

Article

An Ordered Flow-State Identification Method for Unconventional Gas Wells Based on a Five-Region Analytical Model and RTA Window Features

Hang Yuan ^{1,2,3,4}, Yuping Sun ^{2,3,4,*}, Wei Xiong ^{1,2,3,4,*}, Deshang Wang ⁵, Yuzheng Gong ⁶, Yong Li ^{1,2,3}, Mingyan Sun ^{1,2,3} and Zejun Tang ^{1,2,3}

¹ University of Chinese Academy of Sciences, Beijing 100049, China

² Institute of Porous Flow and Fluid Mechanics, Chinese Academy of Sciences, Langfang 065007, China

³ Research Institute of Petroleum Exploration & Development, PetroChina, Beijing 100083, China

⁴ National Energy Shale Gas R&D (Experimental) Center, Langfang 065007, China

⁵ Chuannan Exploration and Development Division, PetroChina Jilin Oilfield Company, PetroChina Company Limited, Zigong 643000, China

⁶ PetroChina Tarim Oilfield Company, Korla 841000, China

* Correspondence: sunyuping01@petrochina.com.cn (Y.S.); xiongwei69@petrochina.com.cn (W.X.)

Abstract

Unconventional gas-well production is jointly controlled by fracture conductivity, stimulated-region supply, matrix replenishment, boundary propagation, and low-pressure fluid-property changes. In practice, RTA diagnostic curves are often affected by variable operating schedules, pressure-measurement errors, and production disturbances, making flow-stage boundaries difficult to define consistently. To reduce the subjectivity of manual interpretation and to capture stage evolution rather than whole-well classes, an ordered flow-state identification method based on a five-region analytical model and RTA sliding-window features is developed. A fully random, large-sample production-response library is generated with the five-region model. Each well production curve is divided into local time windows, from which dynamic features, including RNP, material-balance time, local slopes, pseudopressure derivatives, and normalized cumulative gas production are extracted. K-means clustering is then used to identify local states, which are reordered by material-balance time to form an ordered S1–S5 sequence. Results from 10,000 synthetic wells yielded 689,394 RTA windows, an inter-cluster separation of 1.8924, a stage-regression rate of 0.0238, and an average of 4.24 states per well. S1–S5 represent early fracture–stimulated-region response, stimulated-region supply development, matrix composite supply transition, enhanced boundary/control-volume effects, and late low-pressure property response, respectively. Application to Well M1 shows that S4 contributes the most gas (37.83%), followed by S5 (23.47%), indicating dominant mid-to-late effective supply and low-pressure long-tail production. The method converts empirical flow-regime division into reproducible and comparable window-state identification results, supporting stage diagnosis and production-strategy adjustment for unconventional gas wells.

Keywords: five-region analytical model; unconventional gas well; rate transient analysis; gas pseudo-pressure; sliding window; unsupervised clustering; ordered flow state

Academic Editor: Matthew Clarke

Received: 8 June 2026

Revised: 29 June 2026

Accepted: 2 July 2026

Published: 3 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Unconventional gas reservoirs generally have low porosity, low permeability, small pore scales, and strong heterogeneity. Their natural deliverability is limited, so horizontal wells with multistage hydraulic fracturing are commonly used to improve reservoir stimulation and contact [1,2]. After fracturing, a composite supply system is formed around the wellbore, consisting of hydraulic fractures, near-fracture stimulated regions, and unstimulated matrix regions. Gas is replenished from the distal matrix into the stimulated region and then enters the wellbore through the fracture system [3,4]. This process is jointly controlled by fracture conductivity, stimulated reservoir volume, matrix replenishment, boundary propagation, adsorption/desorption, diffusion/slip, and stress sensitivity [5,6].

Production data analysis and rate transient analysis (RTA) provide important tools for dynamic interpretation of unconventional gas wells. They can use daily gas rate, bottom-hole flowing pressure, cumulative gas production, and pseudopressure response to identify stage characteristics such as fracture linear flow, bilinear flow, stimulated-region supply, matrix replenishment, and boundary response [7,8]. Classical linear-flow analysis indicates that the relationship between normalized pressure or pseudopressure and the square root of time in fractured-well production curves can be used to identify linear flow and to estimate fracture half-length, permeability, and effective drainage area [9]. However, field production data are commonly affected by variable-rate schedules, shut-in and restart events, intermittent pressure tests, water-production disturbances, and pressure-measurement errors. Typical diagnostic curves, therefore, do not always show clear plateaus or standard slopes. Different interpreters may obtain different results when selecting start and end times, smoothing scales, and flow-regime boundaries, reducing the repeatability and comparability of stage division [10]. In addition, the early fracture response of some unconventional gas wells may last only a short time, and linear or bilinear flow may deviate from standard type-curve behavior because of operating schedules, multiphase flow, heterogeneity, and stress sensitivity [11].

These uncertainties make manual flow-stage division strongly interpreter dependent. Different choices of start and end times, smoothing scales, outlier treatment, and diagnostic type curves may lead to different flow-regime boundaries and may affect judgments regarding fracture conductivity, effective stimulated volume, matrix supply capacity, and boundary-response time. For studies that compare many production curves, it is difficult to establish a unified standard based only on manual type-curve interpretation [12]. A more suitable approach is to retain the physical basis of RTA flow-regime interpretation while converting local curve morphology into computable and comparable dynamic features, so that stage division is supported by similarity in feature space rather than by visual interpretation alone.

Analytical and semi-analytical models can provide physical references for this type of dynamic interpretation. Multilinear-flow models simplify the complex seepage process around fractured horizontal wells into several coupled linear-flow regions and can retain the main control relationships among fractures, stimulated regions, and external matrix at relatively low computational cost [13,14]. The five-region analytical model proposed by Stalgorova and Mattar further divides the drainage volume of a single fracture in a multistage fractured horizontal well into the fracture region, near-fracture stimulated region, and external supply regions in different directions. It describes composite supply under finite stimulated volume through pressure-continuity and rate-continuity conditions [15]. This model is naturally connected with RTA flow-regime interpretation: the fracture region corresponds to early fracture response, the near-fracture stimulated region to stimulated-region supply, the external matrix to matrix replenishment and composite supply,

the outer boundary to boundary control, and low-pressure property corrections to late adsorption and compressibility response.

High-fidelity numerical simulation can describe complex fracture networks, reservoir heterogeneity, multiple gas-transport mechanisms, and geomechanical responses in greater detail and is suitable as a physical benchmark for complex production dynamics [16,17]. However, numerical simulation is computationally expensive under multiparameter, large-sample, and repeated-trial conditions [18]. When the objective shifts from fine single-well history matching to stage identification for many dynamic curves, the rapid forward-modeling advantage of analytical models becomes more prominent. Previous studies have also shown that samples generated by physical models or numerical simulation can be combined with machine learning methods for production prediction, parameter optimization, and extraction of development patterns [19]. Therefore, constructing a large-sample dynamic response library with the five-region analytical model provides a practical balance between computational efficiency and physical constraints.

Data-driven methods provide new tools for batch interpretation of complex production dynamics. Machine learning has been applied to shale-gas-well production evaluation and productivity prediction, showing that nonlinear relationships can be extracted from geological, engineering, and early production data [20–22]. Physics-constrained, domain-adaptation, and ensemble-learning methods have also been used to improve the stability of production forecasting and estimated ultimate recovery (EUR) evaluation [23–26]. Nevertheless, if black-box models are used directly to predict pressure or rate, their results often cannot explain the physical relationships among fractures, stimulated regions, and matrix supply. In the present study, data mining is more suitable for stage identification and similarity summarization than for replacing the five-region model in pressure prediction. Unsupervised learning is appropriate for production-dynamic samples lacking reliable manual labels, principal component analysis can reduce high-dimensional feature redundancy, and K-means clustering can divide similar samples in feature space [27,28]. In unconventional oil and gas development evaluation, clustering methods have been used for type-well construction, analog-well identification, and EUR prediction improvement [29–31]. However, if features are derived only from whole-well statistics, the results tend to remain at the level of high/low productivity or fast/slow decline and cannot readily explain the flow mechanisms behind different stages.

Based on these considerations, this study uses the five-region analytical model as the physical backbone, RTA sliding-window features as the dynamic representation, and unsupervised clustering as the identification tool to develop an ordered flow-state identification method for unconventional gas wells. The focus is not to assign one category to an entire well, but to divide the production history into comparable stage sequences. The five-region model generates large-sample responses satisfying material balance and seepage mechanisms; RTA windows extract local dynamic morphology; the clustering algorithm provides comparable state groups; and ordering by material-balance time restores the production sequence. Because wellbore storage and the earliest fracture response are very short in daily production data and are difficult to identify reliably as separate stages, this study combines F1 wellbore/fracture rapid response with F2 fracture-stimulated-region bilinear flow, and further identifies stimulated-region linear supply, matrix composite supply transition, boundary/control-volume effects, and late low-pressure property response. The resulting ordered S1–S5 state sequence can be used for real-well stage diagnosis, evaluation of effective supply duration, and analysis of late production-strategy adjustment.

<https://doi.org/10.3390/en19133172>

linear flow. When the near-fracture stimulated region continuously feeds the fracture, the production curve is closer to the stimulated-region linear supply. As pressure propagates further into the external matrix, the response gradually enters a transition of matrix composite supply. After pressure reaches the control boundary, the boundary and control volume begin to decline. As pressure decreases further, adsorption/desorption, porosity changes, water-saturation changes, and changes in total compressibility enhance the low-pressure property response.

Accordingly, the five-region model can provide physical interpretations for six theoretical flow regimes. F1, wellbore/fracture rapid response, mainly corresponds to the wellbore and fracture region; F2, fracture–stimulated-region bilinear flow, corresponds to coupled supply from the fracture and near-fracture stimulated region; F3, stimulated-region linear supply, mainly corresponds to the near-fracture stimulated region; F4, matrix composite supply transition, corresponds to gradual participation of the matrix outside the stimulated region; F5, boundary/control-volume influence, corresponds to supply attenuation after pressure reaches the outer boundary; and F6, late low-pressure property response, corresponds to enhanced adsorption, compressibility, and saturation-change effects under low-pressure conditions.

Because gas viscosity and the compressibility factor vary markedly with pressure during unconventional gas-reservoir production, direct use of pressure produces strong nonlinearity in the governing equations. Therefore, real-gas pseudopressure is introduced to transform pressure into a pseudopressure variable that accounts for gas viscosity and the compressibility factor:

$$m(p) = 2 \int_{p_{ref}}^p \frac{p}{\mu_g(p)Z(p)} dp \quad (1)$$

where $m(p)$ is the real-gas pseudopressure, MPa^2/cP ; p is reservoir pressure, MPa ; p_{ref} is the reference pressure, MPa ; $\mu_g(p)$ is gas viscosity, cP ; and $Z(p)$ is the gas compressibility factor. The pseudopressure treatment weakens the nonlinear influence of pressure-dependent gas properties and enables the five-region analytical model to calculate gas-reservoir production dynamics.

Considering pore-volume changes caused by pressure decline during production, porosity is corrected as a function of pressure. The porosity correction used in this study is

$$\frac{\phi(p)}{\phi_i} = 1 - C_f(p_i - p) \quad (2)$$

where $\phi(p)$ is the porosity at pressure p , ϕ_i is the initial porosity, C_f is the rock compressibility, and p_i is the initial reservoir pressure, MPa . This equation represents the influence of pore-volume shrinkage on storage space as pressure decreases.

Because retained fracturing fluid and water-phase expansion can change the gas-water saturation distribution, the model further corrects water saturation. The water-phase volume change is expressed as

$$\frac{V_w(p)}{V_{wi}} = e^{[C_w(p_i - p)]} \quad (3)$$

where $V_w(p)$ is the water-phase volume at pressure p , m^3 ; V_{wi} is the initial water-phase volume, m^3 ; and C_w is the water compressibility, $1/\text{MPa}$. Combining this relation with pore-volume change gives the water-saturation correction:

$$S_w(p) = S_{wi} \frac{V_w(p)/V_{wi}}{\phi(p)/\phi_i} \quad (4)$$

where $S_w(p)$ is the water saturation at pressure p and S_{wi} is the initial water saturation. The gas saturation is

$$S_g(p) = 1 - S_w(p) \quad (5)$$

On this basis, the total compressibility consists of gas compressibility, water compressibility, rock compressibility, and the adsorbed-gas storage term:

$$C_t(p) = S_g(p)C_g(p) + S_w(p)C_w + C_f + C_{ads}(p) \quad (6)$$

where $C_t(p)$ is the total compressibility, 1/MPa; $C_g(p)$ is gas compressibility, 1/MPa; and $C_{ads}(p)$ is the equivalent compressibility term for adsorbed gas, 1/MPa, calculated from the Langmuir adsorption parameters, pressure, compressibility factor, and porosity.

To handle nonlinearities caused by pressure-dependent gas properties, porosity, water saturation, and total compressibility, the model adopts an improved pseudotime. The pseudotime factor is defined as

$$F(p) = \frac{\phi_i \mu_i C_{t,i}}{\phi(p) \mu_g(p) C_t(p)} \quad (7)$$

where μ_i and $C_{t,i}$ are the gas viscosity and total compressibility at the initial pressure, respectively. Using trapezoidal integration gives the cumulative pseudotime:

$$\tau_n = \sum_{r=2}^n \frac{F_{r-1} + F_r}{2} (t_r - t_{r-1}) \quad (8)$$

where τ_n is the cumulative pseudotime at the n th time point, and t_r is physical time, days. This pseudotime accounts for changes in gas viscosity, total compressibility, porosity, and saturation; it is therefore more suitable than a pseudotime based only on gas compressibility for describing production dynamics under gas-water coexistence and changing storage space.

The cumulative pseudotime is further converted into the dimensionless time required by the five-region analytical model:

$$t_D = 86400 \frac{\eta_i}{y_f^2} \tau \quad (9)$$

where the initial hydraulic diffusivity is

$$\eta_i = \frac{k_1 \times 9.86923 \times 10^{-16}}{\phi_i \mu_i C_{t,i}} \quad (10)$$

where t_D is dimensionless time, η_i is the initial hydraulic diffusivity, m^2/s ; y_f is the fracture half-length, m; and k_1 is the permeability of the near-fracture stimulated region, mD.

The five-region analytical model obtains the dimensionless bottom-hole pseudopressure response in Laplace space and then uses Stehfest numerical inversion to obtain the time-domain result. For the n th time point, the inversion can be written as

$$f_D(t_D) = \frac{\ln 2}{t_D} \sum_{v=1}^{N_s} V_v \bar{f}_D \left(\frac{v \ln 2}{t_D} \right) \quad (11)$$

where $f_D(t_D)$ is the time-domain dimensionless bottom-hole pseudopressure response, $\bar{f}_D(s)$ is the Laplace-space kernel function of the five-region model, s is the Laplace variable, N_s is the number of Stehfest inversion terms, and V_v is the Stehfest weight coefficient.

Because the program used in this study adopts a mixed SI and field-unit calculation system, the conversion between dimensionless pseudopressure and real pseudopressure does not use the imperial-unit constant form, but instead uses the unit-conversion coefficient in the program:

$$\Gamma = \frac{Tq_g}{760k_1h} \quad (12)$$

where Γ is the pseudo-pressure conversion coefficient, q_g is the gas rate, m^3/d , and h is the reservoir thickness, m . Based on the dimensionless pseudo-pressure response calculated by the five-region model, the bottom-hole pseudo-pressure can be back-calculated as

$$m(p_{wf}) = m(p_i) - \Gamma f_D \quad (13)$$

The real bottom-hole flowing pressure is then obtained from the bottom-hole pseudo-pressure:

$$p_{wf} = m^{-1}[m(p_i) - \Gamma f_D] \quad (14)$$

where m^{-1} denotes the inverse relation used to convert pseudo-pressure back to real pressure.

Through the above pseudo-pressure treatment, porosity-saturation correction, total-compressibility calculation, improved pseudotime, and Laplace-Stehfest inversion procedure, bottom-hole flowing pressure and related production-dynamic responses can be obtained at different production times. The forward process preserves the physical supply relationships among the fracture region, stimulated region, and matrix region, while maintaining high computational efficiency, and therefore provides the basic data for large-sample dynamic response-library construction, feature extraction, and unsupervised dynamic zoning.

2.2. Construction of a Fully Random Large-Sample Dynamic Response Library

The large-sample dynamic response library is the basis for ordered flow-state identification. To avoid overly idealized clustering caused by strong preassignment of classes, this study constructs the synthetic sample library by fully random parameter sampling. The sampled parameters are constrained within physically meaningful engineering ranges, including initial pressure, daily gas-rate schedule, matrix and external-supply-region permeability, near-fracture stimulated-region permeability, fracture conductivity, fracture geometry, fracture spacing, stimulated-region width, boundary range, rock and water compressibility, water saturation, porosity, and Langmuir adsorption parameters. The wellbore-storage coefficient and skin factor are kept uniform and are not treated as random variables.

The sample library is constructed using a geological reserve closure concept. The control volume of a single fracture is determined by fracture spacing, control boundary, and reservoir thickness; the stimulated volume is determined by fracture half-length, stimulated-region width, and reservoir thickness. On this basis, free-gas and adsorbed-gas geological reserves are calculated, and the average reservoir pressure is updated through the material-balance relationship. Bottom-hole flowing pressure is obtained from the five-region analytical model, improved pseudotime, and variable-rate superposition. This workflow maintains consistency among initial pressure, rate schedule, reserve scale, and pressure decline, and avoids pressure curves that violate material balance.

During sampling, the permeabilities of the external matrix supply regions are kept the same, that is, $k_2 = k_3 = k_4 = k_a$. Gas density at standard conditions is treated as a fixed property parameter rather than an adjustable scaling coefficient. The main analysis samples are not assigned hidden labels such as high-conductivity, stimulated-region supply, or adsorption-contribution types. Instead, the unsupervised algorithm identifies stage structure in continuous parameter space and dynamic-feature space. This treatment better reflects the coupled multiple mechanisms in field production data; the goal is not to assign a unique label to an entire well, but to determine the dominant response in different time intervals.

It should be emphasized that the synthetic sample library is not intended to use machine learning to rediscover unknown flow physics, nor is it used as a substitute for field-

reservoir behavior. In this study, the five-region analytical model serves as a physics-constrained sample generator and provides a physically interpretable dynamic-feature space. The role of clustering is to convert the experience-based separation of RTA flow stages into reproducible and comparable window-state identification. Therefore, the synthetic samples are used to establish a controlled physical reference space, whereas the real-well case is used to test the engineering applicability of the ordered-state workflow.

2.3. Construction of Sliding-Window RTA Dynamic Features

Based on the correspondence between the five-region model and typical flow regimes, the sample unit for dynamic identification is not the entire well but a local sliding window on the production curve. This is because the same well may be controlled successively by fracture conductivity, stimulated-region supply, matrix replenishment, boundary propagation, and low-pressure property changes at different production times. If clustering is performed only with whole-well statistics, early flow regimes, mid-stage supply, and late decline are compressed into one mixed label, making it difficult to explain the dominant mechanisms of different stages.

Window segmentation provides only local production segments and does not directly assign flow-regime labels. The RTA features extracted from each window enter the clustering algorithm, several dynamic states are formed according to feature similarity, and the states are then ordered by material-balance time to obtain an ordered state sequence. In other words, theoretical flow regimes guide the choice of window scale and the interpretation of results, but F1–F6 labels are not input into the clustering process, preventing manual classification from being embedded directly in the algorithm.

F1 wellbore/fracture rapid response usually occurs at the earliest stage of production and may last only from several hours to several days. The samples in this study use daily production data with a minimum time resolution of 1 d. Separately identifying F1 therefore provides too few valid data points and may incorrectly treat opening schedules, pressure-measurement errors, or early noise as an independent flow state. Accordingly, F1 is combined with F2 fracture–stimulated-region bilinear flow in the sliding-window design and is treated as an early fracture–stimulated-region response class.

The window settings are kept consistent with the five-region/flow-regime mapping. The combined F1–F2 stage corresponds to 0–15 d, with a window length of 5 d and a step of 1 d. F3 stimulated-region linear supply corresponds to 5–180 d, with a window length of 30 d and a step of 10 d. F4 matrix composite supply transition corresponds to 60–730 d, with a window length of 120 d and a step of 60 d. F5 boundary/control-volume influence corresponds to 365–3000 d, with a window length of 365 d and a step of 180 d. F6 late low-pressure property response corresponds to time after 1000 d, with a window length of 730 d and a step of 365 d. Thus, the minimum window length is 5 d, the maximum window length is 730 d, and the minimum sliding step is 1 d.

$$W_{i,r} = \{t \mid t_{i,r}^{\text{start}} \leq t \leq t_{i,r}^{\text{end}}\} \quad (15)$$

where $W_{i,r}$ is the local dynamic window in a single-well production curve, i denotes the well number, r denotes the window index, t_{start} and t_{end} are the start and end times of the window, and t is the production time within the window. Multiscale window lengths are adopted: short early windows preserve rapid fracture and near-well responses as much as possible, whereas longer mid-to-late windows reduce the influence of daily fluctuations on supply and boundary features.

Because gas viscosity and the compressibility factor vary significantly with pressure, the RTA features are constructed using gas pseudo-pressure drawdown:

$$RNP(t) = \frac{m(p_i) - m[p_{wf}(t)]}{q_g(t)} \quad (16)$$

where RNP is the pseudo-pressure-based normalized pressure response, $m(p_i)$ is the gas pseudo-pressure at initial pressure, $m(p_{wf})$ is the gas pseudo-pressure corresponding to bottom-hole flowing pressure, and q_g is the daily gas rate, m³/d. Using pseudo-pressure drawdown rather than pressure difference reduces the nonlinear influence of pressure-dependent gas viscosity and compressibility factor. Material-balance time is constructed from cumulative gas production and instantaneous gas rate, converting a variable-rate production history into a rate-related equivalent time scale and reducing the influence of different rate schedules on dynamic-stage identification:

$$t_{mb}(t) = \frac{G_p(t)}{q_g(t)}, \quad G_p(t_n) = \sum_{j=1}^n q_g(t_j) \Delta t_j \quad (17)$$

In each sliding window, additional features are extracted, including the local slope between RNP and $\sqrt{t_{mb}}$, the log-log slope between RNP and t_{mb} , the median pseudo-pressure derivative, the daily gas-decline slope, gas-production capacity per unit pseudo-pressure drawdown, normalized cumulative gas production, and median window time. These features are all derived from the production-dynamic curve itself and can reflect differences in pressure-propagation speed, supply capacity, and decline intensity under different flow states.

$$\mathbf{x}_{i,r} = [RNP, t_{mb}, \beta_{\sqrt{t}}, \beta_{\log}, D_p, I_q, G_p/G_{p,\max}, q_g, \Delta q_g]^T \quad (18)$$

where $\mathbf{x}_{i,r}$ is the dynamic feature vector of the r th sliding window of the i th well; $\beta_{\sqrt{t}}$ is the local slope between RNP and $\sqrt{t_{mb}}$; $\beta_{\log}$ is the log-log slope between RNP and t_{mb} ; D_p is the representative pseudo-pressure derivative in the window; I_q is gas-production capacity per unit pseudo-pressure drawdown; $G_p/G_{p,\max}$ is normalized cumulative gas production; q_g is the representative daily gas rate in the window; and Δq_g is the change in daily gas rate within the window. The feature vector covers five types of information: pressure response, time scale, derivative morphology, production efficiency, and rate variation.

The local slope between RNP and $\sqrt{t_{mb}}$ in a window is defined as

$$\beta_{\sqrt{t_{mb}}, i, r} = \frac{\sum_{j \in W_{i,r}} (\sqrt{t_{mb,j}} - \overline{\sqrt{t_{mb}}}) (RNP_j - \overline{RNP})}{\sum_{j \in W_{i,r}} (\sqrt{t_{mb,j}} - \overline{\sqrt{t_{mb}}})^2} \quad (19)$$

where $\beta_{\sqrt{t_{mb}}}$ characterizes the local response intensity of RNP with respect to the square root of material-balance time and is commonly used to capture early-to-middle linear-flow or near-well supply-response changes.

The log-log slope is defined as

$$\beta_{\log, i, r} = \frac{\sum_{j \in W_{i,r}} (\ln t_{mb,j} - \overline{\ln t_{mb}}) (\ln RNP_j - \overline{\ln RNP})}{\sum_{j \in W_{i,r}} (\ln t_{mb,j} - \overline{\ln t_{mb}})^2} \quad (20)$$

where $\beta_{\log}$ reflects the local curve morphology of RNP and t_{mb} on log-log coordinates. Its variation can be used to distinguish different stages of pressure propagation and supply decline.

The representative pseudo-pressure derivative is defined as

$$D_{p, i, r} = \text{median}_{j \in W_{i,r}} \left(\frac{d \Delta m_{p,j}}{d \ln t_j} \right), \quad \Delta m_{p,j} = m(p_i) - m(p_{wf,j}) \quad (21)$$

where D_p is the median derivative of pseudo-pressure drawdown with respect to logarithmic time within the window, and Δm_p is the difference between the initial pseudo-pressure and the pseudo-pressure corresponding to bottom-hole flowing pressure. D_p characterizes the strength of pressure response and derivative morphology and is an important dynamic indicator for identifying transitions between flow states.

Gas-production capacity per unit pseudo-pressure drawdown and normalized cumulative gas production are defined as

$$I_{q,i,r} = \frac{\Delta G_{p,i,r}}{\Delta m_{p,i,r}^{end}}, \quad \frac{G_p}{G_{p,max}} = \frac{G_p(t)}{\max(G_p)} \quad (22)$$

where $\Delta G_{p,i,r}$ is the cumulative gas-production increment within the window, and $\Delta m_{p,i,r}^{end}$ is the pseudo-pressure drawdown at the end of the window. I_q denotes the gas production obtained per unit pseudo-pressure drawdown. A larger value indicates that more gas is produced under the same pseudo-pressure drawdown. $G_p/G_{p,max}$ represents the degree of recovery.

The representative daily gas rate and rate variation in a window are written as

$$\bar{q}_{g,i,r} = \frac{1}{n_r} \sum_{j \in W_{i,r}} q_{g,j}, \quad \Delta q_{g,i,r} = q_{g,e} - q_{g,s}, \quad b_{q,i,r} = \frac{d \ln q_g}{dt} \quad (23)$$

where $\bar{q}_g$ is the average daily gas rate within the window, Δq_g is the change in daily gas rate between the start and end of the window, and b_q is the logarithmic daily gas-decline slope. In actual clustering features, $\bar{q}_g$ or b_q can be selected according to data quality to represent the intensity of rate variation.

The components of the dynamic feature vector do not directly correspond to individual five-region model parameters, but provide a low-dimensional summary of production-curve morphology. Local slopes mainly reflect pressure propagation and supply response; derivative indicators reflect changes in flow state; gas-production capacity per unit pseudo-pressure drawdown reflects production efficiency under drawdown consumption; and normalized cumulative gas production and median window time characterize production-stage progression.

2.4. Unsupervised Identification of Ordered Flow States

Before sliding-window features enter clustering, they are standardized to remove the influence of different dimensions and magnitudes on distance calculation. During clustering, no manual flow-regime labels are input, and five-region model parameters are not used directly to calculate sample distances, thereby avoiding forced constraints from preset physical classes. The five-region model parameters are used only after clustering to interpret the dominant controls of different ordered states.

$$z_{i,j} = \frac{x_{i,j} - \mu_j}{\sigma_j} \quad (24)$$

where x_{ij} is the original value of the j th feature of the i th window, μ_j and σ_j are the mean and standard deviation of the j th feature among all window samples, and z_{ij} is the standardized dimensionless feature value. Standardization prevents variables with different dimensions, such as daily gas rate, cumulative gas production, slopes, and derivatives, from receiving unreasonable weights in Euclidean distance.

In the standardized feature space, K-means groups samples by minimizing the squared distance between each window sample and its assigned cluster center. The objective function is:

$$J = \sum_{k=1}^K \sum_{\mathbf{z}_i \in C_k} \|\mathbf{z}_i - \boldsymbol{\mu}_k\|_2^2 \quad (25)$$

where J is the clustering objective function, K is the number of clusters, C_k is the k th cluster, z_i is the standardized feature vector of the i th window, and μ_k is the center of the k th cluster. A smaller objective function indicates that dynamic features within the same cluster are more similar, but the physical meaning of the clustering result must still be interpreted with the time sequence and flow features.

The choice of $K = 5$ is based on both physical interpretability and data-resolution considerations. RTA interpretation of fractured unconventional gas wells may involve wellbore or fracture-storage response, bilinear flow, fracture or stimulated-region linear flow, matrix or compound supply, radial or boundary-controlled response, and late low-pressure property effects. However, not all of these stages can be resolved stably from daily production data. In particular, the wellbore/fracture rapid response may last only several hours to a few days and is easily mixed with opening schedules and pressure-measurement noise. Therefore, the earliest rapid response and fracture–stimulated-region bilinear response are combined into one early state, while the remaining states represent stimulated-region supply, matrix/compound supply transition, boundary or control-volume influence, and late low-pressure property response. In this way, $K = 5$ is not selected only as a numerical clustering parameter, but as a compromise among RTA flow-regime theory, the five-region model structure, and the resolution of the available production data.

The labels obtained by K-means have no inherent time meaning: label 1 does not necessarily represent an early stage, and label 5 does not necessarily represent a late stage. To interpret clustering results as a production-evolution process, the clusters are further ordered from small to large according to the median material-balance time and are defined as S1, S2, S3, S4, and S5. After this treatment, a single well can be represented by an ordered state sequence rather than a fixed well-type label. The sequence describes the transition from early fracture response to mid-to-late supply decline and low-pressure property control.

$$S_k = \text{rank}[\text{median}(t_{mb} | \mathbf{z}_i \in C_k)], \quad S_{i,r}^* = \max_{1 \leq l \leq r} S_{i,l} \quad (26)$$

where S_k is the k th ordered flow state after sorting, $\text{median}(t_{mb} | C_k)$ is the median material-balance time within the k th cluster, $S_{i,r}$ is the original state index of the r th window of the i th well, $S_{i,r}^*$ is the ordered state index after monotonic constraints, and l denotes windows of the same well that are not later than the r th window. The max operation reduces stage regression caused by label jumping between adjacent windows and makes the state sequence more consistent with the understanding that production evolves from early to late.

Identification quality is evaluated from four aspects: state separability, temporal ordering, life-cycle coverage, and gas-production contribution. Inter-cluster separation measures the degree of distinction among states in feature space. The stage-regression rate tests whether the state sequence follows the direction of production from early to late. The number of states experienced by each well describes the completeness of the dynamic process. The proportion of wells entering S5 indicates whether the sample set covers late low-pressure states. Gas-production contribution is further calculated for each state to evaluate its contribution to total gas production.

$$f_{S5} = \frac{N_{\text{well}}(S5)}{N_{\text{well}}}, \quad C_s = \frac{\sum_{\Delta t_r \in S_s} \Delta G_{p,r}}{\sum_r \Delta G_{p,r}} \quad (27)$$

where f_{S5} is the proportion of wells entering the S5 state, $N_{\text{well}}(S5)$ is the number of wells that experience the S5 state at least once, and N_{well} is the total number of sample wells. C_s is the gas-production contribution ratio of the s th ordered flow state, $\Delta G_{p,r}$ is the cumulative gas-production increment within the r th window, and $\Delta t_r \in S_s$ indicates that the win-

dow belongs to the sth ordered flow state. f_{ss} is used to judge whether the sample covers late low-pressure response, and C_s quantifies the contribution of different flow states to total gas production.

2.5. Physical Interpretation of Stages and Real-Well Application

The goal of ordered-state identification is not to assign a static class to a well, but to determine the occurrence time, duration, and gas-production contribution of different stages during production. State interpretation follows two principles: first, the production sequence is kept from early to late according to material-balance time; second, the dominant mechanisms are interpreted using the spatial supply relationships in the five-region model. In general, S1 corresponds to early fracture and near-well rapid response; S2 corresponds to fracture–stimulated-region response and early stimulated-region supply; S3 corresponds to the transition from stimulated-region linear supply to matrix-participating supply; S4 corresponds to matrix composite supply and enhanced boundary influence; and S5 corresponds to late low-pressure properties, adsorption/desorption, and enhanced compressibility effects. In actual samples, the physical meaning of each state must also be evaluated together with gas-production contribution, duration, and dynamic-feature intensity.

$$T_s = \sum_{\Delta t_r \in S_s} \Delta t_r, \quad Q_s = \sum_{\Delta t_r \in S_s} \Delta G_{p,r} \quad (28)$$

where T_s is the duration of the sth ordered state, Q_s is the cumulative gas production within the sth ordered state, C_s is the gas-production contribution ratio of the sth state, and $\Delta G_{p,r}$ is the gas-production increment of window r . T_s reflects how long a flow state lasts during production, and C_s reflects its contribution to total gas production. Together, they can be used to determine which flow state provides the strongest support for productivity.

For real-well application, field daily gas rate, bottom-hole flowing pressure, and cumulative gas production are converted into the same RTA window features and then passed through the state-identification workflow trained with the sample library to obtain the S1–S5 sequence. If S1 or S2 lasts for an abnormally short time, early fracture conductivity or near-fracture stimulated response may be insufficient. If S3 is short, the effective stimulated-region supply capacity may be limited. If S4 appears early, boundary influence or constrained matrix supply may be strong. If S5 occupies a much larger proportion, adsorption contribution, compressibility changes, and pressure-controlled production in the low-pressure stage should be evaluated. Thus, the clustering result is not merely a grouping label, but can be translated into a basis for fracturing-effect evaluation, stable-production-stage identification, and late production-strategy adjustment.

3. Results and Discussion

3.1. Fully Random Sample Library and Parameter Ranges

A total of 10,000 synthetic well samples with geological reserves closure were constructed. During sample generation, initial pressure, daily gas-rate schedule, fracture geometry, stimulated-region width, fracture spacing, control boundary, permeability, porosity, water saturation, compressibility, and Langmuir adsorption parameters were randomly sampled within engineering ranges (Table 1). The initial pressure P_i was randomly sampled from 40 to 110 MPa. The initial daily gas rate was related to P_i : high-pressure wells corresponded to higher initial rates, and low-pressure wells corresponded to lower initial rates. The daily gas rate generally ranged from 200 to 1600 m³/d for a single fracture, was positively correlated with P_i , and decreased with time. The wellbore-storage coefficient

cient and skin factor were not used as adjustable scaling coefficients, and the maximum development time was set to 7300 days (20 years).

Table 1. Parameter ranges and treatments for the fully random large-sample set.

Parameter	Physical Meaning	Range or Treatment
P_i	Initial reservoir pressure	40–110 MPa
q_g	Daily gas rate	200–1600 m ³ /d
k_1	Near-fracture stimulated-region permeability	1.0×10^{-5} – 1.0×10^{-1} mD
k_2, k_3, k_4	External matrix supply-region permeability	1.0×10^{-6} – 1.0×10^{-4} mD
k_f	Fracture permeability	1–10,000 mD
x_f	Fracture width	0.001–0.010 m
y_f	Fracture half-length	20–300 m
y_e	Single-fracture control half-length	120–600 m
L	Fracture spacing or control-length parameter	8–100 m
X_w	Stimulated-region width	1–40 m
C_f	Rock compressibility	1.0×10^{-5} – 1.0×10^{-3} 1/MPa
S_w	Initial water saturation	0.20–0.60
φ	Porosity	0.02–0.10
C_w	Water compressibility	1.0×10^{-5} – 1.0×10^{-3} 1/MPa
VL	Langmuir volume	0.2–14 cm ³ /g
PL	Langmuir pressure	2–16 MPa

The parameter ranges cover the main factors controlling fracture conductivity, stimulated-region supply capacity, matrix replenishment capacity, boundary-control range, adsorption parameters, and compressibility parameters. Because no hidden classes are preset in the samples, the subsequent identification results better reflect multi-factor coupling and continuous transition characteristics in real production data.

3.2. Identification Quality of the Ordered Five-State Sequence

The sample library generated 689,394 RTA sliding-window samples. Each window represents the dynamic response of a well over a local time interval; subsequent clustering is performed on window states rather than on entire wells. Consecutive windows with the same state are further merged into automatic stage segments, yielding 42,378 stage segments for statistics of state duration, state-transition direction, and gas-production contribution. After unsupervised clustering with $K = 5$, the cluster labels were reordered from small to large according to the median material-balance time of each cluster, producing five ordered flow states, S1–S5. It should be emphasized that S1–S5 are not manually predefined flow-regime labels; they are a state sequence obtained by clustering RTA windows in feature space and then ordering them according to production progression. The five-region model and theoretical flow regimes guide window-scale design and result interpretation, but F1–F6 labels are not directly input into the clustering algorithm (Table 2).

Table 2. Quality evaluation of ordered flow-state identification.

Evaluation Index	Result	Interpretation
Number of dynamic windows	689,394	Number of local production-dynamic samples used for clustering
Number of synthetic wells	10,000	Number of fully random five-region model samples
Number of automatic stage segments	42,378	Number of stages obtained after merging consecutive windows with the same state
Stage-regression rate	0.0238	Low value, indicating that the state sequence generally follows production progression
Average number of states per well	4.24	Indicates relatively complete dynamic-state coverage for single wells
Proportion entering S5	1.0000	Most wells enter the late low-pressure response state

Inter-cluster separation	1.8924	The five states are separable in dynamic-feature space
--------------------------	--------	--

The stage-regression rate is 0.0238, indicating that most window states progress from S1 to S5 and only a small number of windows show local regression. Such regression usually arises from overlap between adjacent windows, rate disturbances, or curve-shape transitions and does not affect the overall ordered-evolution characteristics. Each well experiences an average of 4.24 states, showing that most sample wells can form a relatively complete sequence of early response, mid-stage supply, boundary influence, and late low-pressure response.

To further examine the robustness of this selection, standard clustering metrics were calculated for $K = 3$ –8 using the same RTA-core feature set. As K increases from 3 to 8, the inertia per sample decreases from 9.8669 to 5.3911, which is expected because more clusters are reduced within-cluster dispersion. The silhouette score is highest at $K = 3$ (0.3270), indicating that three broad groups are more separated in a purely geometric sense; however, $K = 3$ merges several physically different responses and cannot represent the ordered evolution from early fracture/stimulated-region response to late boundary and low-pressure behavior. For $K = 5$, the inertia per sample is 7.2805, the silhouette score is 0.2740, the Davies–Bouldin index is 1.1954, and the Calinski–Harabasz index is 16,515. Larger K values further reduce inertia but split transitional responses into less stable subclasses. Therefore, $K = 5$ is selected as a physically interpretable compromise rather than as the mathematical optimum of a single clustering index (Table 3).

Table 3. Standard clustering metrics for different numbers of ordered states.

K	Inertia Per Sample	Mean Silhouette	Davies–Bouldin Index	Calinski–Harabasz Index	Interpretation
3	9.8669	0.3270	1.0648	17,819.73	Best geometric separation, but too coarse for flow-state evolution
4	8.2393	0.2893	1.1207	17,518.52	Still merges several transitional flow responses
5	7.2805	0.2740	1.1954	16,514.79	Selected compromise between separability and physical interpretability
6	6.4589	0.2714	1.0907	16,164.02	Lower inertia, but begins to split transitional subclasses
7	5.8394	0.2785	1.0721	15,782.81	Additional classes improve local fitting but reduce state simplicity
8	5.3911	0.2640	1.1192	15,246.55	Lowest inertia among tested cases, but over-fragments continuous transitions

Figure 2 shows that the reduction in inertia is monotonic as K increases, which is a normal result of K -means because additional clusters reduce within-cluster dispersion. The silhouette coefficient and Calinski–Harabasz index are higher at $K = 3$, indicating that three broad groups are more compact and better separated in a geometric sense. This solution, however, merges early fracture–stimulated-region response, middle supply transition, and late boundary or low-pressure response into overly coarse states. When K is greater than 5, inertia continues to decrease, but transitional responses are split into smaller subclasses and the physical meaning becomes less stable. Therefore, $K = 5$ is used as a physically constrained compromise between statistical separability, daily data resolution, and flow-state interpretability.

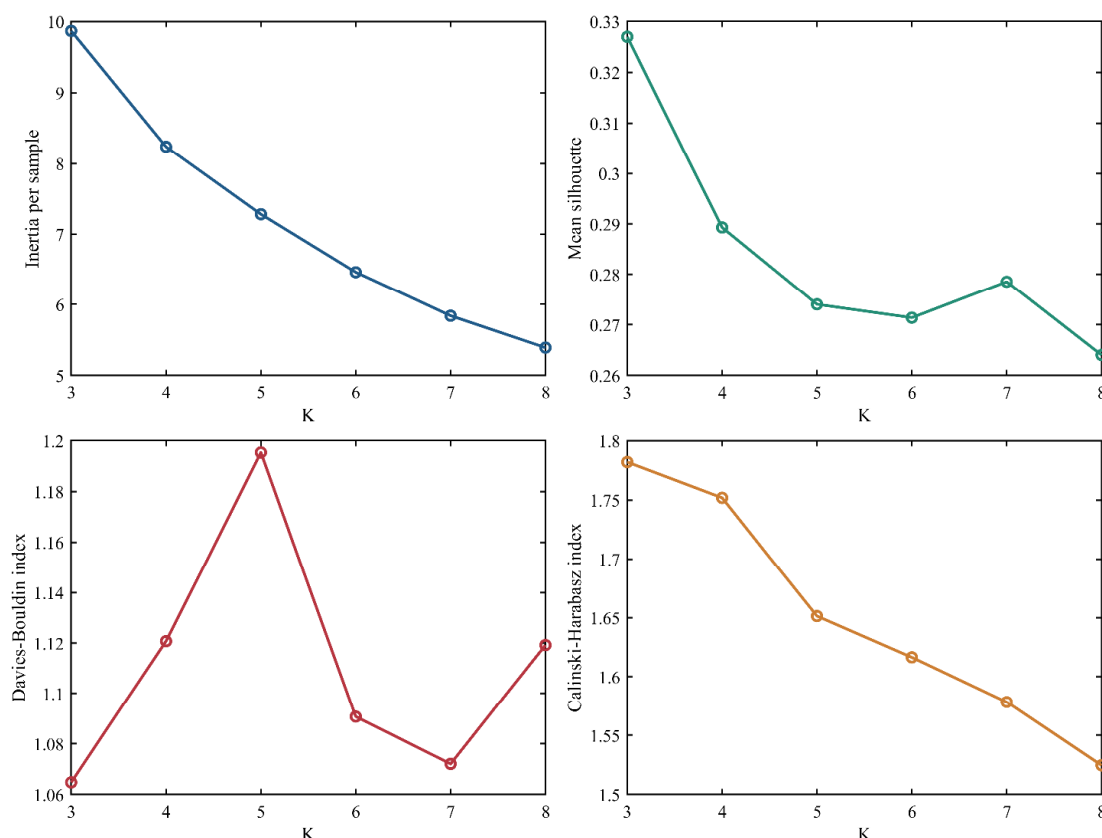


Figure 2. Standard clustering metrics used to support the selection of K = 5.

Figure 3 is used to test whether the ordered states follow the basic understanding that production progresses from early to late. In the figure, nodes represent the five states S1–S5, links represent state transitions between adjacent stages of the same well, and the numbers on the links denote transition counts. The main path is concentrated along S1 → S2 → S3 → S4 → S5, indicating that the clustering result is not a disordered grouping but can recover the direction of gradual evolution in pressure propagation and supply structure during production. A small number of cross-level transitions reflect differences among wells in reserve scale, rate schedule, and boundary-propagation speed.



Figure 3. Ordered-state transition network.

3.3. Physical Consistency Between Ordered States and Typical RTA Slopes

Judging the reasonableness of the identification results based only on clustering separation and state-transition order is insufficient. To further verify whether S1–S5 have RTA physical meaning, the log–log slope $\beta_{\log}$ between RNP and tmb for each well in different ordered states was extracted and compared with typical RTA slopes. Each scatter point represents the $\beta_{\log}$ value of one well in an ordered state, the median point represents the overall representative value of that state, and the vertical line denotes the interquartile range. The typical reference lines include the 1/4 slope commonly used for bilinear flow, the 1/2 slope commonly used for linear flow, and the unit-slope response.

Figure 4 shows that the median $\beta_{\log}$ value of S1 is approximately 1.01 and the distribution is relatively concentrated, indicating that the earliest windows mainly show a rapid transient response after production start-up. This state may include wellbore storage, rapid fracture response, and strong initial drawdown effects. The median $\beta_{\log}$ values of S2 and S3 are approximately 0.37 and 0.41, respectively; both are mainly distributed between 1/4 and 1/2, indicating that the early-to-middle states exhibit a transition from fracture–stimulated-region bilinear flow toward stimulated-region linear supply. The median $\beta_{\log}$ of S4 is approximately 0.71, clearly higher than 1/2, suggesting that the response has deviated from pure linear supply and that boundary propagation, control-volume effects, or constrained matrix replenishment begin to intensify. The median $\beta_{\log}$ of S5 is approximately 0.66 and occurs in the late stage of long-term production; it is better interpreted as a late response jointly controlled by low-pressure property changes, adsorption/desorption, compressibility, and long-tail production.

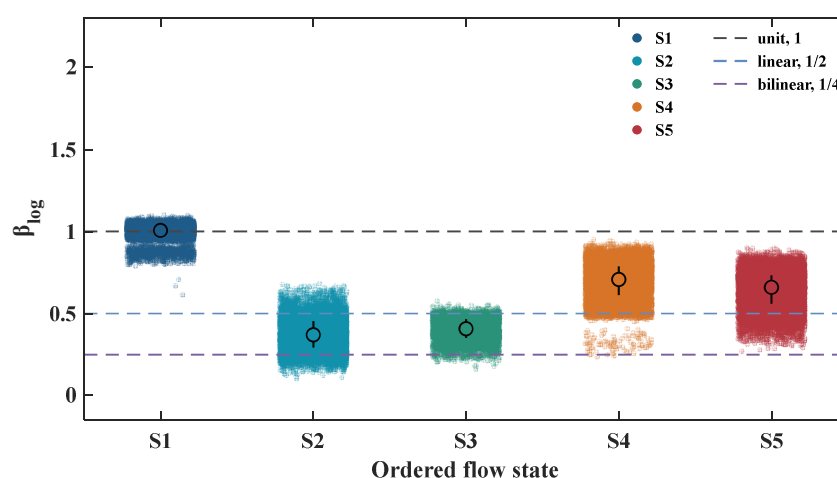


Figure 4. Distribution of ordered flow states in the RTA diagnostic phase space.

Therefore, S1–S5 are better related to theoretical flow regimes by interval correspondence rather than by one-to-one correspondence. S1 mainly corresponds to the earliest rapid response after the F1–F2 combination; S2 corresponds to the transition from early fracture–stimulated-region bilinear response to the development of stimulated-region supply; S3 is closer to stimulated-region linear supply and matrix-replenishment transition; S4 corresponds to matrix composite supply and enhanced boundary/control-volume effects; and S5 corresponds to late low-pressure property response and long-tail production. The slope distributions show that the ordered states obtained by unsupervised clustering are not purely mathematical groups, but have good physical consistency with typical RTA flow-regime slopes.

Table 4 indicates that the clustering states have a continuous correspondence with theoretical flow regimes. Early states are not simply merged into one class, but are separated into S1 and S2 in data feature space. The middle state S3 reflects expansion of the supply range, whereas the mid-to-late states S4 and S5 are controlled more strongly by boundary propagation, control volume, low-pressure properties, and long-term decline. This result is consistent with the understanding of continuous multi-mechanism evolution in unconventional gas wells.

Table 4. Dynamic features and physical interpretation of the five ordered states.

State	Median Time/d	Median $\beta_{\log}$	Main Dynamic Meaning	Approximate Relation to Theoretical Flow Regimes
S1	2.5	ca. 1.01	Rapid transient response at production start-up	Earliest part of F1–F2
S2	10.5	ca. 0.37	Development of fracture–stimulated-region supply	Late F1–F2 to early F3
S3	160	ca. 0.41	Transition from stimulated-region supply to matrix replenishment	F3 to F4 transition
S4	480	ca. 0.71	Enhanced matrix composite supply and boundary/control-volume influence	F4 to F5 transition
S5	3190	ca. 0.66	Late low-pressure property response and long-tail production	Mainly F6

The correspondence in Table 4 should be understood as an interval-type physical interpretation rather than a strict one-to-one mapping between a cluster and an ideal analytical flow regime. This treatment is necessary because real production windows often contain transitional responses. To reduce ambiguity, the interpretation combines material-balance time, RNP level, local log–log slope, pseudo-pressure derivative, gas-production contribution, and the position of the state in the ordered sequence. In the $\log(\text{RNP})$ – $\log(\text{tmb})$ diagnostic space, bilinear flow is commonly associated with a slope close to 1/4, linear-flow response with a slope close to 1/2, and closed-boundary or pseudo-steady-state response with a slope approaching 1. The identified states are therefore interpreted according to their relative position with respect to these slope references and their production-time sequence, rather than by slope alone (Table 5).

Table 5. RTA slope references used for physical interpretation of ordered states.

Theoretical Response	Typical Slope in $\log(\text{RNP})$ – $\log(\text{tmb})$	Diagnostic Implication	Relation to Ordered States
Bilinear flow	Approximately 1/4	Early coupled fracture–stimulated-region response	Reference for S2 transition
Linear flow	Approximately 1/2	Fracture or stimulated-region linear supply	Reference for S2–S3
Boundary-controlled or pseudo-steady response	Approaches 1	Pressure propagation reaches limited control volume	Reference for S4–S5 strengthening
Low-pressure long-tail response	Usually transitional, affected by adsorption and compressibility	Non-ideal late response controlled by pressure-dependent properties	Mainly used to interpret S5

3.4. Dynamic Features, Gas-Production Contributions, and Durations of the Five States

After confirming the physical consistency of the state sequence and slope distributions, the development significance of each state is further analyzed from three aspects: relative intensity of dynamic features, gas-production contribution, and duration. Relative dynamic-feature intensity is used to judge curve-shape differences among states, gas-production contribution is used to identify the main contributing stages, and duration distribution reflects state stability and inter-well differences.

Figure 5 shows the relative intensity of the five ordered flow states in terms of the main dynamic features. The columns correspond to RNP, $\beta\sqrt{t_{mb}}$, $\beta_{\log}$, D_p , q_g , and $G_p/G_{p,max}$. The values of each feature are normalized to 0–1 among S1–S5, so the color intensity indicates relative strength among states rather than the magnitude of the original physical quantity. The results show that S1 and S2 mainly reflect early rapid drawdown, higher gas rate, and near-well response; S3 represents the transition from stimulated-region supply to matrix replenishment; S4 is enhanced in slope and derivative features, indicating that the mid-to-late supply structure begins to change; and S5 is more prominent in re-

covery degree and long-term drawdown response, reflecting low-pressure long-tail production.

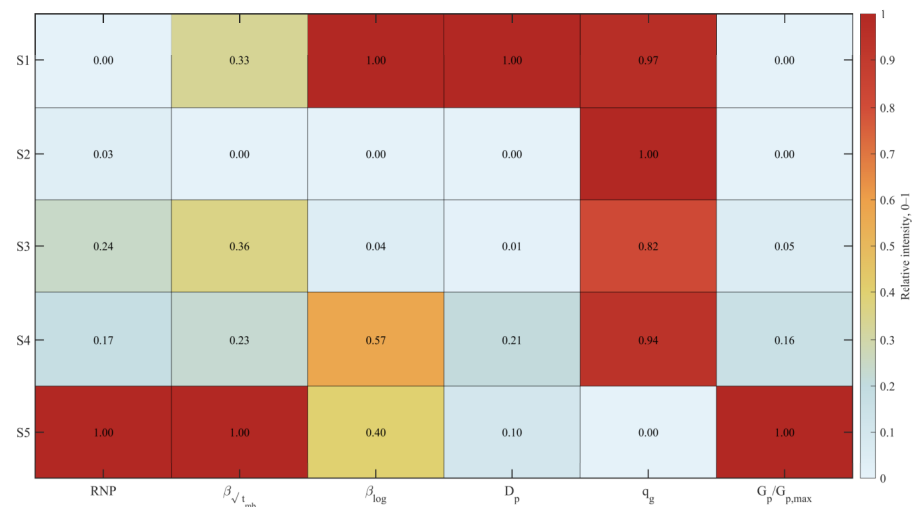


Figure 5. Relative intensity of dynamic features for the five ordered flow states.

After identifying the ordered S1–S5 flow states, it is necessary to further evaluate whether these states have production significance. For this purpose, they are verified from three aspects: gas-production contribution, stage duration, and feature-discrimination ability. Gas-production contribution is used to identify the main gas-producing stages, duration distribution reflects the stability and inter-well variability of different states, and feature-contribution ranking explains the main dynamic basis for state division.

In Figure 6, the bars represent the gas-production contribution ratio corresponding to each state, and the line represents the median duration of continuous state segments. The results show that S1 and S2 have relatively low gas-production contributions and mainly support early dynamic identification. S3 represents the gradual establishment of the early-to-middle supply system but contributes only a limited proportion of life-cycle cumulative gas production. S4 and S5 contribute significantly more, indicating that mid-to-late supply decline and low-pressure response have important effects on cumulative gas production under geological reserves closure and long-term production constraints. S5 has the highest cumulative gas-production contribution mainly because late long-tail production lasts a long time, not because its instantaneous productivity is highest.

Figure 7 shows the distribution of stage duration for different states. S1 and S2 have short durations, indicating that the earliest fracture and stimulated-region responses transition quickly in daily data. S3 begins to show clear inter-well differences, reflecting differences in stimulated-region supply capacity and matrix-replenishment conditions. S4 and S5 last longer and have broader distributions, indicating that after wells enter the mid-to-late stage, boundary propagation, control volume, reserve scale, and low-pressure properties have increasingly significant effects on production dynamics.

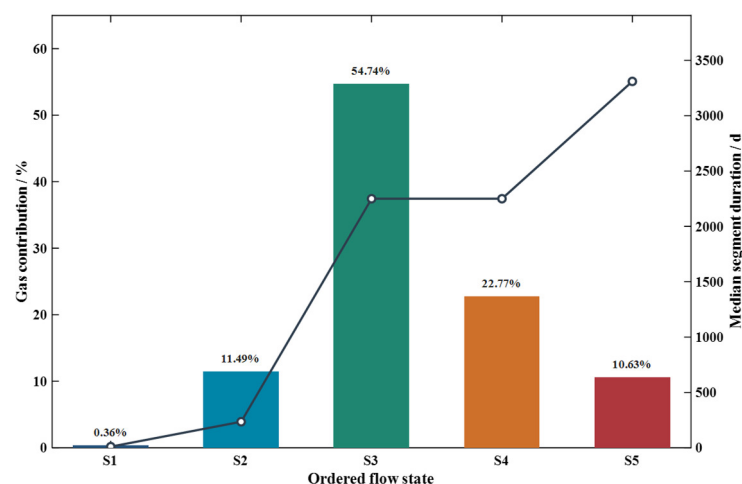


Figure 6. Comparison of gas-production contribution and duration of each ordered flow state.

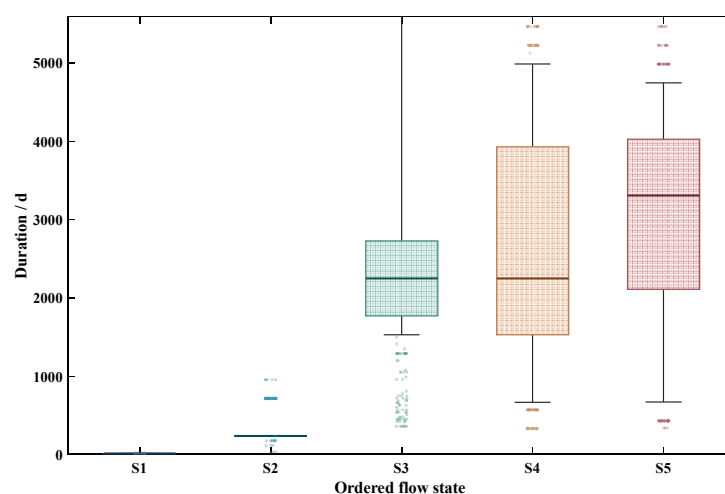


Figure 7. Duration distribution of each ordered flow state.

3.5. Discriminatory Contribution of Dynamic Features and Development Interpretation

To determine which dynamic features play the main roles in ordered-state division, the between-group variance ratio is used to evaluate the degree of difference for each feature among S1–S5. This indicator is not the magnitude of the original feature value; rather, it measures how much of the feature variation comes from differences among states. A higher between-group variance ratio indicates that the feature can better distinguish different ordered states.

In Figure 8, the horizontal axis is the between-group variance ratio, which measures the degree of difference in a dynamic feature among the five ordered states. This metric is not the magnitude of the original physical quantity, but the discriminatory ability of the feature for state division. Let $x_{i,j}$ denote the value of the j th feature in the i th window, n_s denote the sample size of state s , $\bar{x}_{s,j}$ denote the within-state mean, and $\bar{x}_j$ denote the mean of all windows. The between-group variance ratio is defined as

$$\eta_j = \frac{\sum_{s=1}^K n_s (\bar{x}_{s,j} - \bar{x}_j)^2}{\sum_{i=1}^N (x_{i,j} - \bar{x}_j)^2} \quad (29)$$

where η_j is the between-group variance ratio of the j th dynamic feature, K is the number of ordered states ($K = 5$ in this study), and N is the total number of sliding windows. As η_j approaches 1, the feature differences mainly originate from differences among states and the feature contributes strongly to state discrimination. As η_j approaches 0, the feature mainly reflects within-state fluctuation and plays a weaker role in distinguishing S1–S5.

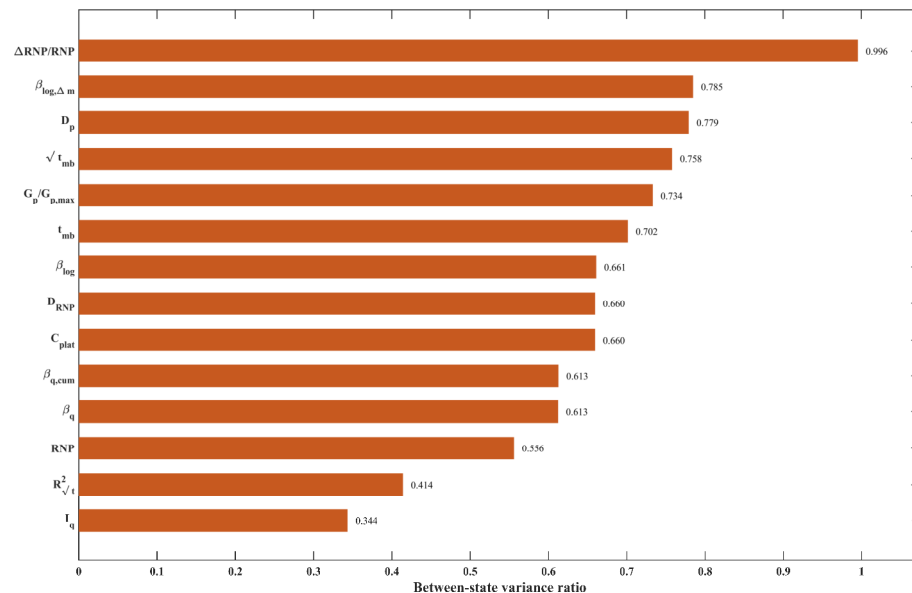


Figure 8. Ranking of the contributions of dynamic features to ordered-state discrimination.

All features in the figure are calculated from RTA dynamic curves within sliding windows. RNP is the median normalized pseudo-pressure response within a window and characterizes the pseudo-pressure drawdown level of that window. $\Delta RNP/RNP$ is the relative change in RNP within the window and characterizes the evolution intensity of pseudo-pressure response over a window:

$$RNP_{i,r} = \text{median}_{j \in W_{i,r}}(RNP_j), \quad \frac{\Delta RNP}{RNP} = \frac{RNP_e - RNP_s}{RNP} \quad (30)$$

where $W_{i,r}$ is the r th sliding window of the i th well, and RNP_s , RNP_e , and RNP_c are the RNP values at the start, end, and center of the window, respectively. A larger value indicates greater pseudo-pressure-response change within the window and usually corresponds to flow-state transition, supply-capacity change, or enhanced late-stage drawdown accumulation.

$R^2_{\log}$ and $R^2_{\sqrt{t_{mb}}}$ are local goodness-of-fit indicators used to judge whether the window curve is close to a log–log linear relationship or a square-root-time linear relationship:

$$R^2_{\log} = 1 - \frac{\sum_{j \in W_{i,r}} [y_j - \hat{y}_j]^2}{\sum_{j \in W_{i,r}} [y_j - \bar{y}]^2}, \quad y_j = \log(RNP_j), \quad x_j = \log(t_{mb,j}) \quad (31)$$

$$R^2_{\sqrt{t}} = 1 - \frac{\sum_{j \in W_{i,r}} [RNP_j - \widehat{RNP}_j]^2}{\sum_{j \in W_{i,r}} [RNP_j - \overline{RNP}]^2}, \quad x_j = \sqrt{t_{mb,j}}$$

These two indicators mainly evaluate whether the local window curve shape is stable. A low between-group variance ratio does not mean that they are useless; rather, it indicates that they mainly serve as quality constraints and auxiliary shape indicators, and

that their direct discrimination ability for the five states is weaker than that of core features such as drawdown variation, recovery degree, local slope, and pseudopressure derivative.

Figure 8 shows that $\Delta RNP/RNP$, β_{log} , D_p , $\beta\sqrt{t_{mb}}$, $G_p/G_{p,max}$, and t_{mb} have relatively high between-group variance ratios, indicating that state identification is jointly controlled by pseudopressure-response variation, local slope, pseudopressure derivative, and production-stage progression. R^2 -type goodness-of-fit indicators contribute less; this does not mean that they are meaningless, but shows that they are more useful for judging whether the curve shape within a window is stable and has a weaker direct contribution to state discrimination than drawdown-change and stage-progression features.

From the perspective of development interpretation, the value of ordered-state identification does not lie in assigning a fixed well-type label to an entire well, but in determining which mechanisms dominate a well during different production intervals. If S2 and S3 last for short times, early-to-middle effective supply is usually insufficient, and attention should be paid to fracture conductivity, stimulated-region width, and near-well supply conditions. If S4 appears early or lasts for a long time, matrix composite supply, boundary propagation, or control-volume constraints have begun to dominate production dynamics. If S5 occupies a high proportion, late low-pressure response and long-tail production contribute greatly to cumulative gas production, and adsorption/desorption, compressibility, water-production changes, and pressure-control schedules should be further analyzed.

Compared with whole-well clustering, ordered-state identification based on sliding windows is more suitable for describing stage evolution in unconventional gas wells. A well may show fracture–stimulated-region response in the early stage, stimulated-region and matrix replenishment in the middle stage, and boundary influence plus low-pressure property control in the late stage. The evaluation focus shifts from which class a well belongs to, to which stage lasts the longest, which stage contributes the most gas, and which stage ends too early. Such results are more useful for fracturing-parameter optimization, well-spacing evaluation, production-schedule adjustment, and dynamic diagnosis of real wells.

3.6. Dynamic-State Identification and Development Interpretation of Well M1

To verify the applicability of the ordered flow-state identification method to actual production data, Well M1 is selected for dynamic-state identification. The dataset includes daily gas rate and pressure data, and the original reservoir pressure is set to 35 MPa. Because the raw data do not provide a complete average reservoir-pressure sequence, measured pressure is used in the analysis to construct pseudopressure drawdown and RNP. Time is recorded in the sequence of production data as day 1, day 2, and so on; cumulative gas production is obtained by integrating the daily gas rate; and the same PVT table as described above is used.

The real-well application is used as an engineering validation of the ordered-state workflow. Independent high-fidelity numerical simulations with more complex fracture networks would provide an additional benchmark, but such simulations are outside the main purpose of this work. The objective here is to establish a unified and reproducible state-identification framework constrained by a five-region analytical model. Field production data contain operational changes, measurement uncertainty, multiphase effects, and geological heterogeneity, and therefore provide a practical test of whether the ordered-state sequence can extract interpretable dynamic stages from real production behavior.

Well M1 contains 874 d of valid production data. Pressure decreases from 33.87 MPa to 4.56 MPa, with the minimum pressure approximately 3.35 MPa. The maximum single-stage daily gas rate is approximately 5285 m³/d, the late-stage rate is approximately 1387

m³/d, and cumulative gas production is approximately 2.03×10^6 m³. Overall, the well shows dynamic characteristics of rapid early drawdown, sustained mid-stage supply, and late low-pressure long-tail production.

Figure 9 shows that pressure declines rapidly during the initial production of Well M1 and that the daily gas rate fluctuates significantly, indicating that early production is jointly affected by the opening schedule, near-well pressure release, and rapid fracture-system response. As production proceeds, the rate of pressure decline gradually slows while the daily gas rate remains at a certain level, suggesting that the stimulated reservoir volume and matrix supply can still maintain effective replenishment during the middle stage. In the late stage, pressure remains in a low range and the daily gas rate decreases overall with fluctuations, indicating that the well gradually enters low-pressure decline and long-tail production.

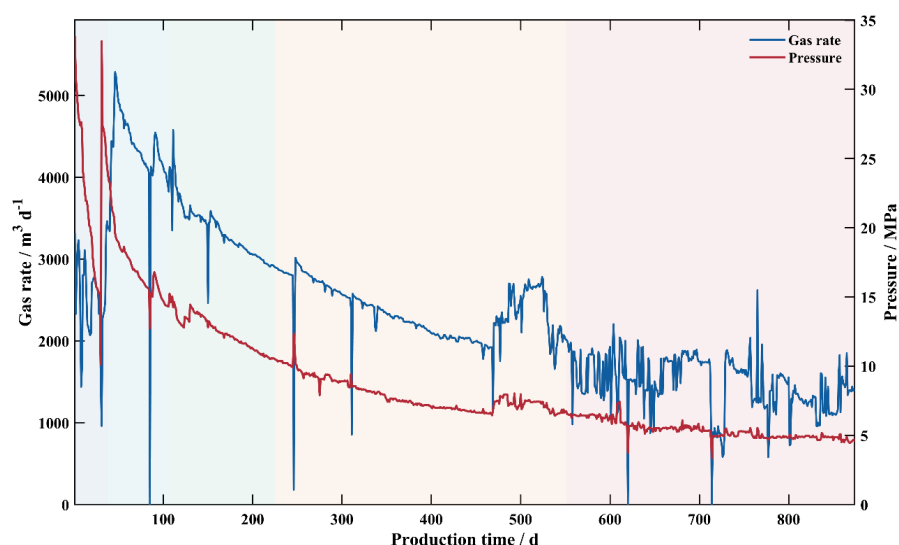


Figure 9. Production dynamics and ordered-state division of Well M1.

Based on the sliding-window RTA feature identification results, Well M1 can be divided into five ordered states, S1–S5. The duration, pressure change, average daily gas rate, and gas-production contribution of each state are listed in Table 6.

Table 6. Ordered flow-state identification results and development interpretation for Well M1.

State	Time Range/d	Pressure Change/MPa	Average Daily Gas Rate/m ³ ·d ^{−1}	Gas Contribution/%	Dynamic Interpretation
S1	1–39	33.87–23.40	2540	4.59	Rapid initial drawdown; strong near-well and fracture response; obvious production fluctuations
S2	39–105	23.40–14.33	4311	14.06	Enhanced early fracture–stimulated-region response; high rate but rapid pressure decline
S3	105–225	14.33–10.50	3344	19.78	Stimulated-region supply and matrix replenishment begin to develop stably
S4	225–551	10.50–6.74	2355	37.83	Main supply stage with long duration and the largest gas contribution
S5	551–874	6.74–4.56	1479	23.47	Late low-pressure production; declining rate but evident long-tail contribution

Figure 10 shows that different states have clearly distinguishable distributions on the RNP-tmb curve. S1 is mainly located in the low material-balance-time interval and has a scattered point distribution, indicating that the earliest production stage is strongly affected by the production schedule and rapid pressure release and should not be over-

divided into stable flow regimes. S2 and S3 move upward and to the right along a relatively stable RNP-tmb curve, and their median $\beta_{\log}$ values are approximately 0.41–0.46, close to the common 1/2 slope range for linear flow, indicating a clear early-to-middle supply response. After S4, the curve gradually becomes steeper and the median $\beta_{\log}$ is approximately 0.76, indicating that RNP increases more rapidly with material-balance-time progression, drawdown consumption increases, and supply capacity begins to weaken. The median $\beta_{\log}$ of S5 is approximately 0.96, close to unit slope, indicating that the well enters a stronger decline or low-pressure response stage in the late period.

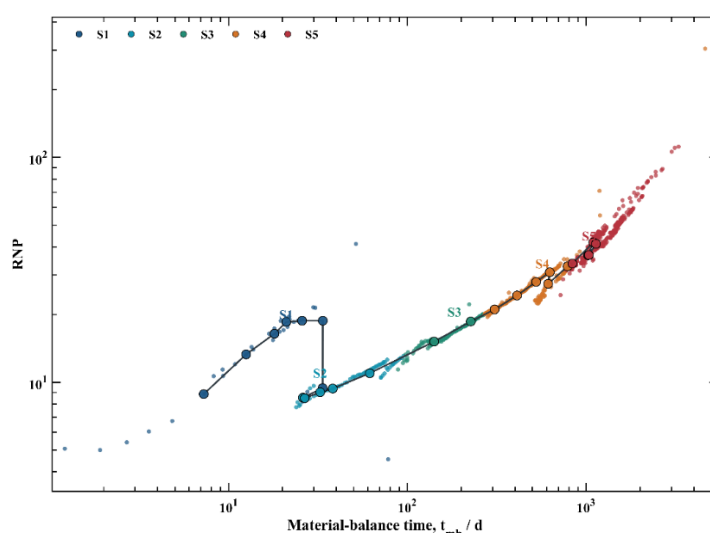


Figure 10. RNP-tmb diagnostic curve and ordered-state comparison for Well M1.

Figure 11 shows that the gas-production contribution of Well M1 is mainly concentrated in S3, S4, and S5. S4 has the highest contribution, 37.83%, indicating that mid-to-late stimulated-region–matrix composite supply is the key contributor to production. Although the average daily gas rate of S5 has decreased, it still contributes 23.47% of cumulative gas production, indicating that late low-pressure long-tail production remains important for ultimate recovery. S1 and S2 together contribute approximately 18.65%; despite high early gas rates, their short durations limit their contributions to cumulative production.

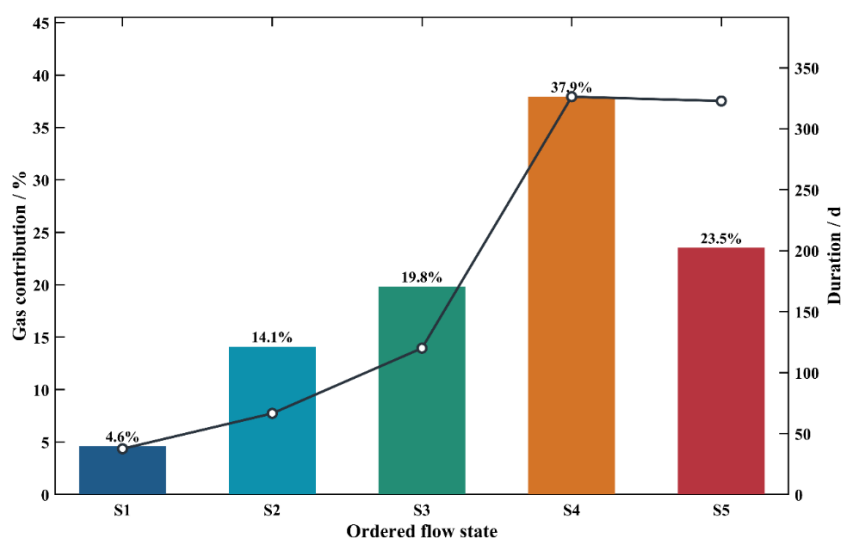


Figure 11. Gas-production contribution and duration of each ordered state for Well M1.

In summary, the development characteristics of Well M1 can be described as rapid early drawdown, sustained mid-stage supply, and significant late low-pressure long-tail production. From a development perspective, S3–S4 are the most important effective-supply stages for this well, and evaluation should focus on stimulated-region supply capacity, matrix-replenishment strength, and pressure-propagation range. After entering S5, the gas-rate decline is jointly controlled by low-pressure conditions, supply attenuation, and production-schedule fluctuations. Late-stage measures should therefore focus on stable-pressure production, reduction in ineffective drawdown, stable liquid unloading, and extension of long-tail gas production. This example shows that the ordered-state identification method based on the five-region model and RTA window features can divide single-well production history into physically meaningful dynamic stages and provide a quantitative basis for real-well production evaluation and late-stage schedule adjustment.

The proposed workflow is also intended to reduce uncertainty caused by the manual selection of RTA stage boundaries. In manual interpretation, shifting the selected boundary by several days may have a limited influence on qualitative flow-regime description, but it can change the slope of the selected RTA segment and therefore affect estimates of parameters such as $A \sqrt{k}$, apparent permeability-related indicators, and subsequent EUR prediction. The sliding-window clustering workflow provides a consistent criterion for boundary identification and makes the resulting stage duration and gas-contribution statistics more comparable among wells.

A formal blind comparison with several independent experts is not included in this study. This limitation should be recognized because manual boundary selection is affected by interpreter experience, smoothing scale, and the diagnostic plot used. The present real-well example is therefore used to demonstrate engineering applicability rather than to quantify the full reduction in human interpretation variability. In future work, the same production dataset should be interpreted independently by several experts, and the spread of boundary time, slope, $A \sqrt{k}$ -related indicators, and EUR estimates should be compared with the ordered-state results.

4. Conclusions

An RTA-ordered flow-state identification method constrained by the five-region analytical model is established. The method does not use machine learning to predict pressure or rate directly. Instead, it uses the five-region model to generate dynamic responses satisfying material balance and seepage mechanisms, and then identifies local flow states through RTA window features. This approach preserves the physical interpretability of the analytical model and reduces the subjective differences associated with manual flow-regime division.

The geological reserves closure sample library provides effective support for ordered-state identification. Based on 10,000 fully random synthetic wells and 689,394 sliding-window samples, K-means clustering identifies five ordered states. The inter-cluster separation is 1.8924 and the stage-regression rate is only 0.0238, showing that the results have good separability in feature space and strong consistency in the direction of temporal progression. Each well experiences an average of 4.24 states, indicating that most samples cover a relatively complete production-evolution process.

The five ordered states correspond clearly to typical flow stages of unconventional gas wells. S1 reflects rapid fracture–stimulated-region response immediately after production start-up; S2 corresponds to the gradual development of fracture and stimulated-region supply; S3 represents the transition from stimulated-region supply to matrix composite supply; S4 reflects enhanced boundary or control-volume effects; and S5 represents low-pressure property response and long-tail decline after long-term production. Local slope, pseudopressure derivative, RNP variation, and normalized cumulative gas produc-

tion jointly control state discrimination, among which RNP variation, normalized cumulative gas production, and local slope contribute strongly to state identification.

Application to real Well M1 verifies the engineering interpretability of the method. The initial reservoir pressure of this well is 35 MPa, and the bottom-hole pressure decreases from 33.87 MPa to 4.56 MPa during production. The identification results show that its production history can be divided into five continuous stages, S1–S5. S4 contributes the most gas, reaching 37.83%, and is the main effective-supply stage of the well. S5 contributes 23.47%, indicating that late production still has some low-pressure long-tail gas-production capacity, although production efficiency per unit drawdown decreases. These results can be used to judge the duration of the stable-production stage, identify the onset of supply decline, and support late pressure-controlled production and drainage-schedule optimization.

The method still depends on production data quality, PVT parameters, and the accuracy of bottom-hole pressure data. Future work can incorporate more real-well validation samples, fracturing-operation parameters, geological sweet-spot evaluation, and water-production dynamics to improve state-interpretation rules. State-transition comparison methods among different blocks can also be developed so that ordered flow-state identification can better support inter-well comparison, development adjustment, and evaluation of remaining supply potential.

The influence of PVT and initial-pressure uncertainty should also be recognized. Gas viscosity, gas compressibility, the compressibility factor, and initial pressure affect the calculation of pseudo-pressure, material-balance time, and pseudo-pressure derivative, and may therefore shift some window features and stage boundaries. For wells in the same block, these properties are usually less variable than production schedules, pressure data quality, and boundary selection; nevertheless, they are not completely negligible. Future sensitivity tests can perturb initial pressure, gas viscosity, and gas compressibility within practical uncertainty ranges and evaluate state-label consistency, boundary-time shifts, and changes in S4–S5 gas-contribution ratios.

Independent numerical simulation is also needed for broader validation. The present workflow is constrained by the five-region analytical model, so it is most reliable for flow behavior that can be represented by composite fracture, stimulated-region, matrix, boundary, and low-pressure property responses. Complex fracture networks, strong fracture interference, multiphase transport, and geomechanical effects may generate responses beyond the current analytical sample space. Future validation should use high-fidelity numerical models with complex fracture geometry and multiphysics transport, and should compare their simulated state sequences with the ordered-state framework proposed here.

Author Contributions: Conceptualization, H.Y., Y.S. and W.X.; methodology, H.Y.; software, H.Y.; validation, H.Y., Y.G. and Z.T.; formal analysis, H.Y. and Y.L.; investigation, H.Y., D.W., Y.G., Y.L. and M.S.; data curation, H.Y., Y.G. and Z.T.; writing—original draft, H.Y.; writing—review and editing, Y.S., W.X., D.W., Y.G., Y.L., M.S. and Z.T.; visualization, Y.L. and Z.T.; supervision, Y.S. and W.X.; project administration, Y.S. and W.X.; funding acquisition, Y.S. and W.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Oil and Gas Major Project, Grant No. 2025ZD1404103; the CNPC's Basic and Foresight-Oriented Science and Technology Program, project "Mechanisms of Shale Oil and Gas Development and Volumetric Development Technology", Grant No. 2023ZZ08; and PetroChina Company Limited, project "Key Technology Research and Demonstration of the 'Ballast Stone Project' for Mature Gas Fields", Grant No. 2023YQX10301.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors gratefully acknowledge the support of the National Energy Shale Gas R&D Center for this work.

Conflicts of Interest: Author Deshang Wang was employed by the Chuannan Exploration and Development Division, PetroChina Jilin Oilfield Company, PetroChina Company Limited. Author Yu-zheng Gong was employed by the PetroChina Tarim Oilfield Company. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The authors declare that this study received funding from the Oil and Gas Major Project, the CNPC's Basic and Foresight-Oriented Science and Technology Program, and PetroChina Company Limited. The funders were not involved in the study design; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Wei, S.-M.; Xia, Y.; Jin, Y.; Guo, X.-Y.; Zi, J.-Y.; Qiu, K.-X.; Chen, S.-Y. Production induced fracture closure of deep shale gas well under thermo-hydro-mechanical conditions. *Pet. Sci.* **2024**, *21*, 1796–1813. <https://doi.org/10.1016/j.petsci.2023.12.010>.
2. Luo, A.; Li, Y.; Wu, L.; Peng, Y.; Tang, W. Fractured horizontal well productivity model for shale gas considering stress sensitivity, hydraulic fracture azimuth, and interference between fractures. *Nat. Gas Ind. B* **2021**, *8*, 278–286. <https://doi.org/10.1016/j.ngib.2021.04.008>.
3. Zhang, L.; Gao, J.; Hu, S.; Guo, J.; Liu, Q. Five-region flow model for MFHWs in dual porous shale gas reservoirs. *J. Nat. Gas Sci. Eng.* **2016**, *33*, 1316–1323. <https://doi.org/10.1016/j.jngse.2016.05.063>.
4. Zhang, Q.; Su, Y.; Zhang, M.; Wang, W. A multi-linear flow model for multistage fractured horizontal wells in shale reservoirs. *J. Pet. Explor. Prod. Technol.* **2017**, *7*, 747–758. <https://doi.org/10.1007/s13202-016-0284-0>.
5. Cui, Q.; Zhao, Y.; Zhang, L.; Chen, M.; Gao, S.; Chen, Z. A semianalytical model of fractured horizontal well with hydraulic fracture network in shale gas reservoir for pressure transient analysis. *Adv. Geo-Energy Res.* **2023**, *8*, 193–205. <https://doi.org/10.46690/ager.2023.06.06>.
6. Xu, Y.; Liu, X.; Hu, Z.; Duan, X.; Chang, J. Bottom-hole pressure drawdown management of fractured horizontal wells in shale gas reservoirs using a semi-analytical model. *Sci. Rep.* **2022**, *12*, 22490. <https://doi.org/10.1038/s41598-022-26978-2>.
7. Anderson, D.M.; Nobakht, M.; Moghadam, S.; Mattar, L. Analysis of production data from fractured shale gas wells. In *Proceedings of the SPE Unconventional Resources Conference/Gas Technology Symposium*; SPE-131787-MS; SPE: Houston, TX, USA, 2010. <https://doi.org/10.2118/131787-MS>.
8. Clarkson, C.R. Production data analysis of unconventional gas wells: Review of theory and best practices. *Int. J. Coal Geol.* **2013**, *109–110*, 101–146. <https://doi.org/10.1016/j.coal.2013.01.002>.
9. Wattenbarger, R.A.; El-Banbi, A.H.; Villegas, M.E.; Maggard, J.B. Production analysis of linear flow into fractured tight gas wells. In *Proceedings of the SPE Rocky Mountain Petroleum Technology Conference/Low-Permeability Reservoirs Symposium*; SPE-39931-MS; SPE: Houston, TX, USA, 1998. <https://doi.org/10.2118/39931-MS>.
10. Nobakht, M.; Clarkson, C. Analysis of production data in shale gas reservoirs: Rigorous corrections for fluid and flow properties. In *Proceedings of the SPE Eastern Regional Meeting*; SPE-149404-MS; SPE: Houston, TX, USA, 2011. <https://doi.org/10.2118/149404-MS>.
11. Yuan, B.; Clarkson, C.R.; Zhang, Z.; Zhu, X. Deviations from transient linear flow behavior: A systematic investigation of possible controls on abnormal reservoir signatures. *J. Pet. Sci. Eng.* **2021**, *205*, 108910. <https://doi.org/10.1016/j.petrol.2021.108910>.
12. Pershin, I.; Pervukhin, D.; Ilyushin, Y.; Afanaseva, O. Design of distributed systems of hydrolithosphere processes management. Selection of optimal number of extracting wells. *IOP Conf. Ser. Earth Environ. Sci.* **2017**, *87*, 032030. <https://doi.org/10.1088/1755-1315/87/3/032030>.
13. Brown, M.; Ozkan, E.; Raghavan, R.; Kazemi, H. Practical solutions for pressure-transient responses of fractured horizontal wells in unconventional shale reservoirs. *SPE Reserv. Eval. Eng.* **2011**, *14*, 663–676. <https://doi.org/10.2118/125043-PA>.
14. Stalgorova, E.; Mattar, L. Practical analytical model to simulate production of horizontal wells with branch fractures. In *Proceedings of the SPE Canada Unconventional Resources Conference*; SPE-162515-MS; SPE: Houston, TX, USA, 2012. <https://doi.org/10.2118/162515-MS>.

15. Stalgorova, E.; Mattar, L. Analytical model for unconventional multifractured composite systems. *SPE Reserv. Eval. Eng.* **2013**, *16*, 246–256. <https://doi.org/10.2118/162516-PA>.
16. Zhao, Y.; Lu, G.; Zhang, L.; Wei, Y.; Guo, J.; Chang, C. Numerical simulation of shale gas reservoirs considering discrete fracture network using a coupled multiple transport mechanisms and geomechanics model. *J. Pet. Sci. Eng.* **2020**, *195*, 107588. <https://doi.org/10.1016/j.petrol.2020.107588>.
17. Li, W.; Liu, J.; Zeng, J.; Leong, Y.-K.; Elsworth, D.; Tian, J.; Li, L. A fully coupled multidomain and multiphysics model for evaluation of shale gas extraction. *Fuel* **2020**, *278*, 118214. <https://doi.org/10.1016/j.fuel.2020.118214>.
18. Sarkar, P.; Yoon, S.; Kim, J.; Baek, S.; Sun, A.; Yoon, H. Surrogate models for development of unconventional shale reservoirs by an integrated numerical approach of hydraulic fracturing, flow and geomechanics, and machine learning. *Geomech. Energy Environ.* **2025**, *43*, 100691. <https://doi.org/10.1016/j.gete.2025.100691>.
19. Wang, H.; Qiao, L.; Lu, S.; Chen, F.; Fang, Z.; He, X.; Zhang, J.; He, T. A novel shale gas production prediction model based on machine learning and its application in optimization of multistage fractured horizontal wells. *Front. Earth Sci.* **2021**, *9*, 726537. <https://doi.org/10.3389/feart.2021.726537>.
20. He, Y.-W.; He, Z.-Y.; Tang, Y.; Xu, Y.-J.; Long, J.-C.; Sepehrnoori, K. Shale gas production evaluation framework based on data-driven models. *Pet. Sci.* **2023**, *20*, 1659–1675. <https://doi.org/10.1016/j.petsci.2022.12.003>.
21. Niu, W.; Lu, J.; Sun, Y.; Liu, H.; Cao, X.; Zhan, H.; Zhang, J. A review of the application of data-driven technology in shale gas production evaluation. *Energy Rep.* **2023**, *10*, 213–227. <https://doi.org/10.1016/j.egyr.2023.06.026>.
22. Niu, W.; Lu, J.; Sun, Y. Development of shale gas production prediction models based on machine learning using early data. *Energy Rep.* **2022**, *8*, 1229–1237. <https://doi.org/10.1016/j.egyr.2021.12.040>.
23. Gou, L.; Yang, Z.; Min, C.; Yi, D.; Li, X.; Kong, B. A novel domain adaptation method with physical constraints for shale gas production forecasting. *Appl. Energy* **2024**, *371*, 123673. <https://doi.org/10.1016/j.apenergy.2024.123673>.
24. Qi, N.; Li, X.; Wu, Z.; Wan, Y.; Wang, N.; Duan, G.; Wang, L.; Xiang, J.; Zhao, Y.; Zhan, H. Machine learning-based research for predicting shale gas well production. *Symmetry* **2024**, *16*, 600. <https://doi.org/10.3390/sym16050600>.
25. Pang, M.; Zhang, Z.; Zhou, Z.; Zhou, W.; Li, Q. Estimated ultimate recovery prediction of shale gas wells based on stacked integrated learning algorithm. *Sci. Rep.* **2025**, *15*, 19258. <https://doi.org/10.1038/s41598-025-03801-2>.
26. Rahmanifard, H.; Gates, I. A Comprehensive review of data-driven approaches for forecasting production from unconventional reservoirs: Best practices and future directions. *Artif. Intell. Rev.* **2024**, *57*, 213. <https://doi.org/10.1007/s10462-024-10865-5>.
27. Abdi, H.; Williams, L.J. Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 433–459. <https://doi.org/10.1002/wics.101>.
28. McQueen, J.B. Some methods of classification and analysis of multivariate observations. In Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability; University of California Press: Berkeley, CA, USA, 1967; pp. 281–297. Available online: <https://projecteuclid.org/ebooks/berkeley-symposium-on-mathematical-statistics-and-probability/Proceedings-of-the-Fifth-Berkeley-Symposium-on-Mathematical-Statistics-and/chapter/Some-methods-for-classification-and-analysis-of-multivariate-observations/bsmsp/1200512992> (accessed on 1 July 2026).
29. Kryvenko, S.; Gorditsa, M.; Siegel, G. Implementing the K-means clustering algorithm for the type well generation workflow in the Eagle Ford Shale. In Proceedings of the SPE/AAPG/SEG Unconventional Resources Technology Conference, Houston, TX, USA, 26–28 July 2021. <https://doi.org/10.15530/urtec-2021-5528>.
30. Chen, Z.; Yu, W.; Liang, J.-T.; Wang, S.; Liang, H.-C. Application of statistical machine learning clustering algorithms to improve EUR predictions using decline curve analysis in shale-gas reservoirs. *J. Pet. Sci. Eng.* **2022**, *208*, 109216. <https://doi.org/10.1016/j.petrol.2021.109216>.
31. Syed, F.I.; Muther, T.; Dahaghi, A.K.; Negahban, S. AI/ML assisted shale gas production performance evaluation. *J. Pet. Explor. Prod. Technol.* **2021**, *11*, 3509–3519. <https://doi.org/10.1007/s13202-021-01253-8>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.