



ARTICLE

An Integrated Multi-Scale Modeling Framework for Gas Entrainment Prediction in Coalbed Methane Production Systems

Qin Zhao¹, Yuxin Wang¹, Gang Chen¹, Hui Zhang¹, Lei Wang¹, Mulin Zhou¹, Songfei Zhang¹ and Yu Weng^{2,*}

¹Petrochina Coalbed Methane Co., Ltd., Beijing, China

²College of Pipeline Engineering, Xi'an Shiyou University, Xi'an, China

*Corresponding Author: Yu Weng. Email: wengyu009@hotmail.com

Received: 30 March 2026; Accepted: 08 June 2026; Published: 30 June 2026

ABSTRACT: This study presents an integrated multi-scale framework for predicting gas entrainment and flow behavior in coalbed methane production systems under gas-liquid two-phase flow conditions. The approach combines three-dimensional computational fluid dynamics simulations, reduced-order modeling, and machine-learning-based prediction to achieve both high physical fidelity and computational efficiency. Such an improved strategy stems from a specific need. As coalbed methane extraction increasingly encounters complex multiphase flow conditions, accurate characterization of gas entrainment has become essential for improving production stability and optimizing downstream gathering and separation systems. In practice, the flow entering rod pumps frequently deviates from the conventional assumption of a purely aqueous phase, leading to strongly coupled gas-liquid interactions that are difficult to capture using traditional modeling approaches. At the same time, the computational cost associated with full three-dimensional simulations limits their applicability to real-time field operations. The present approach is introduced to map the three-dimensional flow dynamics onto a one-dimensional numerical manifold while retaining the dominant physical characteristics of the system. Building upon this reduced representation, a Stacking ensemble learning framework is developed using Support Vector Regression, Back-Propagation Neural Networks, and Random Forests as base learners, combined through an Adaptive Neuro-Fuzzy Inference System meta-learner for rapid field-scale prediction. The results show that the suction effect generated by the rod pump is the dominant mechanism governing gas entrainment. Notably, the proposed prediction model achieves high predictive accuracy, with a root mean square error of 1.75 L/min, a mean absolute error of 0.82 L/min, and a coefficient of determination of 0.98. Furthermore, analysis of the surface pipeline network indicates that terminal flow rates are characterized by low-frequency pulsations that promote stable gas-liquid interface formation within downstream separators.

KEYWORDS: Coalbed methane; gas containing produced water; reduced order analysis; large model; transport and separation

1 Introduction

As a clean and efficient unconventional energy resource, coalbed methane (CBM) is typically extracted via dewatering and pressure drawdown, a process designed to facilitate the desorption and subsequent recovery of methane from coal seams. Under ideal operating conditions, the pumping system is intended to lift only CBM produced water, while natural gas is recovered separately through the wellbore annulus. However, field monitoring across major global CBM basins—including the San Juan Basin (USA), the Bowen Basin (Australia), and the Qinshui Basin (China)—reveals that gas-liquid two-phase flow is prevalent within

the production stream. This indicates significant gas entrainment in the produced water, deviating from the intended single-phase extraction [1–3].

The natural gas produced along with water mainly exists in two forms: first, dissolved methane, which is affected by reservoir temperature, pressure, and water chemical properties [4]. Second, dispersed bubbles, which are difficult to separate from water due to the negative pressure environment at the pump suction inlet [5]. Current investigations have shown that the natural gas entrainment in produced water from a single well can reach 0.5–3.2 m³/d. Based on the estimation of the global annual CBM production of 70 billion m³, the annual loss of natural gas resources along with produced water exceeds 3 billion m³ [6,7]. Moreover, most unrecovered natural gas is released directly into the atmosphere during produced water treatment. Because methane's 100-year GWP is 28 times higher than that of CO₂, its emission is a major source of greenhouse gas pollution [8,9]. A study by Omara et al. of over 1000 sites found that episodic events, particularly wellhead fluid discharge, are the leading cause of emissions in gas fields [10]. Similarly, studies at the Shengli Oilfield show associated gas emissions from produced water reaching 0.415 g/(m³·h) [11].

The current research on tools used for water drainage during coalbed methane (CBM) extraction emphasizes the development and application of various pump technologies, including rod pumps and rod-less alternatives. Traditional rod pumps remain a significant component in CBM drainage, with studies highlighting their operational mechanisms and technological advancements [12]. These pumps are widely used due to their reliability and effectiveness in lifting water and gas from coal seams, although challenges such as equipment durability and operational efficiency persist.

To elucidate the mechanisms governing gas ingress into produced water pipelines, we delineate the fluid transport trajectory from the coal seam to the wellbore (Fig. 1). Initially, the millennial-scale coexistence of water and methane under reservoir pressure establishes a state of saturated solubility. As extraction-induced pressure drawdown occurs, this dissolved gas exsolves into a free phase. During the recovery process, fluctuations in the wellbore liquid level or the positioning of production intervals below the pump intake—typically a reciprocating rod pump—can trigger the mechanical suction of gas. Furthermore, the structural integrity of wellbore tubulars is frequently compromised by recurrent mechanical wear and corrosion; such fractures not only facilitate internal fluid leakage but also permit the infiltration of extraneous gas from the surrounding formation. Consequently, characterizing the gas content in produced water presents a multi-faceted challenge. Current research is largely confined to regional assessments based on atmospheric gas escape, leaving a critical void in well-specific, high-resolution diagnostics. While high-fidelity Computational Fluid Dynamics (CFD) offers sophisticated insights into gas-liquid multiphase flow, the inherent complexity of 3D numerical simulations—compounded by dynamic boundary conditions and phase transitions—results in prohibitive computational latencies, often spanning several weeks for a single solution [13–15]. This bottleneck precludes real-time prognostic assessment of volatile downhole conditions and hinders the fundamental investigation of gas-bearing mechanisms. To bridge this gap, reduced-order modeling (ROM) and data-driven reconstruction techniques have emerged as transformative paradigms, offering a robust balance between physical rigor and computational efficiency in resolving complex multiphase dynamics [16,17].

In addition, for the recovery of natural gas from produced water, it is difficult to carry out separation and recovery directly at the wellhead due to the high cost. Actually, gas-liquid mixtures are usually transported from individual wells to processing stations through pipelines for centralized separation. During pipeline transportation, the flow regime of the gas-liquid mixture changes with the ups and downs of the pipeline route, resulting in unstable transportation flow rates, and fluctuations in the gas-liquid flow rates entering

the centralized separators, even slug flow. These phenomena will have a significant impact on the design of the gas-liquid separator [18].

Dottore et al. [19,20] elucidated the pump performance dynamics in gas-bearing wells by formulating analytical expressions for pump chamber pressure fluctuations, volumetric efficiency, and the critical thresholds for gas lock initiation. From a mechanistic perspective, Das et al. [21,22] investigated gas-liquid flow transitions within vertical annular pipes, identifying gas velocity fluctuations at constant pressure as the primary driver of regime shifts. Furthering this, Gomez et al. [23] developed a unified steady-state mechanistic model to predict flow regimes, liquid holdup, and pressure gradients. Despite these advancements, a significant knowledge gap persists: existing frameworks seldom integrate the complex interplay between two-phase pipeline pulsations and the transient flow phenomena within gas-liquid separators.

In recent years, with the development of techniques such as order reduction and machine learning, scholars have begun to conduct large-scale model predictions based on well measured data and mathematical models for pipe pumps. Jamalbayov et al. [24] adopted a discrete modeling approach to adjust and decouple complex control equation systems, proposing a comprehensive model for the development of rod pump wells and reservoirs. Zuo et al. [25] based on a large amount of measured data on the electric power of rod pumps, used the MSISSA-BiLSTM learning model to estimate effective liquid production, forming an efficient prediction method. Therefore, techniques such as order reduction and machine learning have strong application potential in this type of problem.

A certain coalbed methane field mainly uses rod pumps to extract produced water. In recent years, the gas field has been plagued by the problem of gas content in produced water. To address the gas-entrainment challenges in CBM produced water, this study develops a multiscale integrated framework combining High-Fidelity CFD, Reduced-Order Modeling (ROM), and Ensemble Learning. We first conduct 3D two-phase flow simulations to characterize the transport mechanisms within single-well extraction systems. To overcome computational bottlenecks, these 3D physics are distilled into an efficient 1D numerical model via ROM, significantly accelerating individual analysis cycles. Leveraging this efficiency, a Stacking-based ensemble model is constructed to predict well-scale gas-liquid flow rates and gas content across the entire extraction block. Furthermore, we establish a network-scale two-phase flow model to analyze the impact of topographical undulations on transport characteristics. Finally, the separation performance of tubular gas-liquid separators is evaluated under realistic pulsating inflow conditions using 3D numerical analysis, providing a holistic understanding of the gas trajectory from the reservoir to surface processing.

2 Research Methodology

2.1 Numerical Analysis Method

Rod pumps are commonly used for extracting produced water in the studied CBM extraction area. In the analysis of downhole produced water flow, it is necessary to consider the displacement and mechanical properties of the pump rod, piston, and valve ball, while investigating the impact of the suction effect generated by the operation of the pump and pipeline. In three-dimensional flow field CFD, the governing equations mainly include the mass, the momentum and the energy conservation equation:

$$\text{Mass conservation equation: } \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{V}) = S \quad (1)$$

$$\text{Momentum conservation equation: } \frac{\partial(\rho \cdot \mathbf{V})}{\partial t} + \nabla \cdot (\rho \mathbf{V} \mathbf{V}) = -\nabla P + \nabla(\mu(\nabla \mathbf{V} + \nabla \mathbf{V}^T)) + \rho \mathbf{g} + \mathbf{T}_\sigma \quad (2)$$

$$\text{Energy conservation equation: } \frac{\partial(\rho T)}{\partial t} + \nabla \cdot (\rho \mathbf{V} T) = \nabla \cdot (\lambda \nabla T) \quad (3)$$

where: ρ is density, t is time, ∇ is the differential Hamiltonian operator, \mathbf{V} is the velocity vector, P is pressure, S is the mass source term, μ is viscosity, T is temperature, g is gravitational acceleration, and T_σ is the surface tension at the phase interface, λ is the equivalent thermal conductivity of the fluid medium. All physical quantities are mixed-phase parameters. The compressibility of the gas is using the ideal gas state equation.

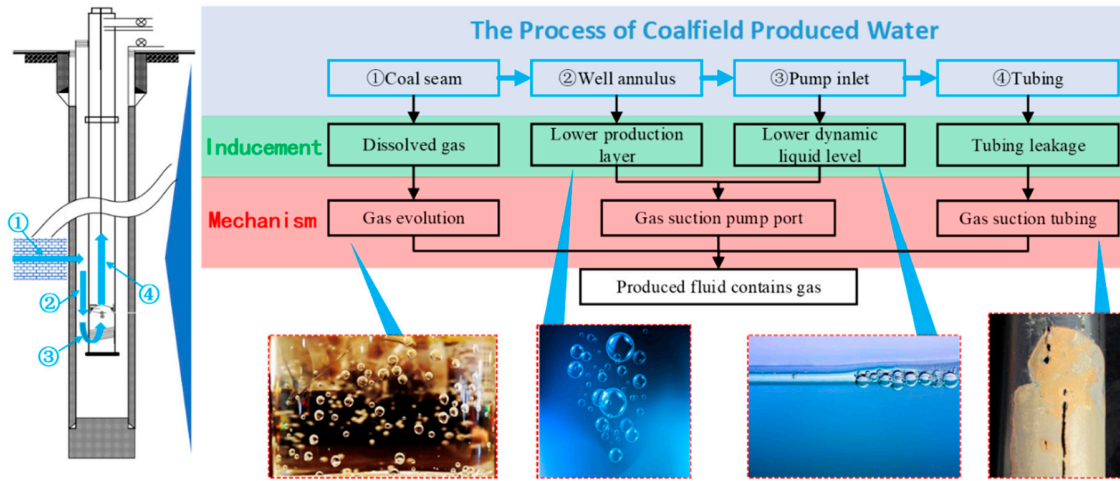


Figure 1: Schematic diagram of gas source in produced water.

Since the natural gas in the studied gas field has an extremely high methane content, the solubility of natural gas in produced water can be calculated using the following formula. Furthermore, the amount of dissolved/precipitated gas caused by pressure changes can be derived in the numerical model:

$$x = \frac{f_{\text{CH}_4}^g}{\gamma_x \cdot H_x^0} \quad (4)$$

where: $f_{\text{CH}_4}^g$ is the fugacity of methane in the gas phase, calculated using the gas state equation. γ_x is the activity coefficient of methane in water. and H_x^0 is the reference-state Henry's law constant.

The produced water in the studied gas field exhibits a salinity of 1760 mg/L, a measured viscosity of 1.02 mPa·s, and a surface tension of 76 mN/m.

Since the flow channel changes in CFD calculations are mainly dominated by the overall lifting and lowering of the plunger and valve ball. Therefore, it is assumed that only the fluid-structure interaction interfaces of the plunger and valve ball deformation during calculations, and the deformation mode is very similar to one-dimensional elastic deformation. Thus, based on Hooke's law, grid layers near the wall are defined as the deformation zone, and the force balance relationship of each grid node in the deformation zone can be expressed by the following formula:

$$F_i = \sum_j^{n_i} K_{ij} (\Delta X_j - \Delta X_i) \quad (5)$$

where: n_i is the number of adjacent nodes of node i . $K_{ij} = \frac{1}{\sqrt{|X_i - X_j|}}$ is the spring stiffness coefficient between nodes i and j . ΔX_i and ΔX_j represent the displacements of grid nodes i and j relative to their initial positions.

Once the displacements of nodes on the fluid-structure interaction interface are obtained, the displacements of nodes in the flow channel are determined by the following formula:

$$X_i^{n'+1} = X_i^{n'} + \gamma \Delta X_i^m \quad (6)$$

where: n' is the time step. $X_i^{n'}$ and $X_i^{n'+1}$ are the positions of node i at the current time step and the next time step. γ is the relaxation factor, which is used to characterize the of the displacement of the node. ΔX_i^m is the displacement of node i at the current time step. and m is the number of iterations.

The above expressions for the natural gas dissolution/precipitation process and fluid-structure interaction process are implemented in the computational model through C language subroutines in CFD calculations, respectively applied to the mass conservation equation and the momentum conservation equation.

The downhole numerical analysis domain includes the area enclosed by the wellbore, plug, pump, and tubing. A three-dimensional geometric model of the pipeline and rod pump was established using 3D CAD methods, and the 3D analysis model generated by the filling method is shown in Fig. 2a. The analysis model of the gas-liquid separator is shown in Fig. 2b.

The flow field analysis model was generated using filling technology to ensure consistency with the solid model at the interface. Structured hexahedral (or quadrilateral) topological structures were used for meshing. The grid-independence of the numerical solution for the temperature was examined to ensure the grid independent solution. By varying the grid scale, 5 grid systems were generated. For the under well model, taking the conditions of pump setting depth of 817 m, pump length of 8.3 m, and pump submergence of 1 m. For the gas-liquid separator model, taking the conditions of an inlet liquid flow rate of 34.25 m³/h, and a gas flow rate of 27.29 m³/h. The results of the last two sets of grids are fairly close to each other (shown in Fig. 3), and the error was less than 1%. This indicates that the resolution of the finer grid was adequate to guarantee computational precision. The final grid number used for under well simulations was 5.54 million, and for separator simulations was 6.27 million.

The detailed model selection is shown in Table 1 [26–29]. In the selection of mathematical models, reference was made to existing research on gas-liquid two-phase flow in vertical pipelines [30]. To ensure sufficient computational accuracy, the selection of the difference scheme was determined by referring to the CFD calculation recommendations of AIAA [31]. In the context of vertical pipeline flow, the VOF model has been extensively employed to simulate and analyze gas-liquid interactions [32]. Therefore, the VOF model was employed in the calculation to simulate two-phase flow.

All boundary conditions for numerical calculations are derived from actual under well sensor measurement data. In the CFD calculations, no less than 3000 time iterations were done to reach a fully converged solution: Stability of all monitoring parameters (flow rate, force), mass balance convergence and maximal residuals about 10⁻⁵.

2.2 Reduced Order Analysis Method

Reduced-order model (ROM) analysis utilizes input-output data from a numerical analysis system to determine a simple model system through a type of complex model system, making the two equivalent in terms of solution capability [33,34]. In this study, the complex model is a downhole numerical analysis based on 3D transient numerical calculations.

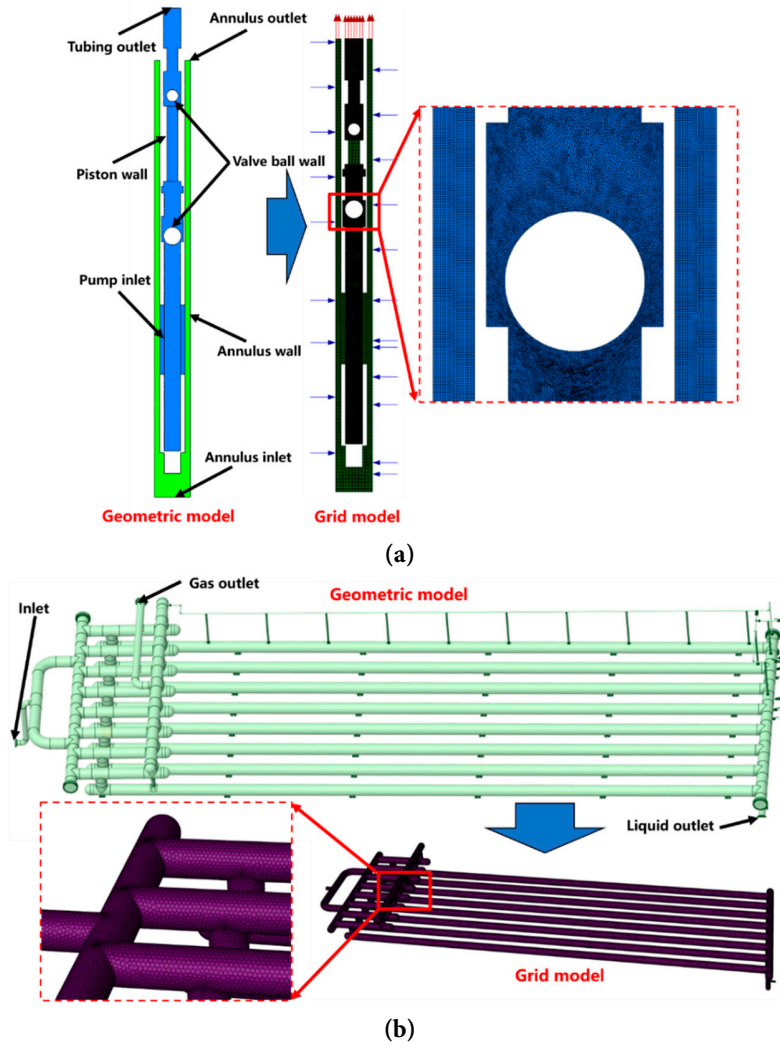


Figure 2: 3D numerical analysis model, (a) Under well, (b) Gas-liquid separator.

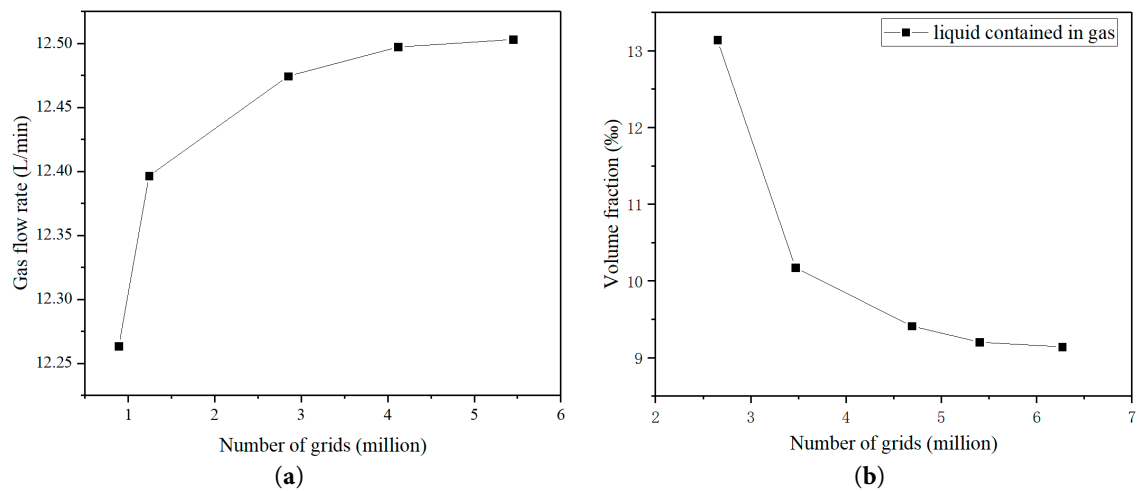


Figure 3: Grid independence assessment, (a) Under well, (b) Gas-liquid separator.

Table 1: Model setting of calculation.

Analysis Conditions	Model Type	Model Settings
Numerical methods	Difference scheme for control equations	Second order upwind
	Turbulence model	SST $k-\omega$
	Multi-phase flow model	Volume of Fluid
	Numerical algorithm	SIMPLE
Boundary conditions	Annulus inlet, outlet	Pressure boundary with given casing pressure and dynamic liquid level height.
	Tubing outlet	Pressure boundary with given tubing back pressure.
	Pump inlet	Porous step-over boundary with given porosity and flow resistance coefficient.
	Piston wall	Forced displacement wall with given axial sinusoidal forced displacement.
	Valve ball wall	Free displacement wall with given valve ball mass and moment of inertia.
	Annulus wall	Fixed wall, hydraulically smooth.

In downhole numerical analysis, the displacement of the valve ball under solid constraints and two-phase countercurrent impact is the input of the system, and the gas-liquid flow rate at the pump outlet is the output of the system. To establish a one-dimensional transient numerical calculation model using this method, it is first necessary to obtain the hydraulic coupling dynamic characteristics of the valve ball. The purpose of model order reduction is to obtain this mechanical characteristic model. The system identification process involves first determining the design of the identification scheme and model type, then determining the parameters of the system based on characteristics of the system, and finally obtaining the model parameters, as shown in Fig. 4.

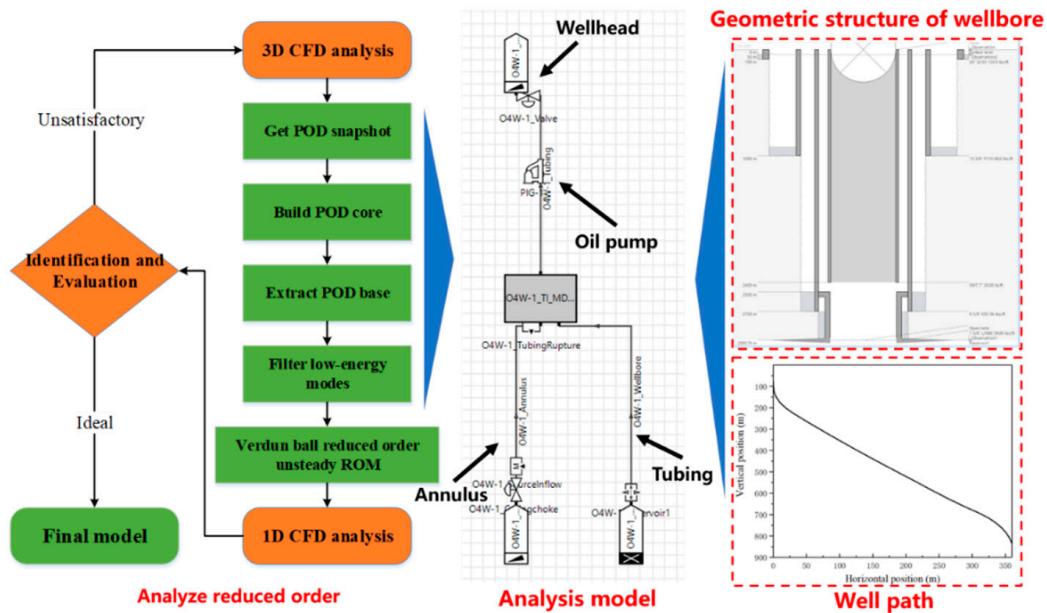


Figure 4: Reduced order analysis process diagram.

Parameter estimation of the model is a key step in the system identification process, which involves using iterative estimation methods to estimate model parameters to minimize the likelihood function. Based on the transient gas-liquid two-phase flow dynamic reduced-order model derived from system identification, the hydraulic coupling state equation of the valve ball can be expressed by the following formula. For a general fluid-structure interaction elastic system:

$$\begin{cases} \dot{x}_a(t) = A_a x_a(t) + B_b \zeta(t) \\ f_a(t) = C_c x_a(t) + D_d \zeta(t) \end{cases} \quad (7)$$

where: x_a is the state vector. f_a is the two-phase flow dynamic coefficient. $\zeta = u$ is the modal displacement. and A_a, B_b, C_c, D_d is the coefficient matrix to be identified.

In the solution process of 3D transient numerical calculations, the displacement of the valve ball can be taken as the input signal, and the corresponding two-phase flow force is obtained as the output signal. Then, system identification technology is used to perform parameter identification on the system. Therefore, for a given calculation condition, only one input excitation signal needs to be calculated to obtain the reduced-order identification model [35].

The study adopts the Proper Orthogonal Decomposition (POD) based on multivariate statistics for system parameter identification. Its basic process is to find a new set of bases in a large multivariate system, such that the dimension of the new system spanned by these new bases is significantly reduced, and the error between the new system and the original system is minimized. Specifically, if the solution obtained by solving the required partial differential equations is $Y = \{y_1, \dots, y_p\}$, where y_i are column vectors, this solution matrix can be used as snapshots. A set of eigen orthogonal bases $\Psi = \{\varphi_1, \dots, \varphi_p\}$ is derived from these snapshots, and the original system is projected onto this set of orthogonal bases, with the error between the projected system and the original system minimized.

The computational load of the hydraulic coupling response using the reduced-order method is far less than that of the direct 3D transient numerical calculation method. After multiple identifications, the identified hydraulic coupling dynamic model of the valve ball can be used to calculate the gas-liquid two-phase flow output rate of the pump.

3 Flow Field Characteristics in Wells

3.1 Flow Field Analysis

During the extraction process, the pump inlet may contain a certain proportion of gas because it is close to the coal seam or the liquid level outside the pump inlet is low, leading to gas being carried into the pump. This is the main cause of gas content in produced water. Taking the typical condition where gas overflows from the underlying reservoir at the pump inlet, a flow field analysis was conducted. The flow field distribution in the wellbore annulus below the pump inlet is shown in Fig. 5. During the upstroke of the pump, due to the pump's suction effect, the gas entering the annulus through the perforations forms a regular gas column distribution and ultimately flows to the pump's suction inlet. In the downstroke, however, the driving force for forced flow in the annulus stops, and the gas then presents an irregular random distribution. Nevertheless, it still moves upward overall under the drive of buoyancy. The flow field near the pump is shown in Fig. 6. Due to the suction effect of negative pressure, gas is entrained into the pump, causing gas carry-over issues.

The time-history curves of gas and liquid discharge flow rates after the pump stabilizes over multiple working cycles are extracted, as shown in Fig. 7a. Due to the gas content at the suction inlet, the pump cham-

ber cannot continuously and stably inhale liquid, resulting in a significant reduction in liquid volume. From the transient gas-liquid flow rate curves, although the gas volume under standard conditions is much higher than the liquid volume, the liquid volume fraction in the fluid still dominates under downhole pressure.

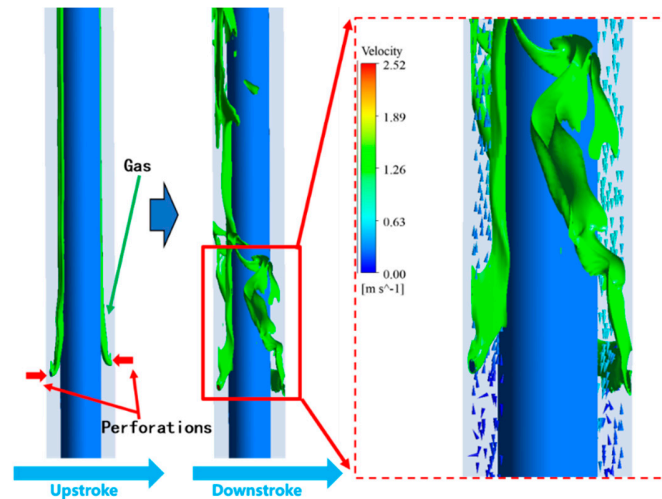


Figure 5: Distribution of 3D flow field in the annulus.

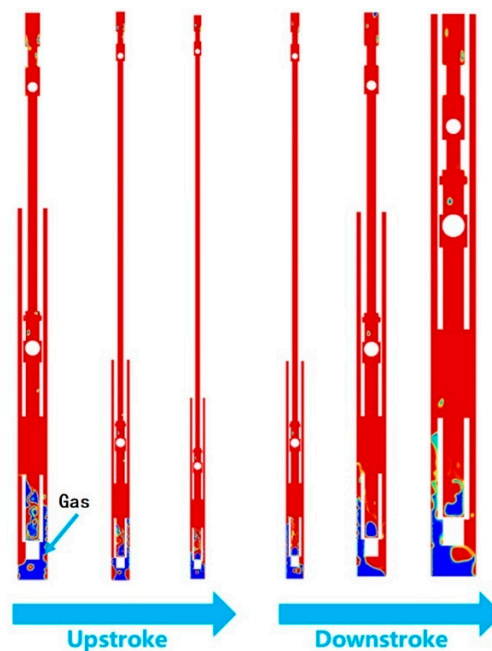


Figure 6: Phase distribution under the condition of gas at the pump suction inlet.

As an important input for reduced-order analysis, the displacement time-history curve of the valve ball is shown in Fig. 7b. During the upstroke, as the piston moves upward, the traveling valve ball closes under the fluid pressure in the annular space, blocking the backflow of fluid in the pump barrel. Meanwhile, the negative pressure inside the pump barrel pushes opens the fixed valve ball, allowing fluid to enter the pump barrel. During the downstroke, the piston moves downward to squeeze the fluid in the pump barrel, and the increased pressure pushes open the traveling valve ball. The fluid then enters the inner channel. The pressure inside the pump barrel is higher than that in the annular space, so the fixed valve ball closes under

the hydraulic pressure to prevent fluid backflow to the casing-tubing annulus. After each valve ball reaches the maximum or minimum displacement point, its displacement should theoretically remain unchanged for a short period. However, the effect of two-phase flow on the ball is far less stable than that of single-phase incompressible liquid, and slight oscillations of the ball are clearly observed during this process.

Using the reduced-order analysis method, after multiple iterations to obtain the hydraulic coupling dynamic model of the valve ball and applying it to the 1D model, the calculation results are very close to those of the 3D calculation. The total calculation time is reduced from nearly 200 h to 27 min.

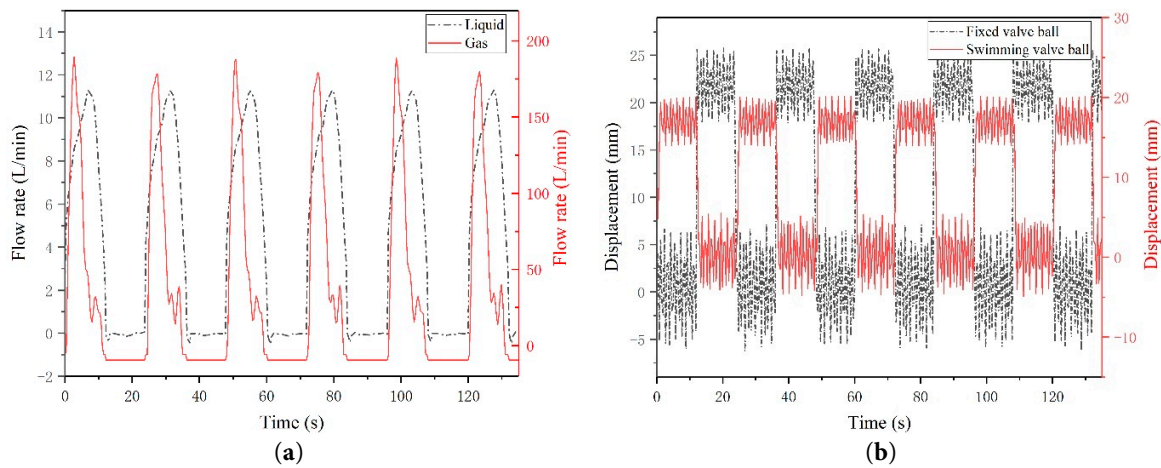


Figure 7: Time history distribution of flow rate and valve ball displacement, (a) Gas liquid flow rate, (b) Valve ball displacement.

To evaluate the accuracy of the reduced-order model, comparative calculations were performed under nine different well conditions, and the proposed 1D reduced-order model was compared with the 3D model. The comparison results for gas and liquid flow rates are shown in Fig. 8, indicating that the two models yield very close results. The root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) were selected as the prediction error evaluation metrics. For the 1D reduced-order model, the RMSE, MAE, and R^2 values for gas flow rate prediction are 5.55 L/min, 1.77 L/min, and 0.96, and for liquid flow rate prediction are 0.18 L/min, 0.21 L/min, and 0.95, respectively.

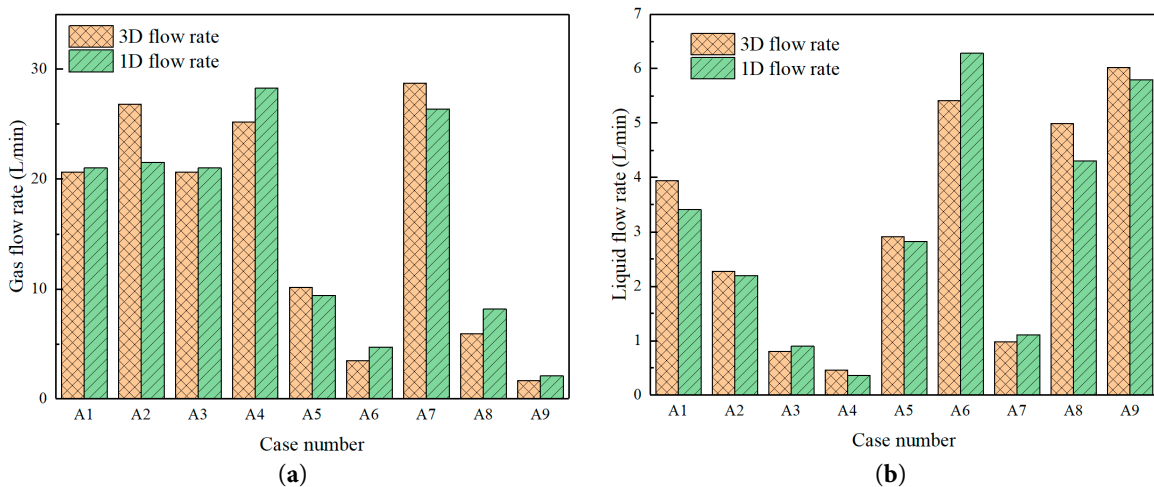


Figure 8: Comparison between 3D and 1D well flow rate calculation, (a) Gas flow rate, (b) Liquid flow rate.

3.2 Predictive Large Model

To ensure model parsimony while bolstering predictive performance, we developed an ensemble framework that integrates heterogeneous base learners to capture diverse data features. Specifically, Support Vector Regression (SVR) [36], Back-Propagation Neural Networks (BPNN) [37], and Random Forests (RF) [38]—representing kernel-based, neural-based, and tree-based architectures, respectively—were selected as base learners. These were coupled with an Adaptive Neuro-Fuzzy Inference System (ANFIS) as the meta-learner [39]. This stacking ensemble strategy [40] facilitates the construction of a robust proxy model for gas volume prediction, the architecture of which is illustrated in Fig. 9. The implementation follows a tripartite process:

(1) Dataset Generation Stage

The sample dataset (X, Y) is divided into a training dataset (X_{tr}, Y_{tr}) containing numerical analysis data and a test dataset (X_{te}, Y_{te}) containing on-site measured data.

$$X = \{x_i, i = 1, 2, 3 \dots, n\} \quad (8)$$

where: x_i represents the input parameters of the i -th dataset sample, including 12 parameters: casing diameter, casing pressure, pump setting depth, dynamic liquid level depth, pump length, pump diameter, stroke length, stroke frequency, screen length, screen outer diameter, total gas production of the underlying reservoir, average depth of the underlying reservoir.

$$Y = \{y_i, i = 1, 2, 3 \dots, n\} \quad (9)$$

where: y_i represents the response value corresponding to the j -th dataset sample, including the gas flow rate and liquid flow rate. $n - k$ samples are selected as the training dataset:

$$(X_{tr}, Y_{tr}) = \{(x_i, y_i), i = 1, 2, 3 \dots, n - k\} \quad (10)$$

According to the on-site conditions of the field, $n - k = 317$ samples are selected as the training dataset, $k = 120$ samples are selected as the test dataset:

$$(X_{te}, Y_{te}) = \{(x_i, y_i), i = 1, 2, 3 \dots, k\} \quad (11)$$

(2) Training Stage

Five-fold cross-validation is adopted to train each base learner (SVR, BPNN, RF). For each base learner, the original training set (X_{tr}, Y_{tr}) is first randomly and equally divided into five subsets (Training Sets 1–5). In each iteration, four subsets are used as the training set to train the base learner, and the remaining one is used as the validation set to evaluate the model's prediction performance and obtain a prediction result P_{tr-j} . This imposes an implicit regularization on the base learners, which greatly reduces the risk of overfitting. This process is repeated five times to obtain the prediction results of the base learner for the entire original training set. For example, for the SVR learner:

$$P_{tr-SVR} = \{P_{tr-j}, j = 1, 2, 3, 4, 5\} \quad (12)$$

The above training process is performed for each base learner to obtain a new training set (P_{tr}, Y_{tr}) constructed by the prediction results of each base learner and the corresponding original response values. This new training set is used as the training samples for the ANFIS meta-learner, where:

$$P_{tr} = \{P_{tr-SVR}, P_{tr-BPANN}, P_{tr-RF}\} \quad (13)$$

(3) Testing Stage

The trained base learners are used to obtain the predicted values P_{te} corresponding to the original test set. A new test set (P_{te}, Y_{te}) is constructed by these predicted values and the response values in the original test set, which is used to test the trained meta-learner, where:

$$P_{te} = \{P_{te-SVR}, P_{te-BPANN}, P_{te-RF}\} \quad (14)$$

Thus, the final prediction results $F_{te} = \{f_i, i = 1, 2, 3, \dots, k\}$ are obtained. Based on the above steps, a Stacking ensemble learning surrogate model for gas content prediction can be constructed. To improve the prediction accuracy of the surrogate model, the improved sparrow search algorithm (ISSA) is further adopted to optimize the parameters of each learner in the model.

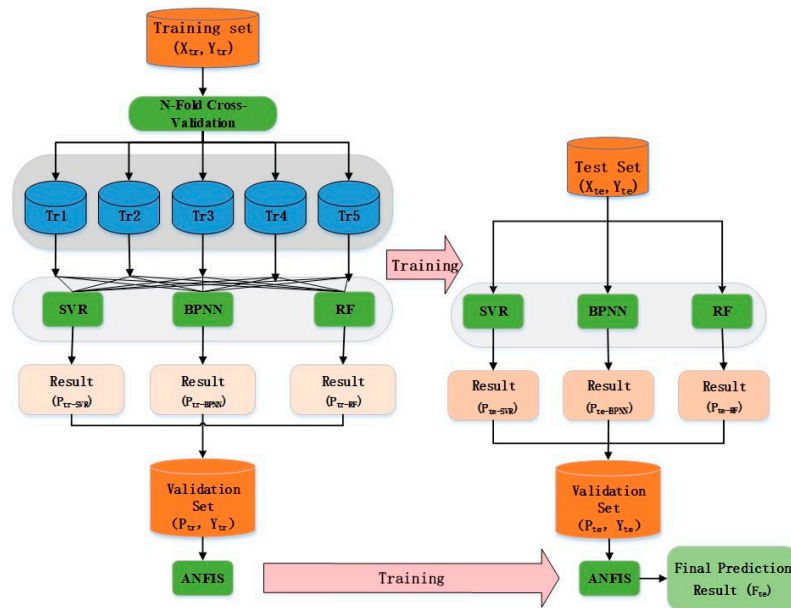


Figure 9: Integrated learning framework.

Using the reduced order model described in the previous section, the learning samples can first be generated. A total of 317 sets of different well operating conditions are produced as training datasets. Each set of data contains 12 input parameters (casing diameter, casing pressure, pump setting depth, dynamic liquid level depth, pump length, pump diameter, stroke length, stroke frequency, screen length, screen outer diameter, total gas production of the underlying reservoir, average depth of the underlying reservoir) and 2 output parameters (gas flow rate, liquid flow rate).

On this basis, operational testing was conducted on 72 wells in the field to obtain gas and liquid flow rates at the wellhead as validation datasets. To demonstrate the accuracy of the developed prediction model, the proposed Stacking learning model was compared with individual base learners (SVR, BPNN, and RF)

based on the gas flow rate obtained from validation datasets. The root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) were selected as the prediction error evaluation metrics. The comparison results of the four models are shown in Fig. 10.

All four models are capable of predicting gas flow rate with reasonable accuracy. Compared with the other three learners, the Stacking learning model achieves the highest prediction accuracy, with RMSE, MAE, and R^2 values of 1.75 L/min, 0.82 L/min, and 0.98. This indicates that the model can complement the strengths of each base learner during the data learning process, demonstrating excellent predictive capability, and performing more robustly and accurately.

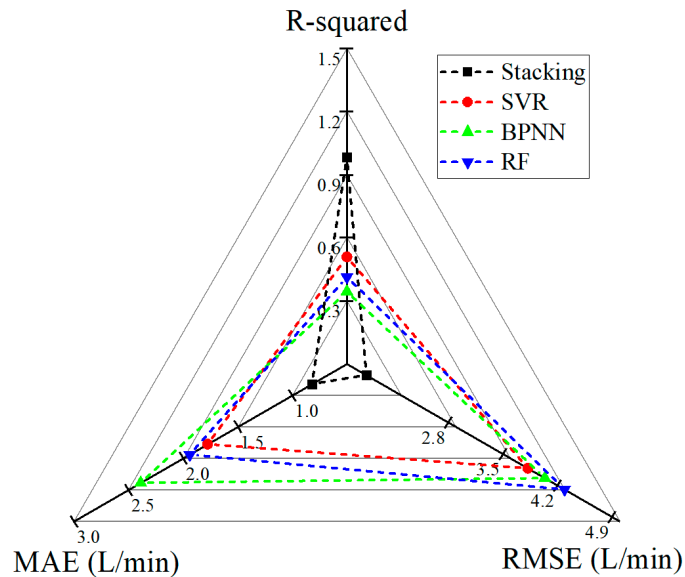


Figure 10: Comparison chart of prediction error.

4 Flow Field Characteristics of Transportation Separation System

4.1 Produced Water Pipe Network

The coalbed methane field under study is in a hilly terrain, with produced water pipelines laid along the surface showing obvious undulations. The entire field includes 101 wells distributed over an area of approximately 50 km². A 1D pipeline network model was established as shown in Fig. 11, which marks the average annual water production of each production point. The pipe material is PE, and the calculation utilizes the average scaling thickness of 3.63 mm and surface roughness of 0.27 mm obtained from on-site investigations. The pipeline network is distributed in a dendritic pattern, converging at the central treatment station. The prediction model obtained in Section 3.2 was used to calculate the water and gas flow rate of each well in a typical year, with the calculation conditions selected as the maximum daily value, minimum daily value, and annual average value of the total regional water production in a year. The two-fluid model is employed to simulate 2 phase flow in pipelines, and the flow pattern state is determined using a region identification method based on flow pattern maps [41].

The pressure at the inlet of each production point's pipeline is marked in Fig. 9. Since the outlet pressure of the pipeline network at the treatment station is close to atmospheric pressure, the pressure at the inlet of each production point's pipeline represents the pressure drop along each pipeline route. The gas holdup entering the treatment station is shown in Fig. 12. From Fig. 12a, it can be found that the gas holdup fluctuates significantly, and there are moments when it is entirely gas or entirely liquid. In

addition, the influence of the maximum or minimum water volume on the gas holdup is not obvious from a time-history perspective.

Fast Fourier Transform (FFT) transformation was performed on the gas holdup time-history data to obtain the pulsation distribution of gas holdup at different frequencies, as shown in Fig. 12b. At low frequencies, the pulsation energy under different water production conditions shows one or more peaks, while the pulsation energy decreases rapidly with the increase in frequency. The peak of the maximum pulsation frequency ranges from 0.008 to 0.007 Hz, corresponding to a pulsation period of 120–150 s. This indicates that the gas-liquid pulsation at the outlet of the pipeline network is dominated by low-frequency pulsation, which will have an adverse impact on the operation of the gas-water separator planned to be installed at the treatment station.

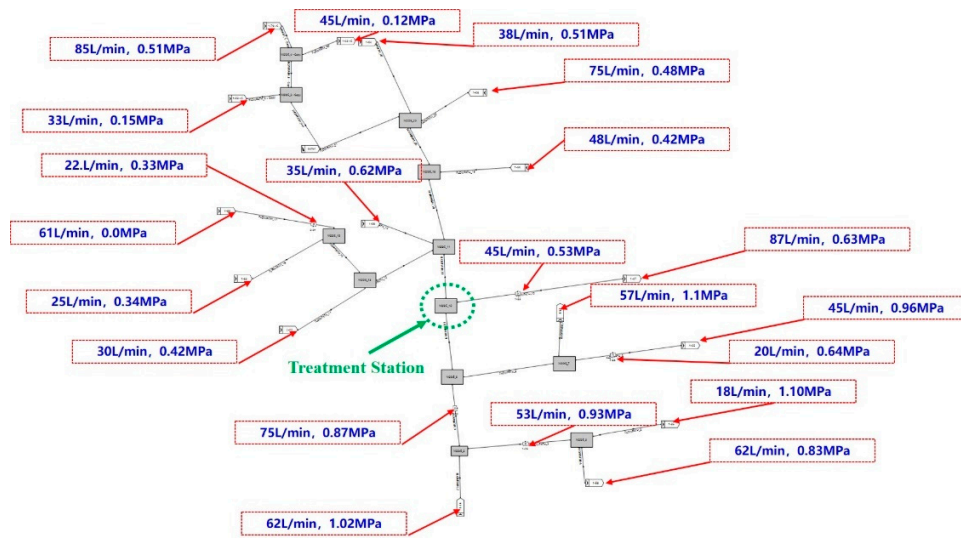


Figure 11: Numerical analysis model for extracted water pipe network.

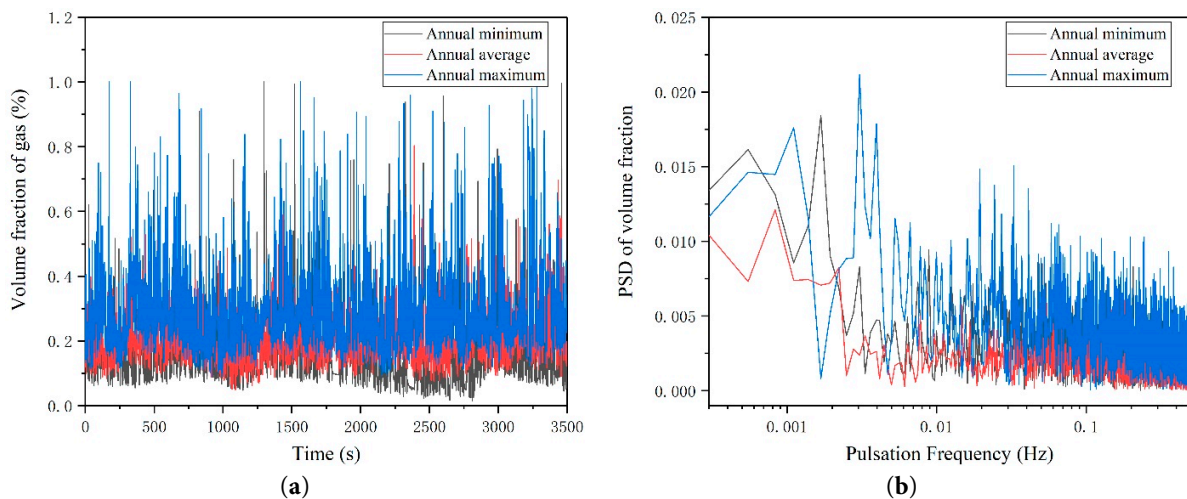


Figure 12: Distribution of gas volume fraction at the end of the pipeline network, (a) Time History, (b) Frequency history.

4.2 Gas-Liquid Separator

This study adopts a tubular separator as the separation equipment for the centralized treatment of produced water. To verify whether the designed separator can operate properly—especially to well accommodate the incoming medium under significant gas-liquid flow rate pulsation. The annual maximum water production condition obtained in Section 4.1 was chosen as the calculation condition for the separator. The time-history gas and liquid flow rates calculated from the pipeline network were used as the inlet boundary conditions for the separator, and the internal flow field distribution under stable working conditions is shown in Fig. 13.

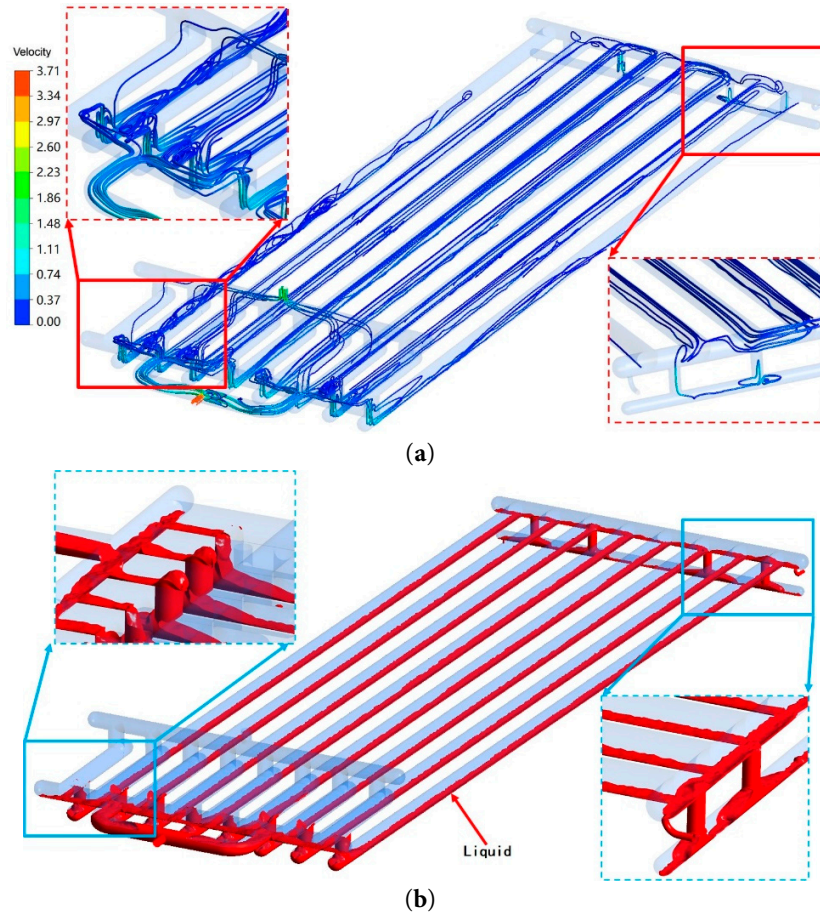


Figure 13: Flow field distribution of gas-liquid separator, (a) Streamline diagram, (b) Liquid phase distribution.

After entering through the separator inlet, the gas-liquid mixture is directed to 8 parallel branch pipes. As the flow channel space expands in the branch pipes, the flow velocity decreases rapidly. Reducing the flow velocity can effectively reduce the inertial force of the fluid, allowing the gas to be gradually dominated by buoyancy. Since the gas phase outlet is located at the highest point of the entire equipment, the gas can flow out smoothly from this outlet. Meanwhile, the liquid continues to flow downstream through branch pipes at the bottom. During operation of the separator, stable stratified flow can be maintained in the branch pipes, ensuring that the downstream liquid phase outlet is always below the liquid level and preventing gas from flowing in. From the perspective of flow field distribution, the separator can maintain a clear gas-liquid interface even under the conditions of maximum flow rate and strongest gas-liquid pulsation, and the interface level does not exceed the mid-plane of the lower branch pipes.

The liquid and gas holdup at the outlet of the separator are extracted as shown in Fig. 14. The liquid holdup at the gas outlet is slightly higher than the gas holdup at the liquid outlet, indicating that a small amount of liquid still fails to fully settle during the flow process. However, both the liquid content at the gas outlet and the gas content at the liquid outlet are maintained below 12‰, the capture efficiency of liquid droplets with a particle size larger than 150 μm can reach 99%, demonstrating that the separator has a good separation effect without obvious gas/liquid carry-over.

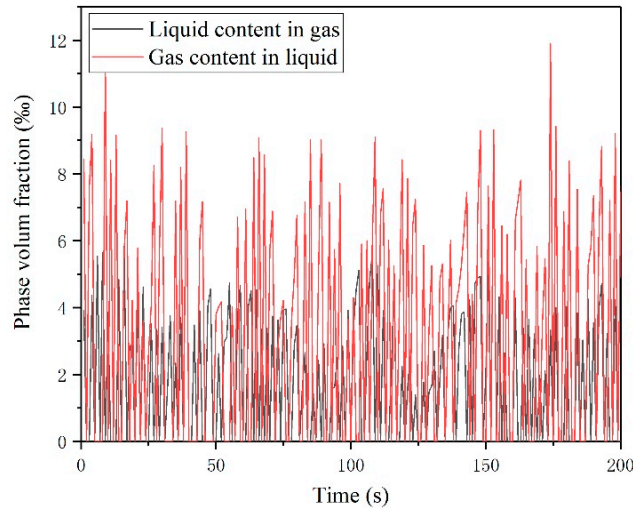


Figure 14: Distribution of phase content at the outlet of gas-liquid separator.

5 Conclusion

This study elucidates the complex dynamics governing gas entrainment in coalbed methane (CBM) produced water through a multi-scale framework that integrates high-fidelity 3D-to-1D reduced-order modeling (ROM) and advanced ensemble learning. By projecting complex 3D fluid dynamics onto a computationally efficient 1D manifold, we have developed a framework capable of rapid, cross-well characterization, providing a definitive depiction of gas entrainment processes that were previously computationally prohibitive to resolve.

Mechanistically, our analysis identifies the ‘suction effect’ of the rod pump as the primary driver for gas ingestion; specifically, intense localized flow gradients formed during the upstroke facilitate the entrainment of gaseous phases from the suction inlet into the pump chamber. To scale these insights for field-wide application, we established an optimized Stacking ensemble learning architecture—synthesizing SVR, BPNN, and Random Forest algorithms via an ANFIS meta-learner—which achieves remarkable predictive fidelity, with RMSE, MAE, and R^2 values of 1.75 L/min, 0.82 L/min, and 0.98.

Furthermore, systemic investigation of the surface infrastructure reveals that while the gathering network is dominated by low-frequency pulsations, the peak of the maximum pulsation frequency ranges from 0.008 to 0.007 Hz. The downstream tubular separators possess sufficient inertial buffering capacity to maintain a stable gas-liquid interface, both the liquid content at the gas outlet and the gas content at the liquid outlet are maintained below 12‰, the capture efficiency of liquid droplets with a particle size larger than 150 μm can reach 99%. This comprehensive modeling suite not only resolves the mechanistic origins of gas-bearing produced water but also provides a robust tool for the real-time prognostic assessment and optimization of CBM extraction and transportation systems.

Acknowledgement: Not applicable.

Funding Statement: This work was supported by the Qinchuangyuan “Scientist + Engineer” Team Construction Project under Grant (2025QCY-KXJ-020) and the Qinchuangyuan “Four Chains” Project under Grant (2025PT-ZCK-76).

Author Contributions: The authors confirm their contribution to the paper as follows: conceived and designed the analysis, collected the data, performed the analysis, and wrote the paper: Qin Zhao, Songfei Zhang. Data curation and investigation, revised the article: Yuxin Wang, Gang Chen. Methodology: Hui Zhang, Lei Wang, Mulin Zhou. Project administration, funding acquisition: Yu Weng. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the Corresponding Author, Yu Weng, upon reasonable request.

Ethics Approval: Not applicable.

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

References

1. Li C, Yang Z, Yan X, Wang G, Huang D, Lu B, et al. Dynamic distribution patterns and release behavior of adsorbed gas/free gas during deep coalbed methane production. *Fuel*. 2026;407:137310. [[CrossRef](#)].
2. Xu F, Hou W, Xiong X, Xu B, Wu P, Wang H, et al. The status and development strategy of coalbed methane industry in China. *Petrol Explor Dev*. 2023;50(4):765–83. [[CrossRef](#)].
3. Cheng C, Gong XP, Cheng XY, Xiao L, Ma XY. Optimum machine learning on gas extraction and production for adaptive negative control. *Front Heat Mass Transf*. 2025;23(3):1037–51. [[CrossRef](#)].
4. Zhou Z, Ballentine CJ, Kipfer R, Schoell M, Thibodeaux S. Noble gas tracing of groundwater/coalbed methane interaction in the San Juan Basin, USA. *Geochim Cosmochim Acta*. 2005;69(23):5413–28. [[CrossRef](#)].
5. Sun X, Yi J, Li J. Characteristics of water occurrence in coalbed methane reservoir. *Unconv Resour*. 2023;3:30–6.
6. Chakhmakhchev A. Worldwide coalbed methane overview. In: *Proceedings of the Hydrocarbon Economics and Evaluation Symposium; 2007 Apr 1–3; Dallas, TX, USA*. Richardson, TX, USA: Society of Petroleum Engineers; 2007. [[CrossRef](#)].
7. Huang H, Bi C, Sang S, Miao Y, Zhang H. Signature of coproduced water quality for coalbed methane development. *J Nat Gas Sci Eng*. 2017;47:34–46. [[CrossRef](#)].
8. Ravindranath NH, Sathaye JA. Greenhouse gas emissions. In: Beniston M, editor. *Climate change and developing countries*. Dordrecht, The Netherlands: Springer; 2002. p. 11–36. [[CrossRef](#)].
9. Howarth RW, Santoro R, Ingraffea A. Methane and the greenhouse-gas footprint of natural gas from shale formations. *Clim Change*. 2011;106(4):679–90. [[CrossRef](#)].
10. Omara M, Zimmerman N, Sullivan MR, Li X, Ellis A, Cesa R, et al. Methane emissions from natural gas production sites in the United States: Data synthesis and national estimate. *Environ Sci Technol*. 2018;52(21):12915–25. [[CrossRef](#)].
11. Liang W, Yan J, Zhang B, Hou D. Review on coal bed methane recovery theory and technology: Recent progress and perspectives. *Energy Fuels*. 2021;35(6):4633–43. [[CrossRef](#)].
12. Yang S, Yang W, Chen G, Fang X, Lv C, Zhong J, et al. Greenhouse gas emissions from oilfield-produced water in Shengli Oilfield, Eastern China. *J Environ Sci*. 2016;46:101–8. [[CrossRef](#)].
13. Herrmann M. A balanced force refined level set grid method for two-phase flows on unstructured flow solver grids. *J Comput Phys*. 2008;227(4):2674–706. [[CrossRef](#)].
14. Di Y, Li R, Tang T, Zhang P. Level set calculations for incompressible two-phase flows on a dynamically adaptive grid. *J Sci Comput*. 2007;31(1):75–98. [[CrossRef](#)].
15. Zhang J, Deng L, Li R, Li M, Liu J, Dai Z. Achieving high performance and portable parallel GMRES algorithm for compressible flow simulations on unstructured grids. *J Supercomput*. 2023;79(17):20116–40. [[CrossRef](#)].
16. Emery A, Dillon H, Mescher A. Analysis of chaotic natural convection in a tall rectangular cavity with non-isothermal walls. *Front Heat Mass Transf*. 2013;4(2):1–9. [[CrossRef](#)].

17. Liu B, Qian J, Jin Z, Huang L, Zou J. Air-side heat transfer performance prediction for microchannel heat exchangers using data-driven models with dimensionless numbers. *Front Heat Mass Transf.* 2024;22(6):1613–43. [[CrossRef](#)].
18. Xiao R, Wang D, Jin S, Yu H, Liu B. Algorithm and influence factor study on flow pattern transition from stratified flow to non-stratified flow of gas-liquid two-phase flow. *Front Heat Mass Transf.* 2021;16:1–9. [[CrossRef](#)].
19. Dottore EJ. How to prevent gas lock in sucker rod pumps (SPE 27010). In: *Proceedings of the SPE Latin America/Caribbean Petroleum Engineering Conference*; 1994 Apr 27–29; Buenos Aires, Argentina.
20. Lea JF, Winkler HW. What's new in artificial lift. *World Oil.* 1990;204(3):31–6.
21. Das G, Das PK, Purohit NK, Mitra AK. Flow pattern transition during gas liquid upflow through vertical concentric annuli: Part II: Mechanistic models. *J Fluids Eng.* 1999;121(4):902–7. [[CrossRef](#)].
22. Das G, Das PK, Purohit NK, Mitra AK. Phase distribution of gas-liquid mixtures in concentric annuli-inception and termination of asymmetry. *Int J Multiph Flow.* 2000;26(5):857–76. [[CrossRef](#)].
23. Gomez LE, Shoham O, Schmidt Z, Chokshi RN, Northug T. Unified mechanistic model for steady-state two-phase flow: Horizontal to vertical upward flow. *SPE J.* 2000;5(3):339–50. [[CrossRef](#)].
24. Jamalbayov MA, Valiyev NA. The discrete-imitation modeling concept of the “sucker-rod pump-well-reservoir” system and the optimization of the pumping process. *Petrol Res.* 2024;9(4):686–94. [[CrossRef](#)].
25. Zuo J, Wang S, Dong S, Li W, Lei M. Liquid flow rate measurement for sucker-rod pumping wells with incomplete fillage using motor power and MSISSA-BiLSTM. *Measurement.* 2025;249:117072. [[CrossRef](#)].
26. Jaiman R, Joshi V. *Computational mechanics of fluid-structure interaction.* Berlin/Heidelberg, Germany: Springer; 2022.
27. Liu M, Zhang Z. Smoothed particle hydrodynamics (SPH) for modeling fluid-structure interactions. *Sci China Phys Mech Astron.* 2019;62(8):984701. [[CrossRef](#)].
28. Andrade DM, de Freitas Rachid FB, Tijsseling AS. A new model for fluid transients in piping systems taking into account the fluid-structure interaction. *J Fluids Struct.* 2022;114:103720. [[CrossRef](#)].
29. Schussnig R, Pacheco DRQ, Kaltenbacher M, Fries TP. Semi-implicit fluid-structure interaction in biomedical applications. *Comput Meth Appl Mech Eng.* 2022;400:115489. [[CrossRef](#)].
30. Nyong OE, Igbong DI, Ebieta CE, Ekpo Ene B, Oluwadare B, Archibong Eso A. Numerical simulation of two-phase gas-liquid flowthrough horizontal annulus pipe. *Arch Thermodyn.* 2024:705–31. [[CrossRef](#)].
31. Cosner R, Oberkampf B, Rumsey C, Rahaim C, Shih T. AIAA committee on standards for computational fluid dynamics: Status and plans. In: *Proceedings of the 44th AIAA Aerospace Sciences Meeting and Exhibit*; 2006 Jan 9–12; Reno, NV, USA. [[CrossRef](#)].
32. Holubenko VP, Doroshenko YV. Study of the liquid-phase impact on gas flow rate during blowdowns of flowlines and wells at gas-condensate fields. *Arch Mater Sci Eng.* 2025;122(1):18–33. [[CrossRef](#)].
33. Hall KC. Eigenanalysis of unsteady flows about airfoils, cascades, and wings. *AIAA J.* 1994;32(12):2426–32. [[CrossRef](#)].
34. Dowell EH. Eigenmode analysis in unsteady aerodynamics—Reduced-order models. *AIAA J.* 1996;34(8):1578–83. [[CrossRef](#)].
35. Lucia DJ, Beran PS, Silva WA. Reduced-order modeling: New approaches for computational physics. *Prog Aerosp Sci.* 2004;40(1–2):51–117. [[CrossRef](#)].
36. Burges CJC. A tutorial on support vector machines for pattern recognition. *Data Min Knowl Discov.* 1998;2(2):121–67. [[CrossRef](#)].
37. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature.* 1986;323(6088):533–6. [[CrossRef](#)].
38. Breiman L. Random forests. *Mach Learn.* 2001;45(1):5–32. [[CrossRef](#)].
39. Jang JR. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Trans Syst Man Cybern.* 1993;23(3):665–85. [[CrossRef](#)].
40. Wolpert DH. Stacked generalization. *Neural Netw.* 1992;5(2):241–59. [[CrossRef](#)].
41. Bendiksen KH, Malnes D, Moe R, Nuland S. The dynamic two-fluid model OLGA: Theory and application. *SPE Prod Eng.* 1991;6(2):171–80. [[CrossRef](#)].