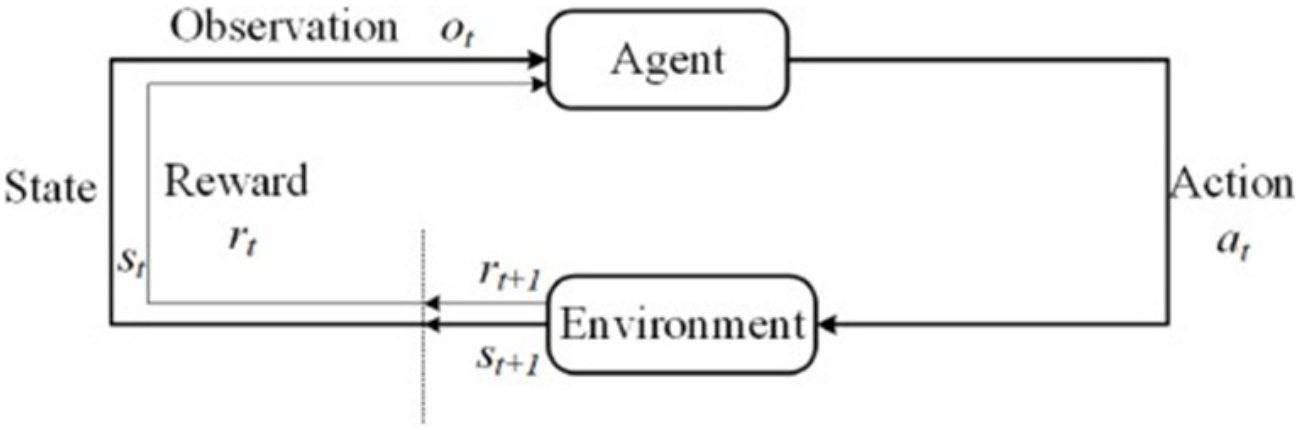


Figure 3. The standard system of reinforcement learning.

The Markov Decision Process (MDP) provides a systematic framework for describing ongoing interactions between the agent and the environment [36]. An MDP is a discrete-time control system that tracks the evolution of various factors over successive discrete intervals. Its basic components are captured in the tuple (S, A, T, r), where S stands for the collection of states, A for the collection of actions, and T for state transition probabilities given by T: S × A × S → [0, 1]. The "×" symbol denotes Cartesian product, so S × A × S includes combinations of the form (s, a, s'), with s as current state, a as chosen action, and s' as resulting state. [0, 1] covers the range for probability values. The reward function r: S × A → Ṙ links a numerical reward to each state-action pair. Policy π maps states to actions via π: S → A. Given a state, the policy dictates the action, and for s in S and a in A, π(s, a) ≥ 0 holds since negative probabilities are invalid. The agent's primary target is acquiring the best policy that maximizes expected cumulative reward.

Within an MDP, at each time step t, the environment begins in state st. The agent then selects an action according to the following policy π(s, a) [36]. In return, the environment provides reward rt and shifts to state st+1 based on transition probability T. A central assumption in MDP states that, given current state st and action at, both reward rt and next state st+1 remain unaffected by any history before t (even if randomness exists in st+1 or at). This property is termed the Markov property.

Consequently, running an MDP with policy π yields the sequence s_0,a_0 , r_0, s_1, a_1, r_1, ... Determining the optimal policy can involve learning a value function that quantifies the benefit of occupying a particular state. Accordingly, the state-action value function (Q function) [22] is formulated to represent expected future rewards from state s after action a, as shown in **Eq. 8.**

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

$$Q^{\pi}(s,a) = E\left(\sum_{k=0}^{\infty}\gamma^{k} r_{t+k} | s_t = s, a_t = a\right) \quad (8)$$

Where γ serves as the discount factor reflecting preference for later rewards and belongs to (0, 1], **Eq. 9** presents the recursive version of **Eq. 8**, known as the Bellman equation. It indicates that the value of action a in state s equals the immediate reward plus the discounted optimal future value after transitioning to s' via best action a'.

$$Q^{\pi}(s,a) = E(r_t + \gamma E(Q^{\pi}(s_{t+1}, a_{t+1}))|s_t = s, a_t = a) \quad (9)$$

Solving the MDP requires finding the optimal Q∗(s, a). A fundamental rule selects the action delivering maximum immediate reward per state while disregarding future consequences, as expressed in **Eq. 10** using transition probability T. This leads to a greedy policy π∗ given in **Eq. 11.**

$$Q^{\pi}(s,a) = \max_{a}\sum_{s'} T(s,a,s')\left(r(s,a) + \gamma \max_{a'} Q^{*}(s',a')\right) \quad (10)$$

$$\pi^{*} = \arg\max_{a} Q^{*}(s,a) \quad (11)$$

Q-learning provides an iterative, model-free approach to approximating the Q function, as detailed in **Eq. 12.** Here, α ∈ (0, 1] is the learning rate that controls the magnitude of the update, and k denotes the planning step. Updates typically follow **Eq. 12** at each step until a terminal state is reached.

$$Q_{k+1}(s_t, a_t) = Q_k(s_t, a_t) + \alpha\left(r_t + \gamma \max_{a} Q_k(s_{t+1}, a) - Q_k(s_t, a_t)\right) \quad (12)$$

At every step, a random value is sampled from U(0,1). When below ϵ, a random action from A is taken; otherwise, the action maximizing Q(s, a) is chosen. User-set ϵ controls the ϵ-greedy balance between exploration and exploitation. Normally, ϵ begins at 1 and gradually declines. Multiple planning cycles, each called a round, are executed. In principle, Q-learning converges to optimality as the number of rounds increases.

To speed up the process, function approximators estimate Q(s, a). Past attempts at training these approximators were often unstable. Deep Q-Learning (DQL) [37] marked a major improvement. It deploys two networks: the main Deep Q Network (DQN) Q(s, a; θ) and a target network. The target computes Y(θ−) = r + maxa′ Q(s′, a′; θ−) with parameters θ−. Initial weights match for both. Training updates θ via gradient descent to minimize squared error loss in **Eq. 13:**

$$\varsigma(\theta) = E\left[\left(\left(Y^{\theta^{-}}\right) - \tilde{Q}(s,a;\theta)\right)^{2}\right] \quad (13)$$

Where ς denotes the loss, target network weights update periodically for stability, either by a full copy of θ every set iterations or through soft updates in **Eq. 14:**

$$\theta^{-} \leftarrow \tau\theta + (1-\tau)\theta^{-} \quad (14)$$

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Where τ ≪ 1 is a small user parameter allowing gradual alignment. This slow change in target parameters greatly enhances training stability. Exploration is further supported by experience replay, which stores (s, a, s', r) tuples in a memory buffer. Training samples are randomly sampled from this buffer rather than in a sequential order, reducing temporal correlations, improving distribution uniformity, limiting overfitting, and accelerating convergence.

Figure 4 shows the fundamental workflow of the Deep Q-Network (DQN) algorithm [38, 39]. The integration of deep networks with Q-learning enables effective handling of policy optimization in large state spaces. This study applied DQN to a thorough evaluation of the action-value function Q associated with the intelligent agent's drainage actions in the CBM setting. DQN loss arises specifically from discrepancies between the prediction network's outputs and the target network's values [38].

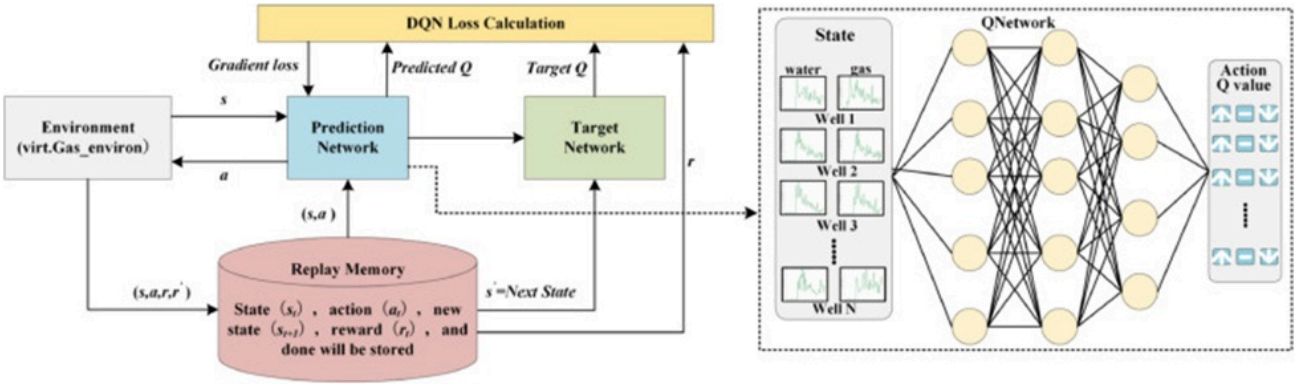


Figure 4. DQN algorithm.

The paired networks jointly refine the action-value function. The prediction network interacts with the environment to select actions from current states, while the target network provides stable references via delayed updates. Replay memory stores full experience records (state, action, reward, next state, terminal flag). Random batch sampling from past experiences eliminates time dependencies, markedly boosting training reliability and speed. This setup allows the agent to master and improve strategies inside the simulated gas production environment.

The present study develops an innovative time-series deep learning surrogate representing coalbed methane drainage actions and corresponding gas production to expedite policy discovery.

Commencing from the standard forward model in **Eq. 15:**

$$F(\mathbf{s}, \mathbf{q}, \boldsymbol{\Lambda}, \boldsymbol{\xi}) = 0 \quad (15)$$

In this equation, F(·) represents a model operator; s refers to all state variables, which can depend on both space and time; Λ is used to represent the set of model parameters; ξ indicates the boundary and initial conditions; and q stands for the forcing terms. For state–action predictions, the system given in **Eq. 16** can be represented in a discrete-time, state-space format as:

$$\mathbf{s}_{t+1} = f(\mathbf{s}_t, \mathbf{q}_t, \boldsymbol{\Lambda}, \boldsymbol{\xi}_t) + \boldsymbol{\eta}_t \quad (16)$$

Where $\hat{f}$ refers to the data-driven surrogate model mentioned above, $\hat{s}_{t+1}=\hat{f}(\hat{s}_t, q_t, \Lambda, \xi_t) + \eta_t$, in which θ represents the network parameters. The complete optimization problem for the drainage strategy of coalbed methane wells based on

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Deep Q Learning is shown in **Figure 5.**

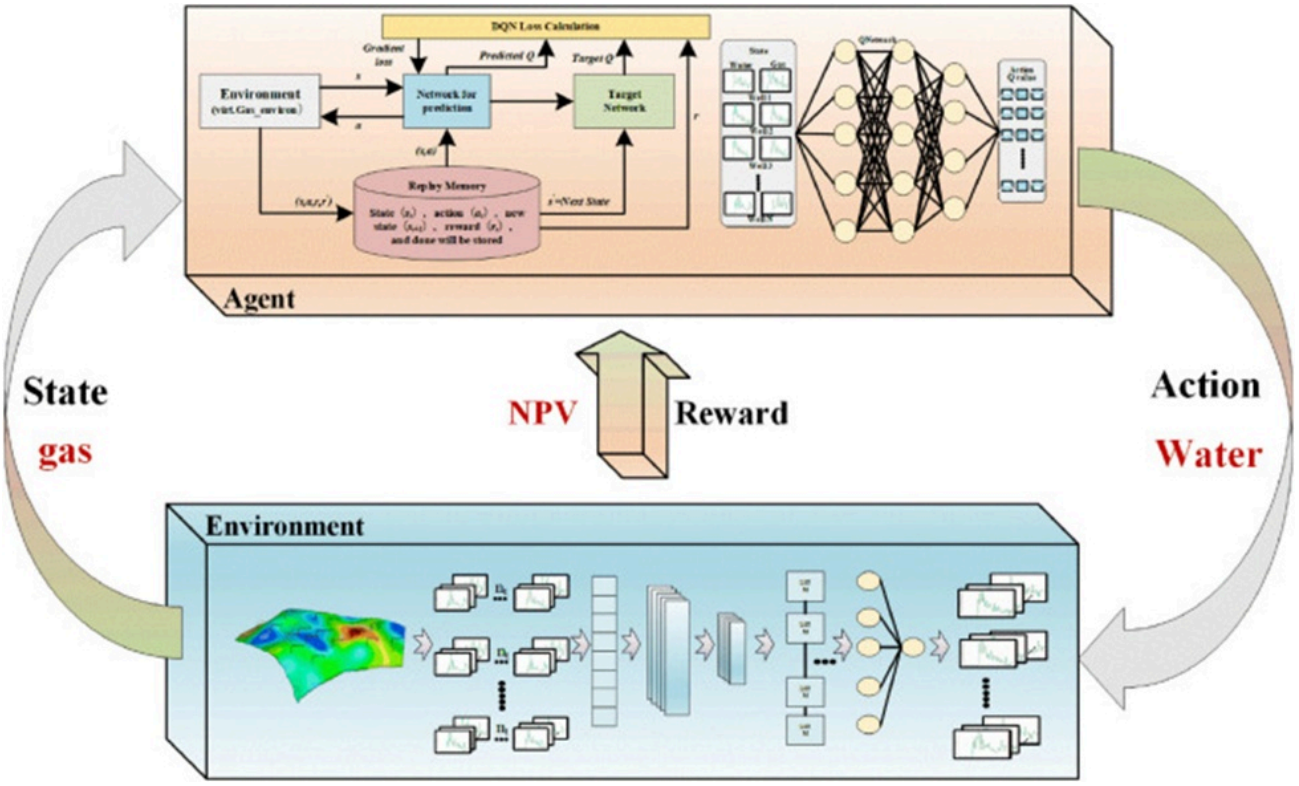


Figure 5. A deep reinforcement learning workflow for coalbed methane wells drainage strategy.

It illustrates the overall workflow of the DRL approach for optimizing CBM well drainage strategies. The core of this workflow is the interaction between the agent and the environment. The environment receives actions from the agent, simulates actual CBM extraction scenarios using a surrogate model, and simultaneously returns state information to the agent. The agent, utilizing the Deep Q-Network (DQN) algorithm and a predefined reward function, continuously learns through interactions with the environment, progressively developing an optimal strategy. Ultimately, the agent maximizes cumulative rewards through optimized strategies, thereby achieving our algorithm.

Results and Discussion

This section demonstrates the use of deep reinforcement learning on a coalbed methane reservoir model incorporating actual production records from 17 wells. Through deep reinforcement learning, drainage strategies for individual wells can be refined based on past performance records, leading to higher overall economic returns and output levels.

Data-driven surrogate model results

Simulations and training datasets were generated using the CBM module in Eclipse to produce gas and water output profiles for the target reservoir. The studied block has already achieved a successful history match in reservoir simulation and includes 4 years of actual production records**.** **Figure 6** and **Table 1,** respectively, show the gas content distribution throughout the region, the well placement, and the key input parameters used in the simulation model.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

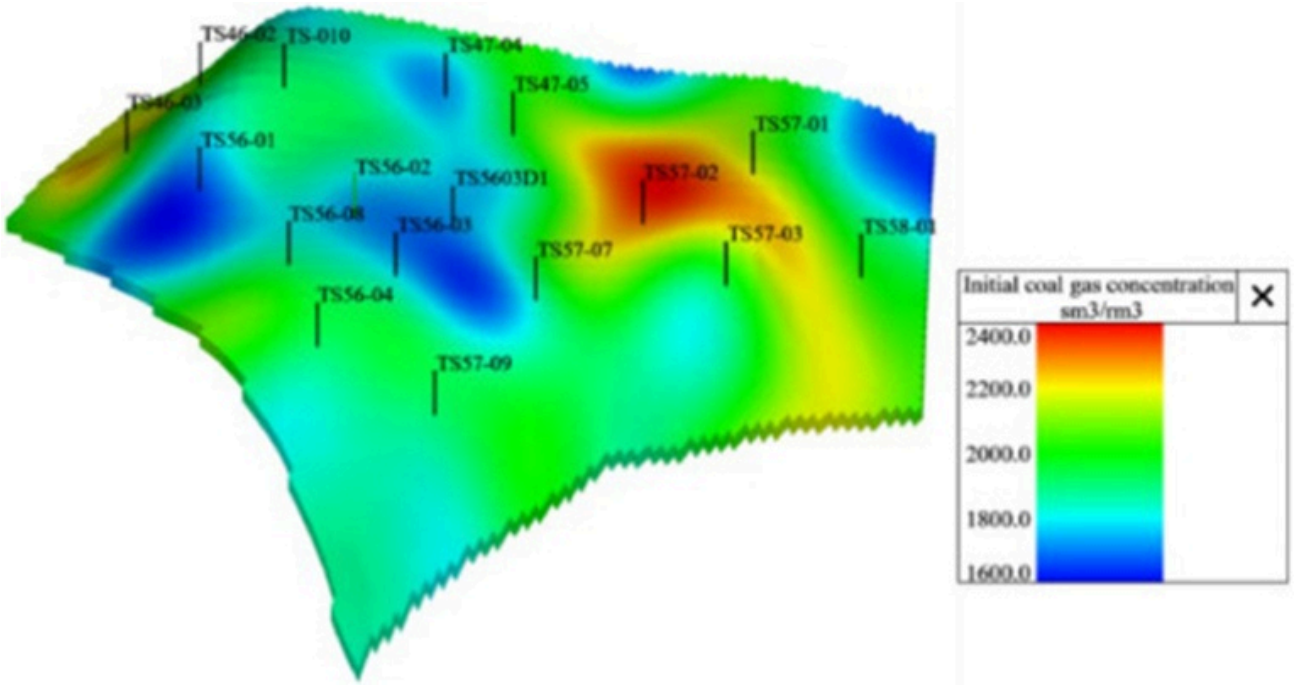


Figure 6. Gas content distribution and 17 wells deployment.

Table 1. Simulation model input parameters.

Parameters	Value	Parameters	Value
Coal cleat porosity, dimensionless	0.01	Grid size, m × m × m	40 × 40 × 5
Coal cleat *x* permeability, mD	100.0	Coal thickness, m	5.0
Coal cleat *y* permeability, mD	100.0	Initial pressure, MPa	5.0
Coal cleat z permeability, mD	10.0	Maximum gas content, m^3/t	500.0
Coal water saturation, %	95.0	Absorbed gas reserve, $10^8 m^3$	17.8
Coal compressibility, $10^{-4} MPa^{-1}$	2.354	Free gas reserve, $10^8 m^3$	0.001

The CNN-LSTM-based surrogate model was trained on historical production records from the 17 coalbed methane wells. A total of 1000 CBM model realizations were produced as the training dataset by varying the historical drainage parameters (by ±50%) via the Monte Carlo method. Corresponding gas production outputs for these 1000 drainage scenarios were then simulated using Eclipse. Drainage data served as the input variable, while gas production served as the output for training the surrogate model. The objective during training was to reduce gas loss, e.g., by optimizing the network parameters θ. **Figure 7** displays the training progression, highlighting how the loss function converged over successive iterations. The loss function stabilized after roughly 40 epochs, confirming that the model successfully captured the intricate relationships between drainage approaches and resulting gas yields.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

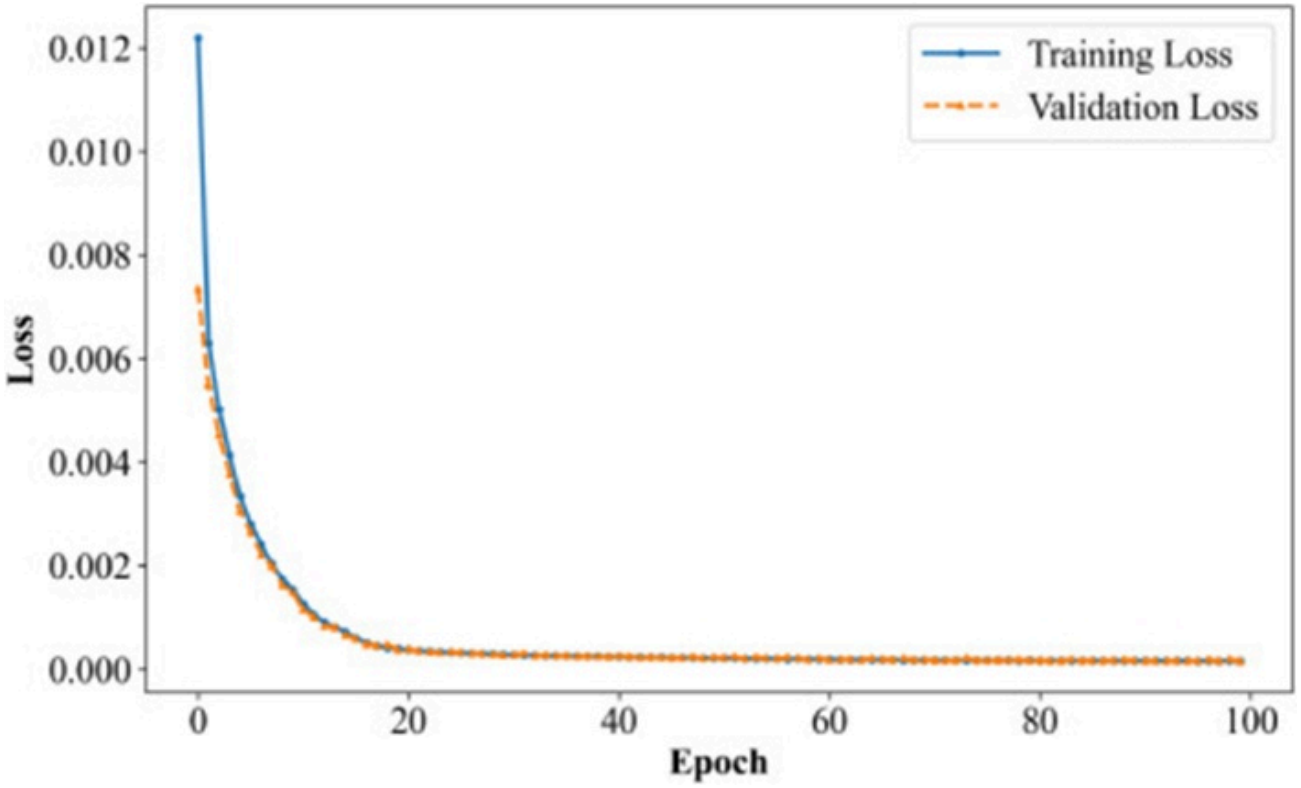


Figure 7. Training and validation loss curve.

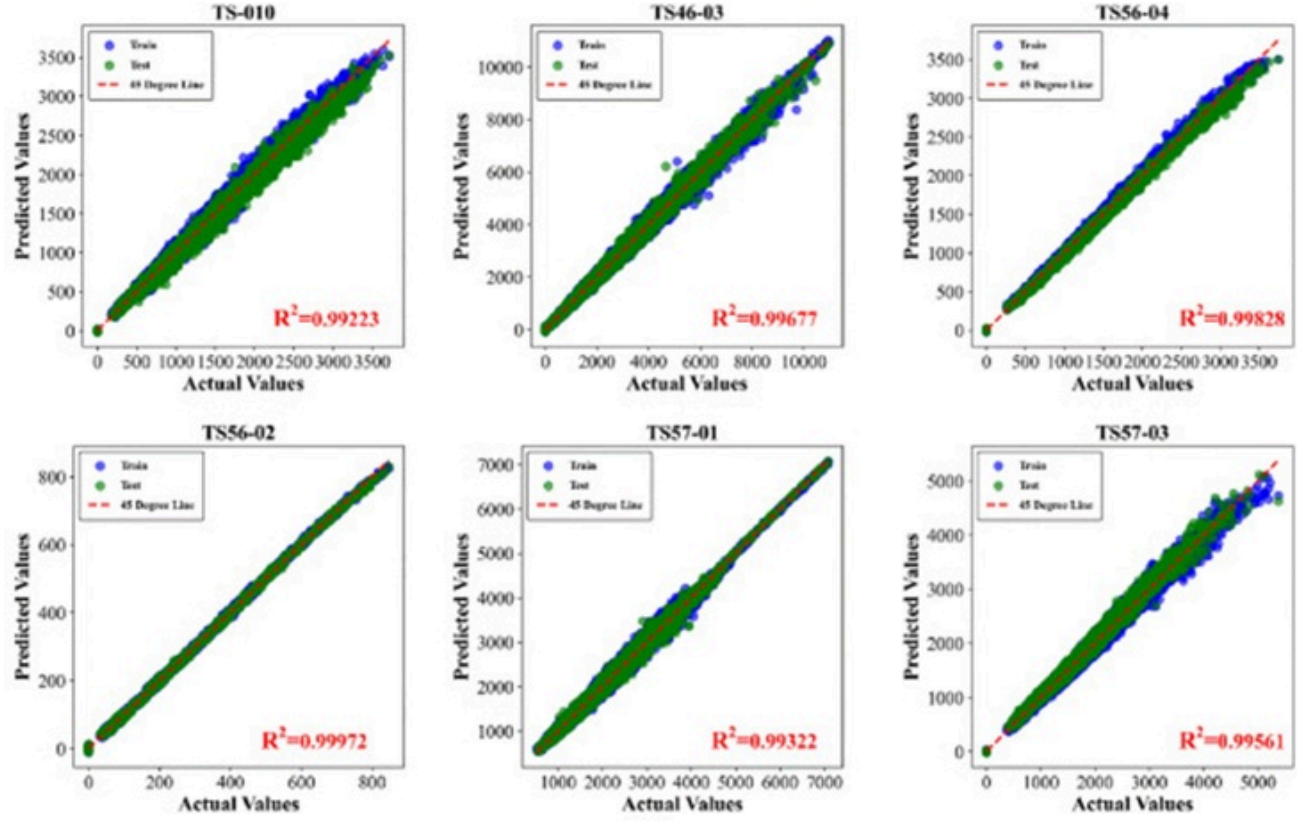
The coefficient of determination (R^2) is an important measure of the extent to which independent variables explain variation in the dependent variable [40], serving as a vital indicator of predictive quality. Its calculation formula is given by **Eq. 17:**

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (17)$$

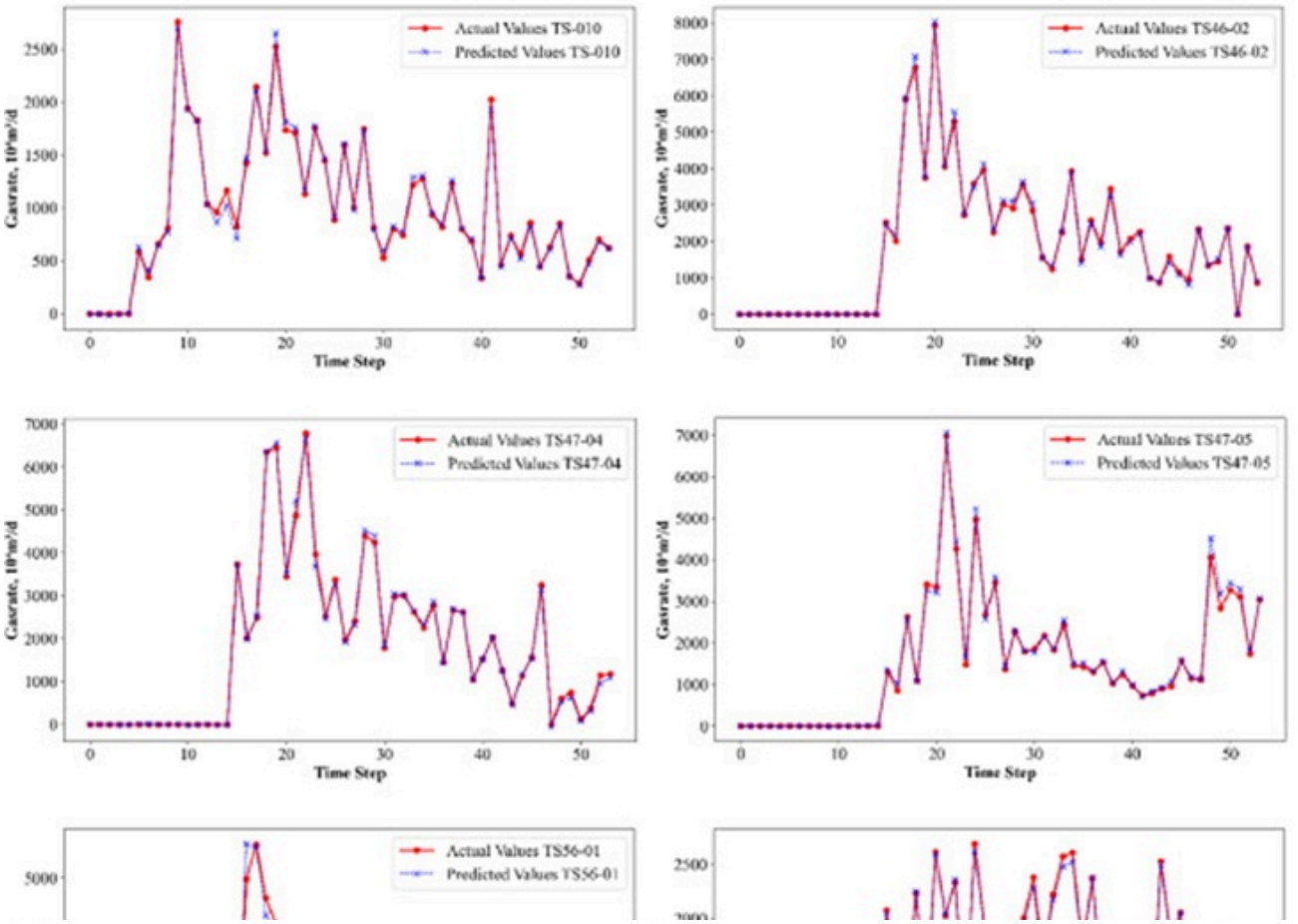
Here, y_i denotes the actual values, y_i represents the predicted values, the mean of the actual values, and n denotes the sample size. The R^2 value lies in the interval [0,1], with values closer to 1 indicating greater model accuracy and better alignment between predictions and observations. Following training, the coefficient of determination (R^2) was computed for each well within the surrogate model, yielding an average of 0.98 and indicating very minimal prediction deviation.

To better evaluate model effectiveness, 45-degree line plots with their corresponding R^2 values were generated for six randomly selected wells **(Figure 8a)**. Both training and test points are well-balanced around the 45-degree line, indicating close agreement between forecasted and observed outputs. Furthermore, surrogate model predictions were benchmarked against numerical model outcomes for four wells **(Figure 8b)**. The findings reveal that the surrogate model offers superior generalization while maintaining strong agreement with numerical simulation outputs, thereby confirming its practical dependability.

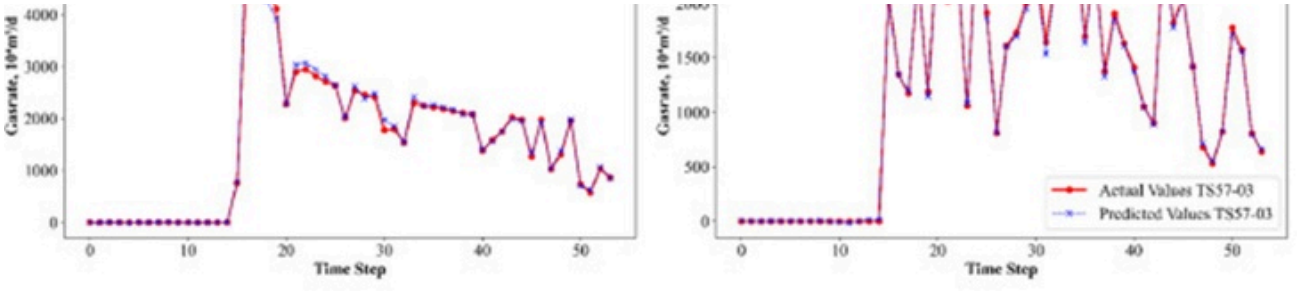
© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press



(a)



© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press



(b)

behaviors in coalbed methane systems effectively.

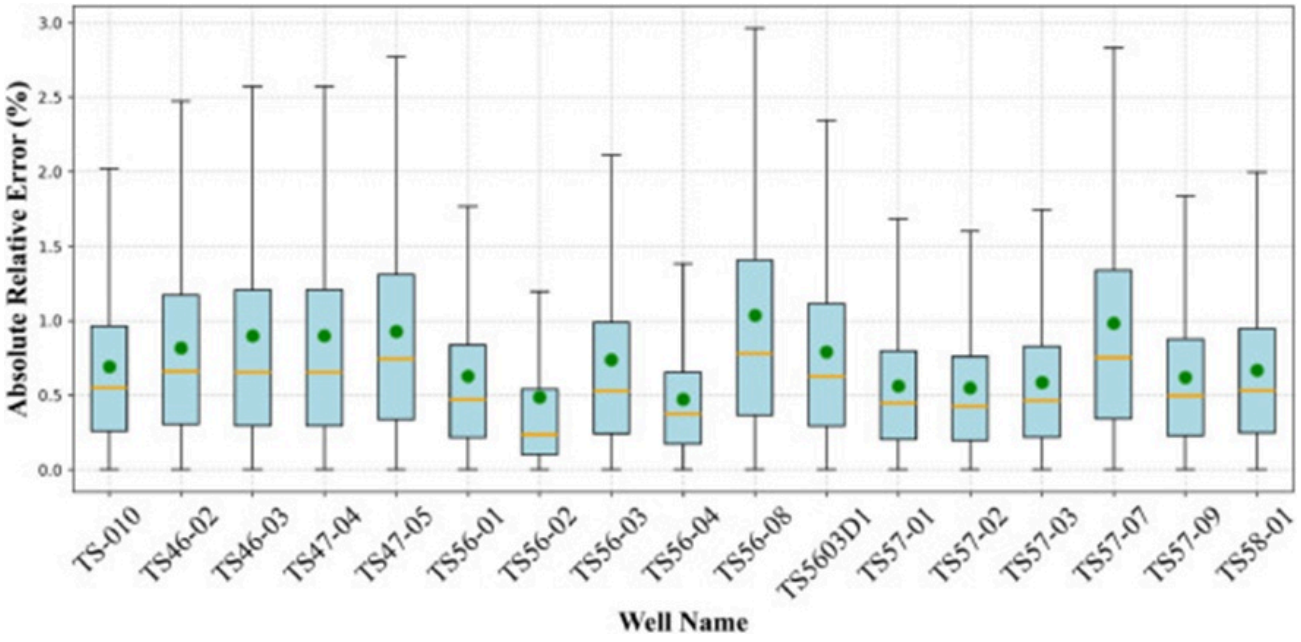


Figure 9. The absolute relative error of gas rate for each well.

Deep reinforcement learning results

A deep reinforcement learning (DRL) system based on the DQN algorithm was implemented to optimize drainage plans across 17 CBM wells. Within this setup, crafting an appropriate reward function is essential for effective strategy development. The current work examined two distinct reward formulations: one centered on total gas output and another on Net Present Value (NPV). Both were utilized to refine drainage approaches for the CBM wells.

The reward function, which focuses on the gas production rate, seeks to maximize total gas recovery from the wells. Its underlying principle emphasizes direct improvement of the production objective by promoting better gas desorption and extraction through refined drainage control. This reward function is formally defined in **Eq. 18:**

$$R_{gas} = \sum_{t=1}^{T} Q_{gas}(t) \quad (18)$$

In this context, Rgas represents the cumulative reward; Qgas(t) denotes the gas production at time step t, m3/d; and T is the total number of time steps. Prioritizing gas production directly provides a straightforward target, ideal when the main

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

priority is achieving maximum output volumes. The agent was trained by modifying water discharge volumes at each time step as actions (increase 10%, maintain, decrease 10%). These actions prompted the environment to provide updated gas production values from the wells in the new state. Training proceeded iteratively with the DQN algorithm guided by the specified reward objective. Computations were performed on a Windows 10 system equipped with an RTX 4070 GPU. A total of 2000 training episodes were executed, with the loss function converging near 750 episodes and the reward curve leveling off at approximately the same point. **Figure 10** presents the DQN agent's training dynamics under this reward structure, featuring both the loss and reward trajectories. The loss curve **(Figure 10a)** illustrates progressive Q-network convergence as values decline consistently. The reward curve **(Figure 10b)** exhibits steady growth in total reward, confirming the agent's success in mastering optimized drainage tactics.

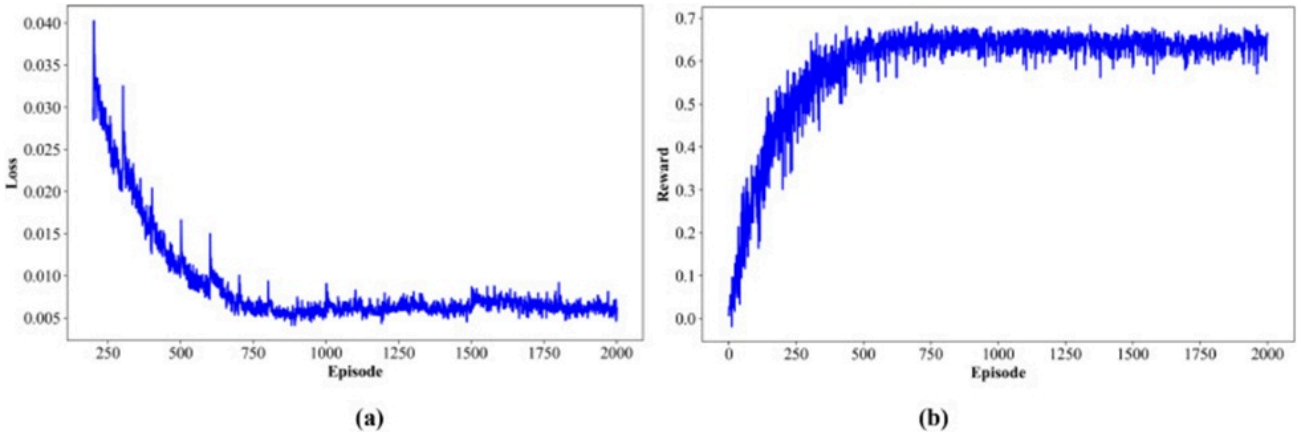


Figure 10. (a) The convergence of the Q-network; (b) Cumulative reward of total gas rate.

Figure 11 provides a 3D representation of how actions evolved spatially for each well throughout the training phases. Data points were captured at 300-episode intervals. The x-axis indicates well identifiers, the y-axis shows time steps, and the z-axis denotes the action type ("→" for no change,"↑" for increased drainage,"↓" for decreased drainage). This plot illustrates the pattern of drainage decisions across wells and time periods.

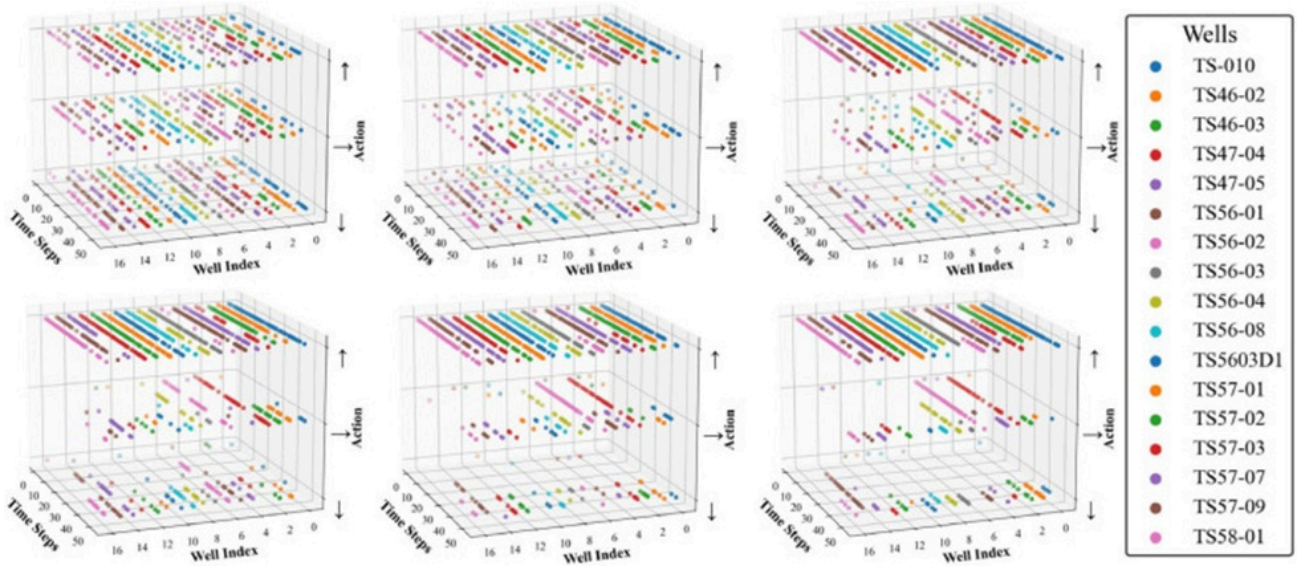


Figure 11. (a) Three-dimensional visualization for the action space of each well.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Results indicate that actions began in a scattered, random pattern but progressively shifted toward configurations that improved rewards as training advanced. The agent continuously refined its drainage adjustments. Given that rewards were tied to CBM output, the agent predominantly increased drainage rates to accelerate pressure drawdown and gas release. This pattern is clearly visible in the visualization and underscores the agent's capacity to select actions that boost overall CBM recovery. While some actions avoided increases, the majority trended upward.

As illustrated in **Figure 12**, real historical rewards are compared with those obtained from the surrogate model and the full simulator when guided by reinforcement learning decisions. Both the surrogate model and the gas-production-focused simulator delivered superior rewards under the learned policies relative to actual historical performance. This suggests that incorporating gas production considerations into reinforcement learning yields improved output in both simulation and surrogate environments. These observations confirm the strength and advantage of the gas-production-based reward approach.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

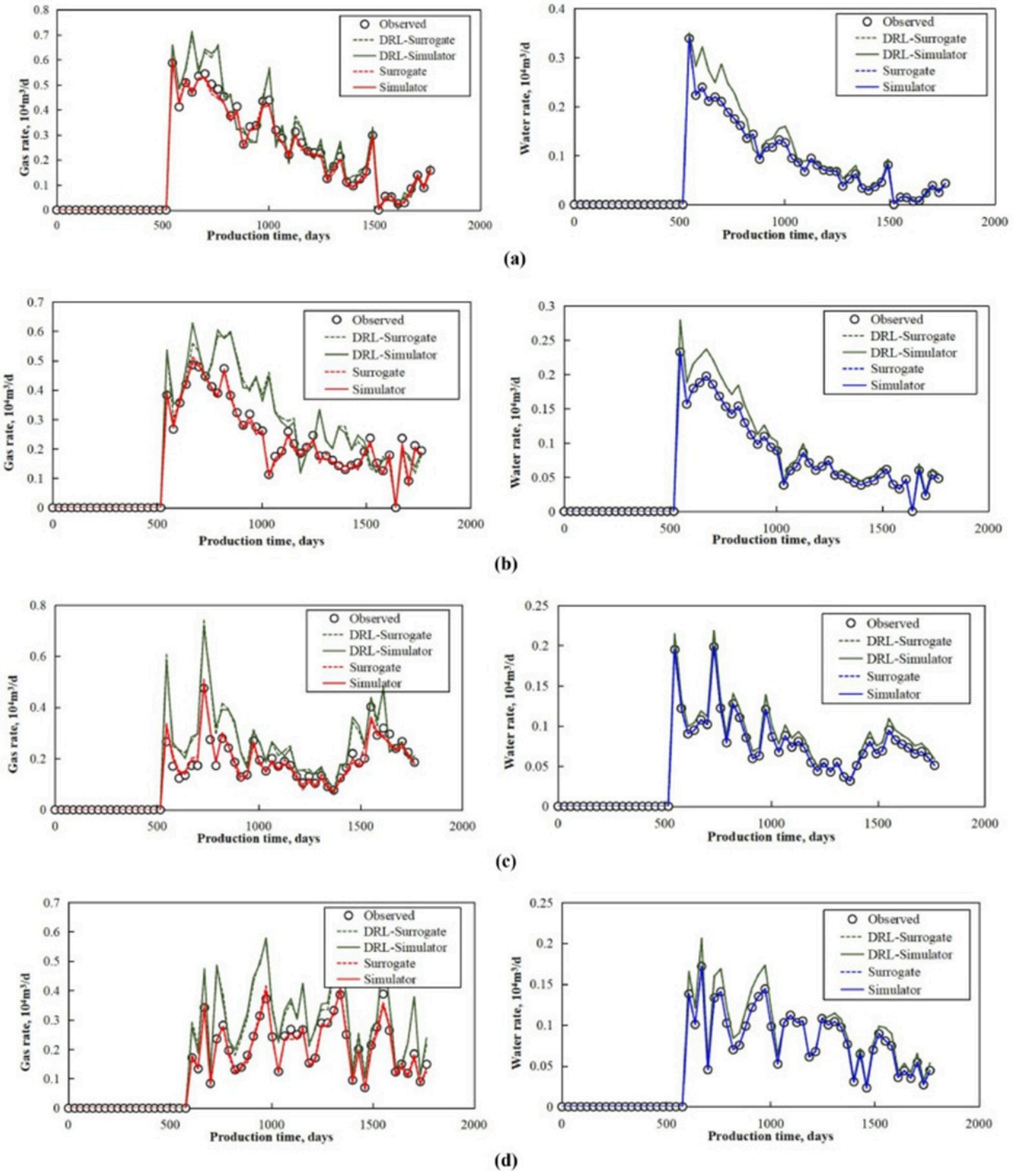


Figure 12. Comparison of drainage strategies for four coal seam wells considering cumulative gas production as a reward: based on deep reinforcement learning vs historical observations. (a)TS46-02; (b)TS46-03; (c)TS47-05; (d)TS56-08.

In **Figure 12,** the solid lines represent numerical simulation outputs, the dashed lines indicate surrogate model results, and the scatter points show historical data. The surrogate model and simulator curves for the four wells demonstrate strong generalization by the surrogate. The close match between green solid and dashed lines (reinforcement learning

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

outputs transferred to the simulator) indicates low discrepancy. Additionally, learned drainage strategies generally exceeded historical levels during the production phase without excessive deviation. This moderation arises from accounting for reservoir energy limits, as overly aggressive early drainage would cause rapid energy loss. Consequently, strategies for all four wells feature higher rates early on and more moderate rates later, ultimately producing greater cumulative gas volumes than historical benchmarks.

A second reward structure was developed in this work to enhance the economic performance of coalbed methane wells by integrating gas output, drainage expenses, and additional financial considerations [41-43]. Net Present Value (NPV) was selected as the reward signal, given its role as a standard indicator of project economic worth. This method balances production gains against operational costs and better reflects actual field conditions, making it especially relevant when profitability is the central concern. The economic calculation formula appears in **Table 2.**

Table 2. Calculation of economic parameters.

Economic parameters	Equation	Economic parameters	Equation
Revenue	$R_t = P_{brent} \times Q_{gas}^t$	Cash Flow	$CF_t = FI_t - CAPEX_t$
Profit Before Taxes	$PBT_t = R_t - OPEX_t$	Net Present Value (NPV)	$NPV = IC + \sum_{t=1}^{n} \frac{CF_t}{(1+i)^t}$
Financial Income	$FI_t = PBT_t - TAXES_t$	Rward	$R_{NPV} = \sum_{t=1}^{T} NPV(t)$

Here, Rt represents the Revenue; is the gas total (108/m3); Pbrent is the gas price ($/m3); i is the Discount Rate; and IC stands for the Initial Investments (disregarded). For NPV computation, the relevant parameters are detailed in **Table 3,** including recent average international natural gas prices, export rates, tax rates, drilling costs, completion costs, and ongoing production costs. These values enable rapid assessment of field-level annual returns.

Table 3. Input parameters for economic calculating.

Parameter Category	Measurement Unit	Value
Brent Gas Market Price	$ / m3	0.40
Share of Gas Export	%	30.0
Taxation:		
• Extraction (Severance) Tax	%	10.0
• Asset (Wealth) Tax	%	2.0
• Corporate Income Tax	%	20.0
Operating Costs (OPEX):		
• Gas Production Operating Cost	$ / m3	0.01
• Water Disposal and Drainage Cost	$ / m3	0.20
• Maintenance Cost for Active Wells	$ / well / year	1350000.00
• Maintenance Cost for Inactive Wells	$ / well / year	350000.00
• Well Perforation Expense	$ / m	67.0
• Miscellaneous Annual Operating Expenses	$ / year	12000000.00

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Drilling and Well Construction Cost	$ / well	1500000.00
Well Closure and Abandonment Cost	$ / well	135000.00
Expected Well Lifetime	years	10
Discount Factor / Rate	%	10
Upfront Capital Investment	$	0.0

Agent training involved modifying discharge volumes per well at each step (increase by 10%, hold, decrease by 10%). The environment responded with updated gas production figures for the state. Iterative DQN training used NPV maximization as the target. The reward curve stabilized near 750 episodes, plateauing at around 0.6 — lower than the plateau observed with gas production as the reward. **Figure 13** depicts the DQN training trajectory under this economic-reward setup, showing both loss and reward progressions. The loss curve **(Figure 13a)** shows steady improvement in the Q-network, while the reward curve **(Figure 13b)** reflects consistent growth, demonstrating the agent's effective strategy optimization.

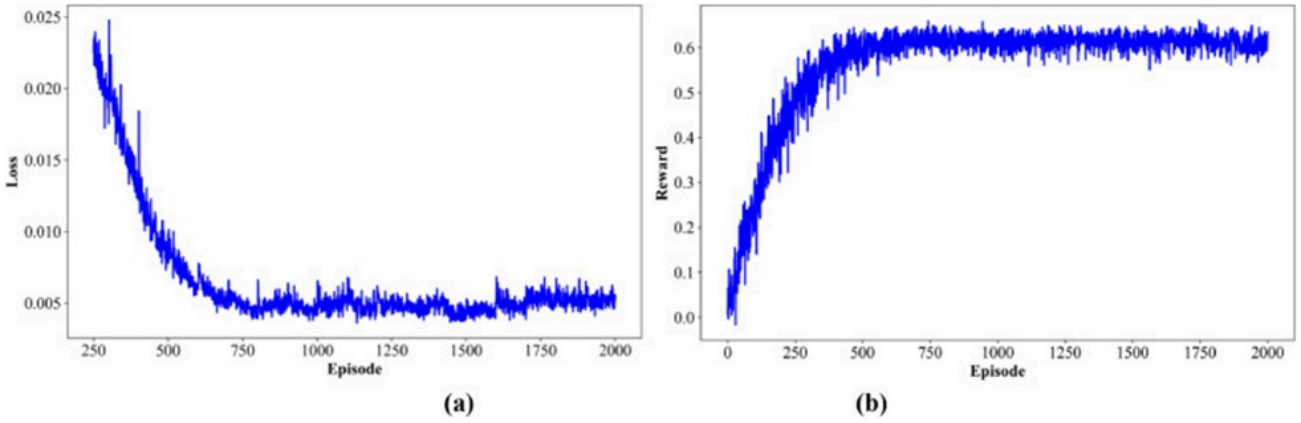


Figure 13. (a) The convergence of the Q-network; (b) Cumulative reward of NPV.

Figure 14 illustrates the three-dimensional evolution of the action space for each well during training, with NPV as the reward. Samples were taken every 300 episodes. The x-axis denotes well names, the y-axis time steps, and the z-axis action categories ("→" for no change,"↑" for increased drainage,"↓" for decreased drainage). This view captures drainage decision patterns across wells and time intervals.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

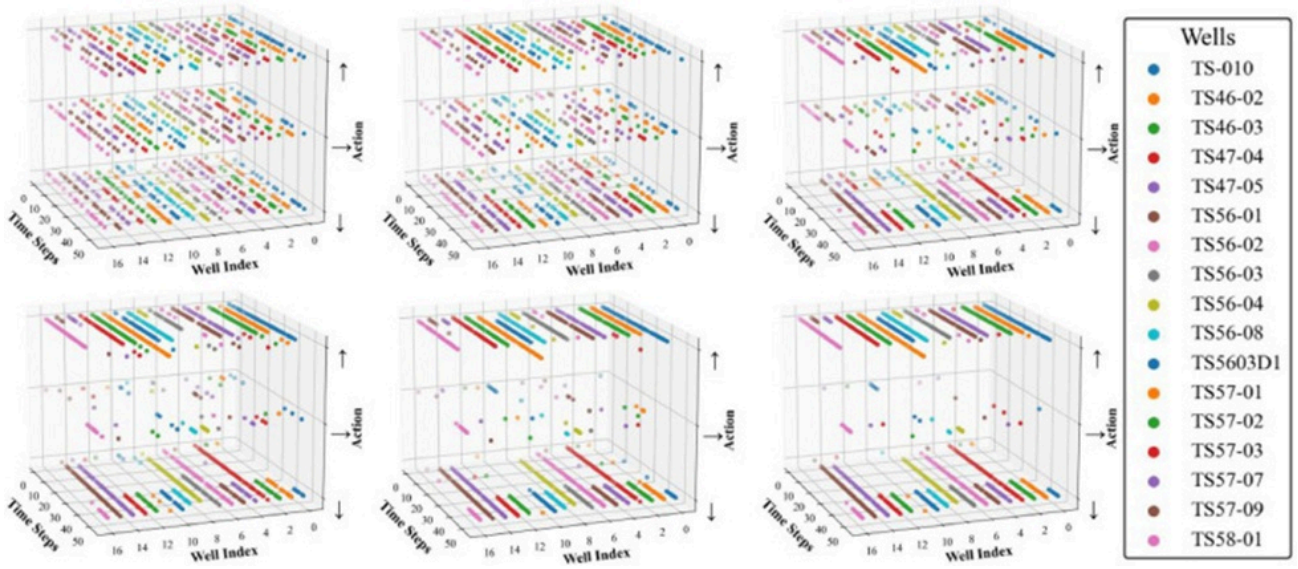


Figure 14. Three-dimensional visualization for the action space of each well.

The outcomes reveal that actions started randomly but evolved toward reward-maximizing patterns over time. With NPV as the objective, the agent initially boosted drainage in most wells to establish flow pathways, then later reduced it to control costs and preserve reservoir energy. This approach differs markedly from the first reward scheme, which relied exclusively on gas volume.

The agent's capacity for adaptive drainage control demonstrates its value in NPV optimization, effectively balancing short-term output with sustained economic and reservoir considerations.

Figure 15 compares historical rewards against those achieved by the surrogate model and simulator when following reinforcement learning policies. Both environments produced higher rewards than historical baselines under the learned strategies. This confirms that integrating economic factors into reinforcement learning enables improved financial outcomes in simulation and surrogate settings alike, validating the NPV-based reward design.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

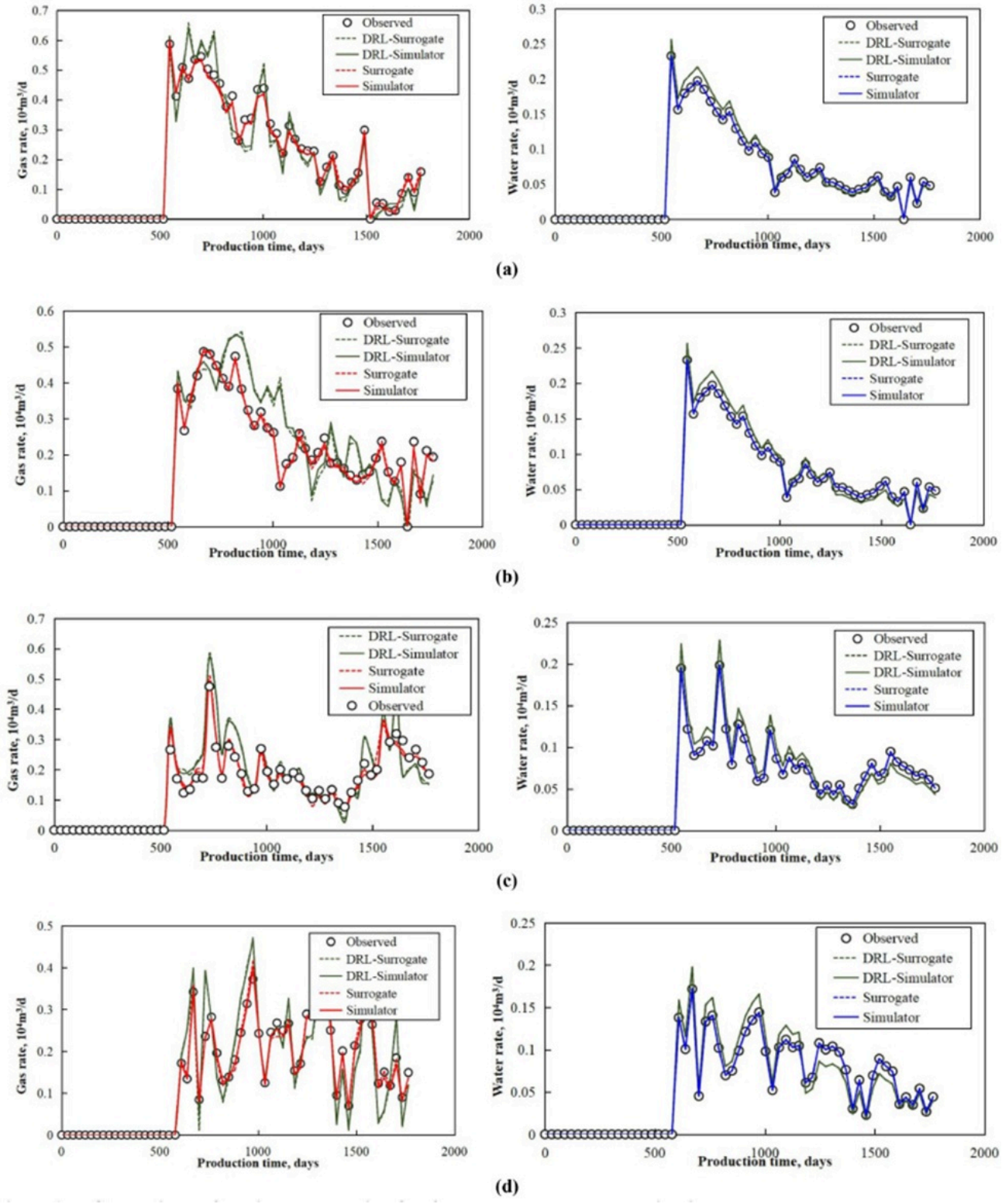


Figure 15. Comparison of drainage strategies for four coal seam wells considering NPV as a reward: based on deep reinforcement learning vs. historical observations. (a)TS46-02; (b)TS46-03; (c)TS47-05; (d)TS56-08.

Comparing **Figure 15** with **Figure 12** under economic reward constraints shows that drainage strategies for the four wells do not consistently surpass historical levels. This stems from higher drainage volumes increasing costs during learning,

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

combined with the generally lower productivity of CBM wells relative to conventional reservoirs, necessitating careful economic evaluation. Across the plots, reinforcement learning strategies typically exceed historical rates in early periods and fall below them later. This aligns with CBM production characteristics, requiring aggressive initial drainage to establish seepage paths followed by moderated rates to sustain output.

Figure 16 presents NPV profiles for the studied CBM field under historical operations and the two reinforcement learning approaches (gas-production-based and NPV-based rewards). The NPV-based strategy clearly delivers superior gains, increasing cumulative NPV by 5.83% relative to historical drainage. In contrast, the gas-production-focused strategy yielded an NPV below historical levels, highlighting the critical role of economic considerations in coalbed methane drainage planning.

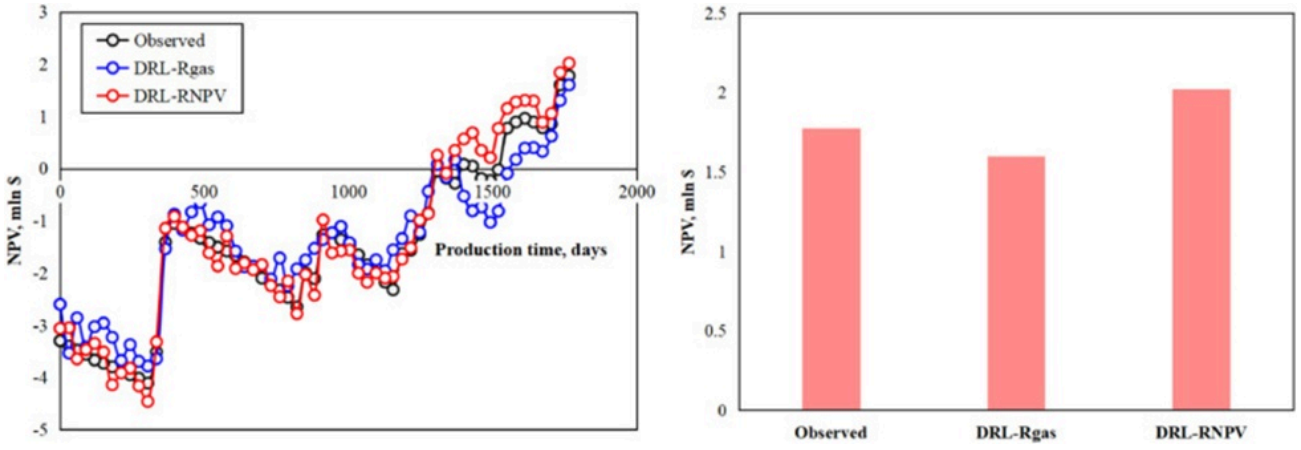


Figure 16. NPV curves of historical drainage strategy results (observations) vs. two reinforcement learning drainage strategy results (based on cumulative gas production vs. based on NPV as a reward).

Conclusion

This study introduces an Auto-Regressive Surrogate Model constructed with deep neural networks to serve as the environment for reinforcement learning in gas production and drainage optimization. The model successfully substitutes for resource-intensive, high-dimensional simulators, greatly speeding up agent-environment interactions. Reward functions were formulated around both gas production quantities and long-term economic returns. Testing under standard conditions confirms the method's practicality and advantages. Compared with conventional techniques, the approach markedly increases Net Present Value (NPV) while refining drainage profiles.

The framework described here establishes that the DRL-driven full-cycle drainage optimization for CBM reservoirs is both efficient and viable, offering a powerful intelligent option for unconventional gas resource development. Results emphasize its robust flexibility and generalization potential, positioning it as a valuable asset for tackling intricate optimization challenges in varying production settings.

Ultimately, the deep reinforcement learning system developed in this research provides a useful and reliable means for addressing diverse sequential decision problems that involve complex, high-dimensional models. Future efforts will aim to broaden the framework to support multiple action types and test it under more challenging geological scenarios, thereby increasing its relevance and utility in unconventional natural gas operations.

Acknowledgements

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

We thank Tracy Energy Technologies for modeling and numerical simulation in this study. We also thank China University of Geosciences and Wuhan Institute of Technology for geological models and suggestions. This work is supported by the National Key R&D Project of “Research and Development of Geological Resource Precision Development Risk Prediction Big Data Analysis Platform”, Fund No. 2022YFF0801203

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 27 Aug 2021 Revised: 19 Oct 2021 Accepted: 05 Nov 2021
Published online: 15 June 2022

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

References

1. BP. bp Energy Outlook 2024 edition [Internet]. 2024 [cited 2026 Jul 2]. Available from: https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/energy-outlook/bp-energy-outlook-2024.pdf

2. Department of Environment and Energy. Australian Energy Update [Internet]. 2024 [cited 2026 Jul 2]. Available from: https://www.energy.gov.au/sites/default/files/2024-08/australian_energy_update_2024.pdf

3. Goraya NS, Rajpoot N, Sivagnanam BM. Coal bed methane enhancement techniques: a review. ChemistrySelect. 2019;4(12):3585-601.

4. Sharma R, Singh S, Anand S, Kumar R. A review of coal bed methane production techniques and prospects in India. Mater Today Proc. 2024;99:177-84.

5. Sarhosis V, Jaya AA, Thomas HR. Economic modelling for coal bed methane production and electricity generation from deep virgin coal seams. Energy. 2016;107:580-94.

6. van Bergen F, Gale J, Damen KJ, Wildenborg AFB. Worldwide selection of early opportunities for CO2-enhanced oil recovery and CO2-enhanced coal bed methane production. Energy. 2004;29(9-10):1611-21.

7. Fedorov AV, Fedorchenko IA. Mathematical modeling of methane flow in coal beds. J Min Sci. 2009;45(1):9-21.

8. Kurlenya MV, Serdyukov SV. Methane desorption and migration in thermodynamic inequilibrium coal beds. J Min Sci. 2010;46(1):50-6.

9. Ruban AD, Zaburdyaev VS, Kharchenko AV. Coal bed methane drainage with long directional boreholes. J Min Sci. 2012;48(3):436-9.

10. Liu H, Sang S, Xue J, Lan T, Xu H, Ren B, et al. Experimental study on the velocity sensitivity of coal reservoir during coalbed methane drainage in southern Qinshui Basin. Energy Explor Exploit. 2018;36(6):1593-608.

11. Liu S, Wei J, Ma Y, Liu H, Liu X, Yan B. Numerical simulation investigations of coalbed methane drainage performance with multilateral well. J Pet Explor Prod Technol. 2021;11(3):1303-21.

12. Sang S, Xu H, Fang L, Li G, Huang H. Stress relief coalbed methane drainage by surface vertical wells in China. Int J Coal Geol. 2010;82(3-4):196-203.

13. Zhang T, Chen S, Tang D. Anisotropic evolution of effective stress and pore pressure during coalbed methane drainage. SPE J. 2023;28(5):2534-46.

14. Zou M, Wei C, Zhang M, Shen J, Shao C. A mathematical approach investigating the production of effective water during coalbed methane well drainage. Arab J Geosci. 2014;7(5):1683-92.

15. Zhao Z, Liu D, Chen M, Wang B, Sun J, Yu L, et al. Gas and water performance from the full-cycle of coalbed methane enrichment-drainage-output: a case study of Daning-jixian area in the eastern margin of Ordos Basin. Energy Rep. 2023;9:3235-47.

16. Clarkson CR, McGovern JM. Optimization of coalbed-methane-reservoir exploration and development strategies through integration of simulation and economics. SPE Reserv Eval Eng. 2005;8(6):502-19.

17. Du S, Wang M, Yang J, Zhao Y, Wang J, Yue M, et al. A novel prediction method for coalbed methane production capacity combined extreme gradient boosting with bayesian optimization. Comput Geosci. 2024;28(5):781-90.

18. Wang D, Li Z, Fu Y. Production forecast of deep-coalbed-methane wells based on long short-term memory and bayesian optimization. SPE J. 2024;29(7):3651-72.

19. Yin L, Yang W, Yongwei G, Shu Z. FOA application in the GRNN optimization for CBM productivity prediction. In: 6th International Conference on Artificial Intelligence and Big Data (ICAIBD 2023). IEEE; 2023. p. 644-9.

20. Yang R, Qin X, Liu W, Huang Z, Shi Y, Pang Z, et al. A physics-constrained data-driven workflow for predicting coalbed methane well production using artificial neural network. SPE J. 2022;27(3):1531-52.

21. Xianglan J, Jianwu L, Zhiyuan Y, Peihe Z. The optimization of the coalbed methane assessment indexes in the high abundance coalbed methane enrichment area in the case of the southern part of the Qinshui Basin, China. Procedia Earth Planet Sci. 2011;3:175-82.

22. Sun AY. Optimal carbon storage reservoir management through deep reinforcement learning. Appl Energy. 2020;278:115660.

23. Li W. Combination of numerical simulation and deep reinforcement learning for Petrole Reservoir management in Daqing Oilfield. In: Proceedings of the 2024 International Conference on Artificial Intelligence, Digital Transformation and Smart Systems (AIDTSS). ACM; 2024. p. 112-7.

24. Mo S, Zabaras N, Shi X, Wu J. Deep autoregressive neural networks for high-dimensional inverse problems in groundwater contaminant source identification. Water Resour Res. 2019;55(5):3856-81.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

25. Geneva N, Zabaras N. Modeling the dynamics of PDE systems with physics-constrained deep auto-regressive networks. J Comput Phys. 2020;403:109056.

26. Zhang L, Lu G, Wu C, Wu J, Zhao Y, Liu W. Numerical simulation of a coupled gas flow and geomechanics process in fractured coalbed methane reservoirs. Energy Sci Eng. 2019;7(4):1095-105.

27. Wang Y, Man Z. Numerical simulation of coalbed methane-water two-phase flow and prediction of coalbed methane productivity based on finite volume method. Int J Heat Technol. 2019;37(4):975-84.

28. Zhang T, Guo J, Zeng J, Zhang H, Zhao Z, Zeng F, et al. A transient source-function-based embedded discrete fracture model for simulation of multiscale-fractured reservoir: application in coalbed methane extraction. Gas Sci Eng. 2025;133:205500.

29. Jiang S, Durlofsky LJ. Use of multifidelity training data and transfer learning for efficient construction of subsurface flow surrogate models. J Comput Phys. 2023;474:111800.

30. Derry A, Krzywinski M, Altman N. Convolutional neural networks. Nat Methods. 2023;20(9):1269-70.

31. Zhao H, Alachiotis N. Data preprocessing methods for selective sweep detection using convolutional neural networks. Methods. 2025;233:19-29.

32. Yao Z, Hu Z, Lai P, Qin F, Wang W, Wu Z, et al. Etching process prediction based on cascade recurrent neural network. Eng Appl Artif Intell. 2025;139:109590.

33. Nikzad MH, Heidari-Rarani M, Rasti R. Long-short-term memory (LSTM)-based modeling of the stiffness of 3D-printed PLA parts. Mater Lett. 2025;379:137636.

34. Li W, Zhou Z, Tang J, Xu M. Spline Versoria prioritization optimization algorithm with modified adaptive moment estimation: analysis and comparison. IEEE Trans Circuits Syst II Express Br. 2024;71(1):485-9.

35. Yatawatta S. Reinforcement learning. Astron Comput. 2024;48:100833.

36. Lee HR, Lee T. Markov decision process model for patient admission decision at an emergency department under a surge demand. Flex Serv Manuf J. 2018;30(1-2):98-122.

37. Wang J, Zhu F, Wang Q, Cui Y, Sun H, Zhao P. An active object detection model with multi-step prediction based on deep q-learning network and innovative training algorithm. Appl Intell. 2025;55(2):1-21.

38. Sunitha D, Latha PH. A secure routing and black hole attack detection system using coot chimp optimization algorithm-based Deep Q Network in MANET. Comput Secur. 2025;148:104166.

39. Lu Q, Huang W, Yin L. Decomposition prediction fractional-order active disturbance rejection control deep Q network for generation control of integrated energy systems. Appl Energy. 2025;377:124773.

40. Rodriguez Sanchez, Salmeron Ainara S, Roman Gomez R, Garcia C. The coefficient of determination in the ridge regression. Commun Stat-Simul Comput. 2022;51(1):201-19.

41. de Freitas VR, Barroso Neto, Pereira E, Sanchez WHC, Fortaleza ELF, Schveitzer A Jr, et al. A sensitivity analysis of the impact of shut-in water cut of reactive control on net present value of waterflooding under reservoir uncertainty. Comput Geosci. 2024;28(6):1459-82.

42. Zanbouri H, Salahshoor K. Development of robust surrogate model for economic performance prediction of oil reservoir production under waterflooding process. J Pet Sci Eng. 2018;165:496-504.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

43. Sandler J, Fowler G, Cheng K, Kovscek AR. Solar-generated steam for oil recovery: reservoir simulation, economic analysis, and life cycle assessment. Energy Convers Manag. 2014;77:721-32.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press