processes

MDPI

Article

Engineering-Constrained Explainable Machine Learning for Development Policy Optimization in the Y284 Tight Oil Block, Ordos Basin

Huanying Yang [1,2], Xiaoping An [1,2], Youwei Xu [3], Yunli Lu [4,*], Gang Hui [4,*], Hongchang Li [1,2], Jing Wang [1,2], Xiaohui Fan [1,2] and Xueying Yu [1,2]

[1] Exploration and Development Research Institute of PetroChina Changqing Oilfield Company, Xi'an 710018, China; yanghy_cq@petrochina.com.cn (H.Y.); axp_cq@petrochina.com.cn (X.A.); lihc_cq@petrochina.com.cn (H.L.); wj920_cq@petrochina.com.cn (J.W.); fxhui_cq@petrochina.com.cn (X.F.); yxy2_cq@petrochina.com.cn (X.Y.)
[2] National Engineering Laboratory for Exploration and Development of Low-Permeability Oil&Gas Fields, Xi'an 710018, China
[3] No. 10 Oil Production Plant, PetroChina Changqing Oilfield Branch, Qingyang 745100, China; cq400042@163.com
[4] State Key Laboratory of Petroleum Resources and Engineering, China University of Petroleum (Beijing), Beijing 102249, China
* Correspondence: yuliyalu@163.com (Y.L.); hui.gang@cup.edu.cn (G.H.)

Abstract

Tight-oil performance is controlled by a coupled reservoir-engineering system: reservoir quality, fracturing design, well pattern, and operating policy jointly determine whether geological potential becomes sustained production. This study develops an engineering-constrained, explainable machine-learning workflow for the Y284 block of the Longdong Chang 6 reservoir and identifies the controls on five-year cumulative oil. Four constraints are imposed: pre-decision and post-hoc variables are separated to avoid leakage; decline fitting, water-breakthrough classification, and injection-effect diagnosis serve as physical constraints; models are selected by accuracy and engineering plausibility; and SHAP controls are turned into field-readable policies. Of 939 well records, 338 with a complete geological and engineering record form the model-ready sample; five-fold cross-validation stratified by well type gives an out-of-fold R^2 of 0.55 for the retained random forest. Engineering communication intensity, set by pump rate, stage count, and per-stage fluid and proppant volume, dominates production (69.9% of the mean absolute SHAP value), followed by geological storage capacity (21.6%) and injection and production control (8.6%); stage count and pump rate rank first. The framework is a block-scale screening and ranking tool, not a high-accuracy predictor.

Keywords: tight oil production decline; energy supplementation; well-network geometry; controlling factors; development-policy optimization; machine learning; explainable AI

Check for updates

Academic Editor: Dicho Stratiev

Received: 25 September 2026
Revised: 22 October 2026
Accepted: 23 October 2026
Published: 30 October 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) license.

1. Introduction

Tight oil reservoirs commonly show strong heterogeneity, small pore-throat systems, and low matrix permeability. In this setting, well performance is not determined by reservoir quality alone. A reservoir interval with relatively favorable porosity, permeability, and oil saturation may still show weak production if fracturing fails to connect an effective drainage volume, if pressure support is delayed, or if injected water rapidly advances through preferential pathways. Conversely, less favorable reservoir intervals may achieve

acceptable productivity when stimulation intensity, well spacing, injection allocation, and operating policy are well matched.

The Y284 block in the Longdong Chang 6 reservoir is a representative case of this coupled development problem. The block contains II/III-type tight-oil reservoirs and includes directional, highly deviated, horizontal, and vertical wells. Field performance shows high early productivity in some wells, rapid water response in part of the block, and large differences in decline behavior and cumulative production among well types and well patterns. These observations indicate that the key question is not whether a single geological or engineering variable is important, but under what reservoir-development conditions a variable becomes controlling.

Classical decline-curve analysis provides the necessary reservoir-engineering baseline. Arps exponential, hyperbolic, and harmonic forms remain widely used because they convert monthly production histories into interpretable indicators such as initial decline rate, b value, and first-year decline intensity [1]. For tight reservoirs, however, the fitted curve can be disturbed by shut-ins, stimulation measures, water breakthrough, pump adjustment, and injection-production policy changes. Therefore, decline fitting must be combined with data cleaning, event screening and engineering plausibility checks rather than used as a purely automatic regression exercise [2–4].

For tight and fracture-dominated wells, decline-curve analysis should be treated as a screened production diagnostic rather than an unconstrained reserve statement, because prolonged transient flow and completion-related heterogeneity can mimic sustained hyperbolic behavior [5,6].

Machine-learning methods provide a complementary way to learn nonlinear interactions from field data. Random forest, XGBoost, LightGBM, and CatBoost have been widely used in petroleum data analytics because they can process mixed numerical-categorical variables and capture feature interactions [7–10]. Recent applications span simulation-assisted history matching, geological-engineering integration, injection-production optimization, and data-driven productivity evaluation [11–15]. In unconventional reservoirs, these models have been used to integrate geological and operational factors into production forecasting and to explain the resulting productivity differences [16]. SHAP further provides a consistent framework for explaining model output [17–19]. Nevertheless, these algorithms are now common tools. The novelty of this study is therefore not the simple combination of tree models and SHAP, but the construction of a reservoir-engineering-constrained workflow that controls feature leakage, embeds field diagnosis into model screening, and converts SHAP results into policy classes that can be checked by reservoir engineers. Reviews of artificial-intelligence applications in oil and gas development reach the same conclusion: the decisive factor is not the choice of algorithm but the engineering constraints placed on the data, the validation design and the interpretation of the result [20].

The analytical design follows established ensemble-learning practice and is implemented with standard tree-ensemble tooling and statistical-learning theory [21,22], but strengthens that practice with an explicit engineering information boundary and interpretation checks for correlated predictors, following established guidance on interpretable machine learning [23] and recent SHAP-based control-factor interpretation of well productivity [24]. Three specific risks motivate this design. First, when the training sample is small, or the model is over-parameterized, the fitted structure is driven by noise rather than by the underlying process [25,26]. Second, the conventional ensemble and local-explanation tools used here—stochastic gradient boosting, random forests and their variable-importance measures, and additive local explanation—can be unstable when feature distributions, categorical structures and spatially related development variables are not controlled [27–32]. Third, a validation scheme that is not matched to the sampling structure of field data can

return optimistic performance even when the model does not transfer [33–37]. For the Y284 case, where neighbouring wells share a development stage, a well pattern and a local geological setting, these three risks are addressed by the grouped validation design described in Section 2.3 and by the engineering screening of the accepted model.

This paper treats Y284 as an independent case study and answers one practical question: how can reservoir condition, energy-supplementation mode, well pattern, well type, and engineering parameters be matched to improve five-year cumulative oil and development-policy design? The main contributions are threefold: (1) a Y284 data hierarchy is established, with explicit distinction among full-field records, complete-feature records and the model-ready sample drawn from the block itself, so that the field population and the modelling sample are never confused;(2) a pre-decision feature system, which is the second operation of the engineering-constrained definition given above, is proposed to separate variables available before or during early development from post-hoc diagnostic variables; and (3) the dominant-control interpretation is converted into a field-readable matching framework for pressure maintenance, water control, pattern correction and stimulation-parameter review. The framework proposed here is methodologically transferable: it requires only pre-decision decision variables and therefore can be applied to other tight-oil blocks with a comparable variable set. Its calibration, however, is block-specific—the policy thresholds and the feature ranking reported below are derived from the Y284 block and must be re-estimated before being applied elsewhere.

2. Materials and Methods

2.1. Study Area and Data Organization

The Y284 horizontal-well area is located in the southern part of the Yishan slope of the Ordos Basin, and the main target interval is Chang 6 (Figure 1). In the compiled field-scale dataset (939 production-well records), the average effective thickness is approximately 21.2 m. The 2025 geological and reservoir survey of the Tenth Oil Production Plant reports an average porosity of 11.7%, an average air permeability of 0.34 mD, and an average initial oil saturation of 56.8%. The block has been developed mainly by water injection, with several well-pattern configurations and multiple well types. This makes Y284 suitable for analyzing the coupled effects of reservoir quality, pressure support, well-pattern geometry, and engineering stimulation. The hierarchy of the three dataset layers used in this study is defined in Table 1. Because these values describe the full-field population, the reservoir-property ranges of the model-ready sample are reported separately in Table 2. The symbols, units, and definitions of the principal variables used in this study are consolidated in Table 3.

The database was organized at the single-well level. Production and injection records were linked with well type, well pattern, well-location and target-coordinate information, reservoir properties, fracturing parameters, and injection-production spatial indicators. To avoid confusing different sample scopes, three data layers were defined. The field-scale development database contains 939 production-well records for dynamic diagnosis. A second layer contains all wells of the block with a valid five-year cumulative-oil label, of which 410 are available. The model-ready sample is the subset of 338 wells that also has a complete record of reservoir properties, well spacing, row spacing, and fracturing parameters. The accepted model and the SHAP interpretation are both based on this model-ready sample, so that the modelled population and the case-study block are identical.

The model-ready sample now covers both well types and the principal pattern classes of the block. The 939-well field population is dominated by directional wells and by the rhombic inverted nine-spot pattern: 505 of the 939 wells (53.8%) belong to that pattern, and among the 662 wells with a recorded pattern the proportion rises to 76.3%. The 338-well model-ready sample contains 104 horizontal wells and 234 directional wells, and its pattern

composition is rhombic inverted nine-spot (234 wells), five-spot (52), seven-spot (49), and nine-spot (3), so that pattern form is no longer degenerate inside the training sample. Well type and pattern form are nevertheless strongly associated in this block, because every directional well is completed on the rhombic inverted nine-spot pattern whereas the horizontal wells are completed on five-spot, seven-spot or nine-spot patterns; the two descriptors are therefore not statistically independent, and only their joint contribution can be interpreted.

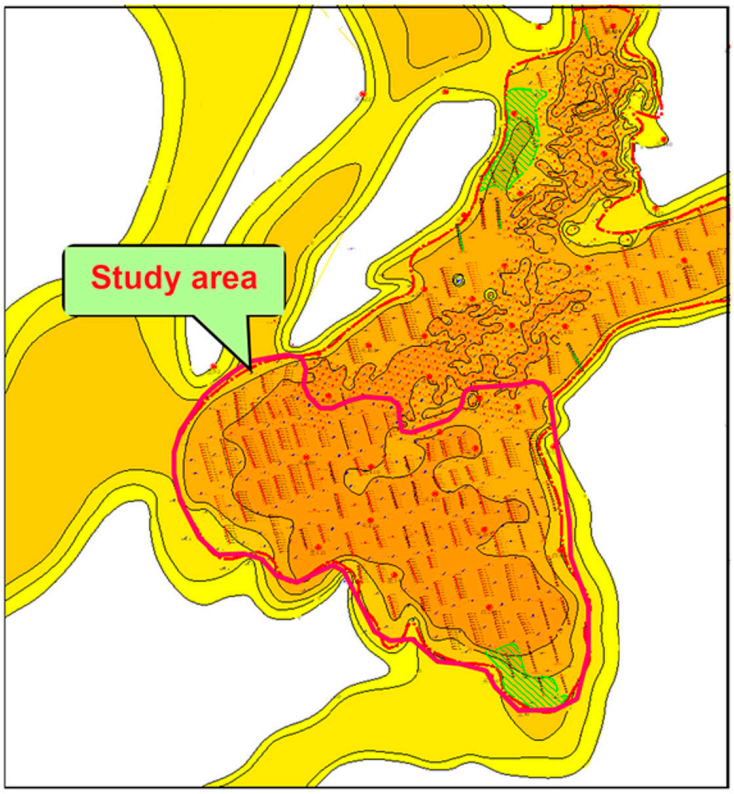


Figure 1. Geological and development setting of the Y284 block in the Longdong Chang 6 tight oil reservoir. The figure defines the single-block study object and the main reservoir-development background. Color indicates effective sandstone thickness, decreasing from orange (thickest) to yellow (thinnest).

Table 1. Y284 data hierarchy, sample scope and representative variables.

Data Layer	Representative Content	Role in the Workflow
Field-scale production database	939 production-well records; monthly production, injection, water response, decline behavior, well type and well pattern	Used for dynamic diagnosis, Arps classification, injection-effect recognition and development-statistics figures
Five-year-label subset	410 wells of the block with a valid five-year cumulative-oil label; production and injection history, well type and well pattern	Used for label screening and for the data-hierarchy audit
Model-ready target subset	338 wells with valid five-year cumulative-oil labels and a complete record of reservoir properties, spacing and fracturing parameters (104 horizontal and 234 directional wells)	Used for model comparison and SHAP interpretation; the model and SHAP samples remain identical
Feature eligibility rule	Pre-decision variables are separated from post-hoc diagnostic variables	Controls data leakage and clarifies whether the model is used for prediction or diagnosis
Quality-control rule	Missing values, disturbed production segments and abnormal operation points are handled by data-source tracing and engineering plausibility checks	Improves reproducibility and prevents unsupported causal interpretation

Table 2. Variable dictionary of the Y284 model-ready sample (n = 338; 104 horizontal and 234 directional wells). Ranges are the minimum–maximum values observed in the modelling set. Variables recorded only for horizontal wells (Horizontal section length, Stage spacing, Stages per 100 m and Clusters per stage) are quoted for the 104 horizontal wells of the sample; together with median grain size they are retained as diagnostic descriptors and do not enter the accepted model.

Variable	Unit	Type	Observed Range	Description	Role
Permeability	mD	Numeric	0.12–6.68	Air permeability of the reservoir interval	Feature (FS-P)
Porosity	%	Numeric	6.21–15.75	Effective porosity of the reservoir interval	Feature (FS-P)
Oil saturation	%	Numeric	40.31–72.26	Initial oil saturation	Feature (FS-P)
Reservoir thickness (vertical)	m	Numeric	2.60–64.50	Net reservoir thickness, measured vertically	Feature (FS-P)
Median grain size	mm	Numeric	0.115–86.108 †	Median grain diameter of the sandstone	Diagnostic descriptor
Horizontal section length	m	Numeric	315–1335	Completed length of the horizontal section	Diagnostic descriptor
Total stages	count	Numeric	1–22	Number of hydraulic-fracturing stages per well	Feature (FS-P)
Stage spacing	m	Numeric	43.24–157.00	Mean spacing between adjacent fracturing stages	Diagnostic descriptor
Stages per 100 m	stage/100 m	Numeric	0.637–2.313	Stage density along the horizontal section	Diagnostic descriptor
Clusters per stage	count	Numeric	1–2	Number of perforation clusters per stage	Diagnostic descriptor
Proppant volume per stage	m^3	Numeric	20.0–285.0	Proppant volume injected per stage	Feature (FS-P)
Injected fluid volume per stage	m^3	Numeric	32.5–4034.9	Fracturing-fluid volume injected per stage	Feature (FS-P)
Pump rate	m^3/min ‡	Numeric	1.3–8.0	Mean pumping rate during the fracturing treatment	Feature (FS-P)
Well spacing	m	Numeric	480–750	Distance between adjacent producers in a row	Feature (FS-P)
Row spacing	m	Numeric	130–175	Distance between adjacent well rows	Feature (FS-P)
Well type	—	Categorical	2 classes: Horizontal (104), Directional (234)	Completion type (horizontal or directional)	Feature (FS-P)
Well pattern	—	Categorical	4 classes: Rhombic inverted nine-spot (234), Five-spot (52), Seven-spot (49), Nine-spot (3)	Injection–production pattern geometry	Feature (FS-P)
Cumulative oil production (5 years)	10^4 t	Numeric	0.0000–1.3484	Target variable: cumulative oil produced by the end of year 5	Target

Note: † Median grain size—this column is a descriptive attribute derived from permeability through an empirical relation (log–log $R^2 \approx 0.97$ against permeability, whereas the directly measured median pore-throat radius of the same database is 0.355 μm) rather than an independent core measurement. It is therefore retained only as a diagnostic descriptor and does not enter the accepted model, and the statement of a previous version that "the model was trained on the values as recorded in the source database" has been deleted. The 15 wells above 1 mm (maximum 86.108 mm) are the upper-tail extension of this permeability–grain-size relation; because the column is neither an independent observation nor a model input, they cannot perturb the feature-importance ranking. ‡ Pump rate—the source fracturing database records this parameter in m^3/min; earlier field lists labelled the column m^3/h. No other model-input variable exceeds its physically plausible range (permeability 0.12–9.38 mD, porosity 6.21–15.75%, oil saturation 40.3–72.3%, net pay 2.6–64.5 m, pump rate 1.3–8.0 m^3/min, stage count 1–22, well spacing 480–750 m), so no outlier filtering or winsorization is required.

Table 3. Symbols, units, and definitions of the principal variables used in this study, with the section of their first appearance. Symbols follow the conventions used in the text and in the figures.

Symbol/Variable	Unit	Definition	First Appearance
Permeability, K	mD	Air permeability of the reservoir interval	Section 2.1
Porosity, φ	%	Effective porosity of the reservoir interval	Section 2.1
Oil saturation, So	%	Initial oil saturation	Section 2.1
Net pay, h	m	Net reservoir thickness measured vertically	Section 2.1
Stage count, Ns	count	Number of hydraulic-fracturing stages per well	Section 2.1
Pump rate, qp	m^3/min	Mean pumping rate during the fracturing treatment (the m^3/h labels of earlier field lists are not used)	Section 2.1
Well spacing, d	m	Distance between adjacent producers in a row	Section 2.1
Row spacing	m	Distance between adjacent well rows	Section 2.1
Normalized cumulative production, Q5	—	Five-year cumulative oil production normalized to a common block-wide basis	Section 2.1
Engineering communication intensity	—	Composite degree of reservoir stimulation determined by pump rate, stage count and the fluid and proppant volumes delivered per stage	Section 1
Group-aware validation	—	Validation scheme in which an entire group of wells (for example one well type) is held out of training	Section 2.3
Cross-validation (CV)	—	Repeated train/test partition used here with folds stratified by well type	Section 2.3

A Y284 model-training database template was further defined to support later model reruns and reproducibility checks. The template keeps one well per row and uses the same field order for geological properties, engineering-stimulation parameters, well-pattern descriptors, and normalized cumulative-production targets. The same field order was applied to the 338-well model-ready table, which was used directly as the model-input table; the same field order is retained so that later model reruns and reproducibility checks can be repeated without redefining variables.

The preprocessing protocol records feature coverage before model fitting, applies physically defensible range checks, and distinguishes unavailable values from true zeros. Missing-data handling, outlier screening, and split definition were specified before model selection because otherwise optimistic performance can arise from reusing correlated observations or from data-dependent preprocessing [38–40].

For well i, the five-year cumulative-oil target was defined as the sum of monthly oil production over the first 60 effective production months, as expressed in Equation (1). Wells without a complete and traceable target window were excluded from the model-ready target subset rather than assigned an imputed cumulative-production label.

$$Y_i^{(5\ y)} = \sum_{t=1}^{60} q_{i,t} \quad (1)$$

2.2. Reservoir-Engineering Diagnosis

Reservoir-engineering diagnosis was placed before machine learning. Its purpose was to convert production history into interpretable labels, screen disturbed data, and prevent the model from learning statistically strong but physically unreasonable relationships. Monthly oil-rate data were first checked for well status, shut-in periods, stimulation measures, pump adjustments, water shutoff, and major injection-production changes. Zero-rate shut-in months and strongly disturbed segments were excluded from decline fitting,

whereas real extreme responses caused by operations or geological conditions were retained and interpreted.

Production decline was fitted using Arps exponential, hyperbolic, and harmonic functions. A valid fitting segment required continuous positive oil-rate records and sufficient duration for stable parameter estimation. This criterion applies to the decline-curve subset used for the reservoir-engineering diagnosis and not to the five-year cumulative-oil target defined in Section 2.1, which is a sum of monthly rates and can legitimately be zero for a well shut in shortly after commissioning. The best decline model was selected using goodness of fit, residual pattern, and engineering reasonableness. The decline type was determined from the fitted b value and model form: exponential decline corresponds to b close to 0, harmonic decline corresponds to b close to 1, and hyperbolic decline covers the intermediate range. Decline intensity was classified as strong, moderate, or slow according to first-year decline rate and early nominal decline behavior. Wells with insufficient continuity or unstable fitting were reported separately rather than forced into a decline class [41].

For the hyperbolic candidate, the rate-time relation is given in Equation (2), where $q(t)$ is oil rate at time t, qi is the rate at the decline origin, Di is the nominal initial decline rate, and b is the decline exponent. The exponential form is obtained as b approaches zero, whereas $b = 1$ gives harmonic decline. This representation permits consistent comparison of the three decline behaviors within the same screening procedure.

$$q(t) = \frac{q_i}{(1 + b\, D_i t)^{\frac{1}{b}}},\ 0 < b \leq 1 \tag{2}$$

For each eligible production segment, parameters were estimated by minimizing the log-rate residual objective in Equation (3). The small positive constant epsilon avoids an undefined logarithm near low-rate observations. The selected model was then checked against fitting stability, residual behavior, and known production disturbances.

$$J(\theta) = \frac{1}{n}\sum_{k=1}^{n}\left[\ln\left(q_{o,k} + \varepsilon\right) - \ln\left(q_{(m,k)(\theta)} + \varepsilon\right)\right]^2 \tag{3}$$

First-year decline intensity was calculated from the fitted rate at the decline origin and after 12 months using Equation (4). This index was used for the strong-, moderate-, and slow-decline diagnosis reported in Table 4.

$$D_{1\,y} = 1 - \frac{q(12)}{q(0)} \tag{4}$$

Table 4. Dynamic-response classification and Arps-diagnosis rules used for Y284.

Label Group	Classes	Operational Threshold
Arps model type	Exponential, hyperbolic and harmonic decline	Determined from the best fitted Arps form and b value after removing shut-in and strongly disturbed production segments
Decline intensity	Strong, moderate and slow decline	Classified by first-year decline rate and early nominal decline behavior; unreliable fits are reported separately
Injection effect	Fast, moderate, slow and unclear/no response	Defined only for wells with interpretable injection-production relationship; natural-energy wells are excluded from injection-effect conclusions
Water response	Fast, moderate and slow water response	Defined by water-breakthrough timing and water-cut rise; unresolved records are not mixed with recognized water-response classes

Water-response and injection-effect labels were defined only for wells with interpretable injection-production relationships. Fast response was defined by short time lag after injection or injection adjustment; moderate and slow response were assigned according to progressively longer lag times and weaker oil-production response. Natural-energy wells were not used for injection-effect conclusions. This revision is important because water breakthrough and oil response are post-development diagnostic variables; they can support engineering interpretation but should not be used as unconstrained predictors for pre-development five-year cumulative-oil forecasting.

2.3. Explainable Machine-Learning Workflow

The modeling workflow used full five-year cumulative oil production as the core target and was run on the Y284 model-ready sample of the block itself (338 wells with valid five-year labels). To control leakage, two feature sets were defined. The pre-decision feature set includes reservoir properties, pore-throat indicators where available, well type, well pattern, well spacing, row spacing, horizontal-section length, fracturing-design parameters, and planned development mode. The diagnostic feature set adds cleaned response indicators such as decline type, water-response type, injection-effect type, and pressure-support descriptors. Annual oil, produced liquid, water cut and response labels derived from the target period are not used to claim pre-development prediction; they are used only for post-hoc diagnosis and policy interpretation.

Accordingly, a cross-validation scheme that respects the structure of the field data is preferred to a single random split when wells share a development stage, well-pattern context, or local geological setting. In this study, the folds are stratified by well type so that every fold preserves the horizontal/directional proportion of the sample, and the results are averaged over 20 random seeds. Such validation does not remove all extrapolation uncertainty, but it provides a more credible test of stability than random resampling alone, whereas a holdout that removes an entire well type forces the model to extrapolate and is therefore reported separately rather than used as the acceptance criterion.

The pre-decision prediction problem and the post-hoc diagnostic augmentation are formalized in Equation (5). Here, xpre contains only information available before development or at the stated early-decision point, whereas z contains dynamic diagnostic variables. This separation defines the permissible information boundary for claims of prospective performance prediction.

$$\hat{Y}_i = f_{\theta(x_i^{pre})},\ x_i^{diag} = \left[x_i^{pre}, z_i\right] \tag{5}$$

Four tree-based algorithms were used as the primary comparison set: random forest, XGBoost, LightGBM and CatBoost. CatBoost was emphasized because the model-ready table contains categorical descriptors such as well type and well pattern. Model performance should be evaluated with R2, RMSE, and MAE on training, validation, and independent test sets, and also checked under group-aware validation by well type, well pattern, or production-start period. A model is accepted only when it satisfies statistical accuracy, validation stability, and reservoir-engineering plausibility [42]. The corresponding design steps, from feature-set construction to policy interpretation, are listed in Table 5. SHAP is then used to quantify feature importance, contribution direction, interaction behavior, and potential engineering thresholds.

Model performance was quantified by the three complementary statistics in Equations (6)–(8), where Y_i and $\bar{Y}i$ denote observed and predicted five-year cumulative oil, respectively. R^2 measures explained variance, whereas *RMSE* and *MAE* retain the production-unit error scale and should be reported jointly for every train-validation-test split.

$$R^2 = 1 - \frac{\left[\sum_{i=1}^{n} e_i^2\right]}{\left[\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2\right]} \quad (6)$$

$$RMSE = \left[\left(\frac{1}{n}\right)\sum_{i=1}^{n} e_i^2\right]^{\frac{1}{2}} \quad (7)$$

$$MAE = \left(\frac{1}{n}\right)\sum_{i=1}^{n}|e_i| \quad (8)$$

Table 5. Engineering-constrained modeling, validation and interpretation design.

Step	Implementation	Engineering Purpose
Feature-set construction	Build FS-P using pre-decision variables and FS-D using post-hoc diagnostic variables	Avoid target leakage and distinguish prediction from diagnosis
Model comparison	Compare random forest, XGBoost, LightGBM and CatBoost using R2, RMSE and MAE	Identify a statistically stable model rather than relying on one algorithm
Strict validation	Use a well-type-stratified cross-validation with repeated seeds and an independent holdout; report a well-type holdout separately as a test of extrapolation rather than as the acceptance criterion	Reduce overestimation caused by neighboring wells or similar development-stage samples
Interpretation	Calculate SHAP ranking, dependence plots and interaction checks for key variables	Identify dominant controls and possible threshold intervals
Policy mapping	Link SHAP controls with reservoir-engineering diagnosis and typical well-group behavior	Convert model interpretation into pressure-maintenance, water-control, pattern-correction and stimulation-review actions

3. Results

3.1. Planar Distribution Analysis of Training-Database Parameters

Before evaluating production dynamics, the variables listed in the Y284 training database were organized for plane-map analysis. The planned spatial analysis includes four groups: reservoir-quality parameters, engineering-stimulation parameters, well-pattern and spacing parameters, and normalized cumulative-production targets. Reservoir-quality maps include permeability, porosity, oil saturation, and oil-layer vertical thickness; engineering maps include cluster number per stage, stages per 100 m, sand volume, injected-fluid volume, displacement rate, total stages, and stage spacing; well-pattern maps include well type, well pattern, development mode, well spacing, row spacing, and horizontal-section length.

The purpose of this plane-map analysis is to identify whether high-production wells are controlled by favorable reservoir distribution, stronger stimulation, better injection-production geometry, or their combined matching. The planar maps are used to compare high-value and low-value zones of five-, six-, and seven-year normalized cumulative oil production. The map-based interpretation will also provide a spatial check for SHAP results, because a factor with high model importance should be consistent with recognizable geological or engineering distribution patterns rather than only with statistical correlation.

3.2. Production Dynamics and Energy Supplementation

Y284 shows a clear staged production and injection history. Annual water injection increased rapidly in the early development stage and reached a high level around 2013–2015,

followed by gradual weakening. After 2021, injection intensity declined markedly, while water cut remained at a relatively high level. This indicates that pressure support and water-control effectiveness must be evaluated together rather than by injection volume alone [43].

Annual oil production increased during early development, entered a fluctuating adjustment stage, and later showed renewed support. The combined curves show that Y284 is not simply constrained by insufficient injection. In some periods, injection support existed but did not translate into proportional oil increase, suggesting ineffective sweep, incomplete pattern control, or preferential water advance. This provides the engineering basis for later feature selection and policy mapping (Figure 2).

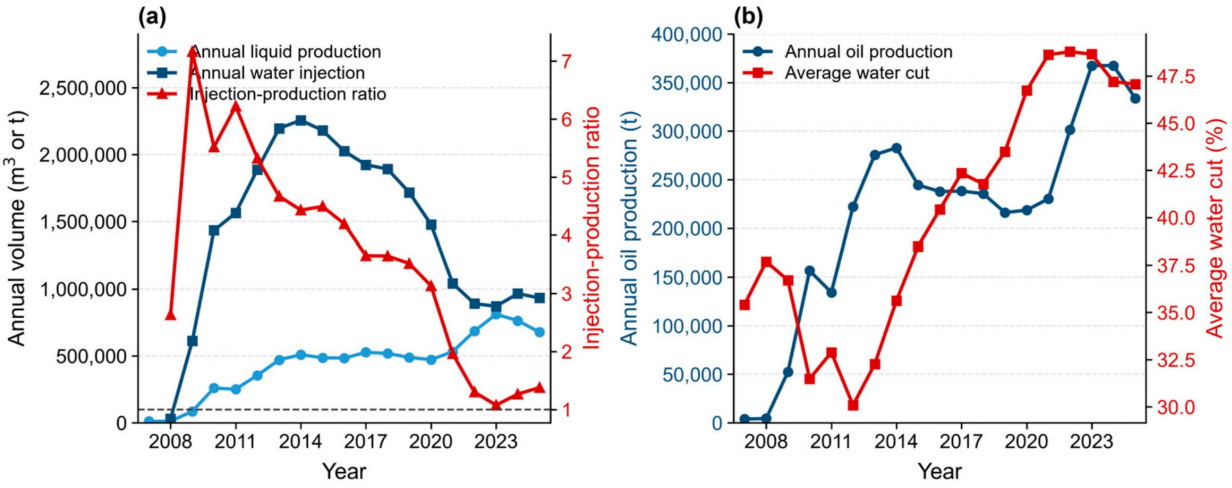


Figure 2. Annual production, injection, and water-cut evolution in the Y284 block. (**a**) Annual liquid production, water injection, and the injection–production ratio, reflecting staged pressure support and a subsequent weakening of injection as the ratio declines toward equilibrium. (**b**) Annual oil production and average water cut, showing a sustained rise in water cut despite continued oil output.

3.3. Decline Behavior and Water Response

The single-well Arps fitting results show that exponential and hyperbolic decline are the main fitted forms in Y284. In the current statistics, 716 wells have valid Arps decline classifications, including 328 exponential, 276 hyperbolic, and 112 harmonic fits; 223 wells with unreliable or insufficient fitting are excluded from model-type statistics. Slow-decline wells dominate the decline-intensity distribution, while moderate- and strong-decline wells remain key targets for production-stabilization diagnosis.

The Arps parameter distributions further clarify the internal heterogeneity of decline behavior. Strong-decline wells have the highest initial decline rate and first-year decline rate, indicating rapid early productivity loss. The b-value ranges of strong- and moderate-decline wells are relatively wide, showing large differences in decline-curve shape. Slow-decline wells generally have lower b values and more stable decline behavior, although part of the slow-decline group shows dispersed fitting R2 because of production-system adjustments, measures, or data fluctuation (Figures 3 and 4).

The response classification reveals a typical mismatch: water response is fast, whereas oil-production response is differentiated. Fast water breakthrough indicates that injected water or water-bearing pathways can quickly influence production wells. However, rapid water arrival does not necessarily mean effective oil displacement. Therefore, simple injection strengthening may increase water-cut risk if injection-production connectivity is too strong or sweep efficiency is poor (Figure 5).

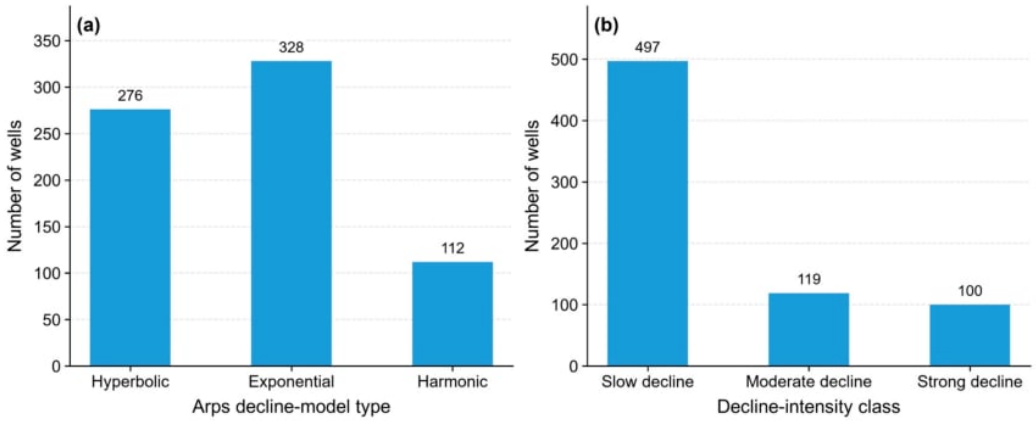


Figure 3. Arps decline-model type and decline-intensity distribution of Y284 wells. (**a**). Frequency distribution of fitted Arps decline-model types. (**b**). Frequency distribution of decline-intensity classes.

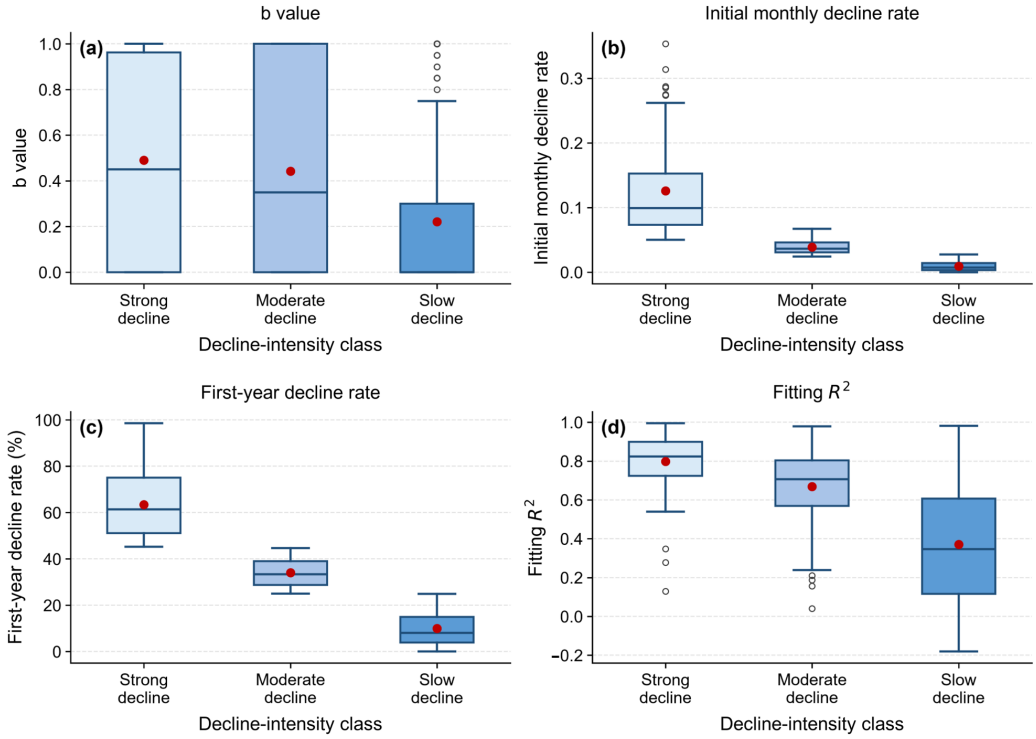


Figure 4. Comparison of Arps parameters among strong-, moderate- and slow-decline wells. (**a**) *b* value decreases systematically from strong to slow decline. (**b**) Initial monthly decline rate is highest for strong-decline wells and falls monotonically toward slow decline. (**c**) First-year decline rate (%) cleanly separates the three classes. (**d**) Fitting R^2 is highest for strong-decline wells and lowest for slow-decline wells, reflecting weaker model fit where decline is weakest.

3.4. Effects of Well Type and Well Pattern

The Y284 well population is dominated by directional wells, with a much larger sample size than other well types. However, the development-effect statistics show that horizontal wells have the strongest productivity advantage. The median peak oil rate of horizontal wells is approximately 9.52 t/d, and the median cumulative oil is approximately 9612 t, both higher than those of directional and highly deviated wells. This indicates a strong productivity potential of horizontal wells under favorable reservoir and engineering conditions (Figure 6).

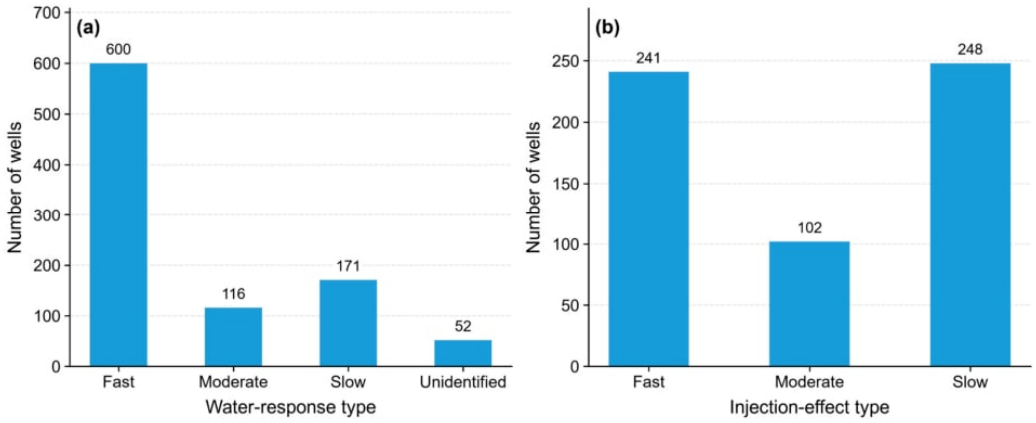


Figure 5. Injection-effect and water-breakthrough response types in Y284. (**a**) Water-response types, with the large majority of wells showing a fast water response. (**b**) Injection-effect types, where fast and slow effects occur in comparable numbers and moderate effects are about half as frequent.

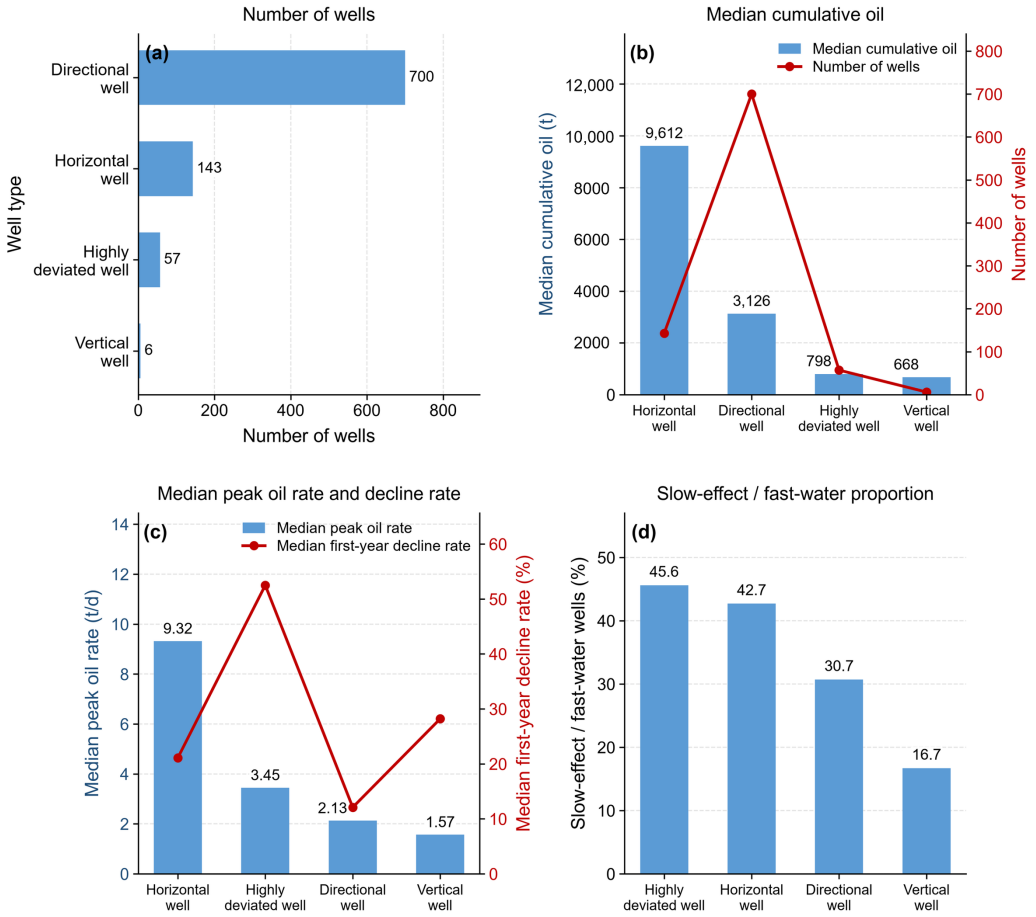


Figure 6. Effect of well type on development performance in Y284. (**a**) Well-type composition, dominated by directional wells. (**b**) Median cumulative oil, with horizontal wells highest despite fewer wells. (**c**) Median peak oil rate and first-year decline rate, where horizontal wells lead in peak rate and highly deviated wells decline fastest early. (**d**) Proportion of slow-effect/fast-water wells, highest for highly deviated and horizontal wells.

The decline characteristics differ among well types. Highly deviated wells show a relatively high first-year decline rate, indicating rapid early production loss. Horizontal wells show better cumulative-production performance with a more controllable decline response. Nevertheless, the well-type conclusion must consider sample bias: horizontal

wells may preferentially target better reservoir intervals, longer effective drainage sections, or stronger stimulation designs. Therefore, same-reservoir-class comparison or propensity-score matching is required before causal conclusions are made.

Observed differences among well types or well-pattern classes are not automatically causal because reservoir quality, completion design, and commissioning time may be jointly selected. Where sufficient common support is available, stratified comparison or propensity-score matching can be used as a sensitivity analysis; otherwise, the policy output should remain an engineering recommendation rather than a causal treatment effect [44–46].

Well-pattern form controls the spatial organization of injection and production. In the 939-well full-field population, the rhombic inverted nine-spot pattern is the dominant pattern form. In the 338-well model-ready sample, the composition is closer to the field population: the rhombic inverted nine-spot pattern contributes 234 wells, five-spot 52, seven-spot 49, and nine-spot 3 (Table 2). The five-spot and seven-spot wells show higher cumulative-oil medians than the rhombic inverted nine-spot wells in the current data, but they are also the horizontal wells, so pattern and well type cannot be separated here. The dominant pattern therefore does not automatically correspond to the highest development effect, and pattern performance must be interpreted together with row spacing, reservoir continuity and water-response class (Figure 7).

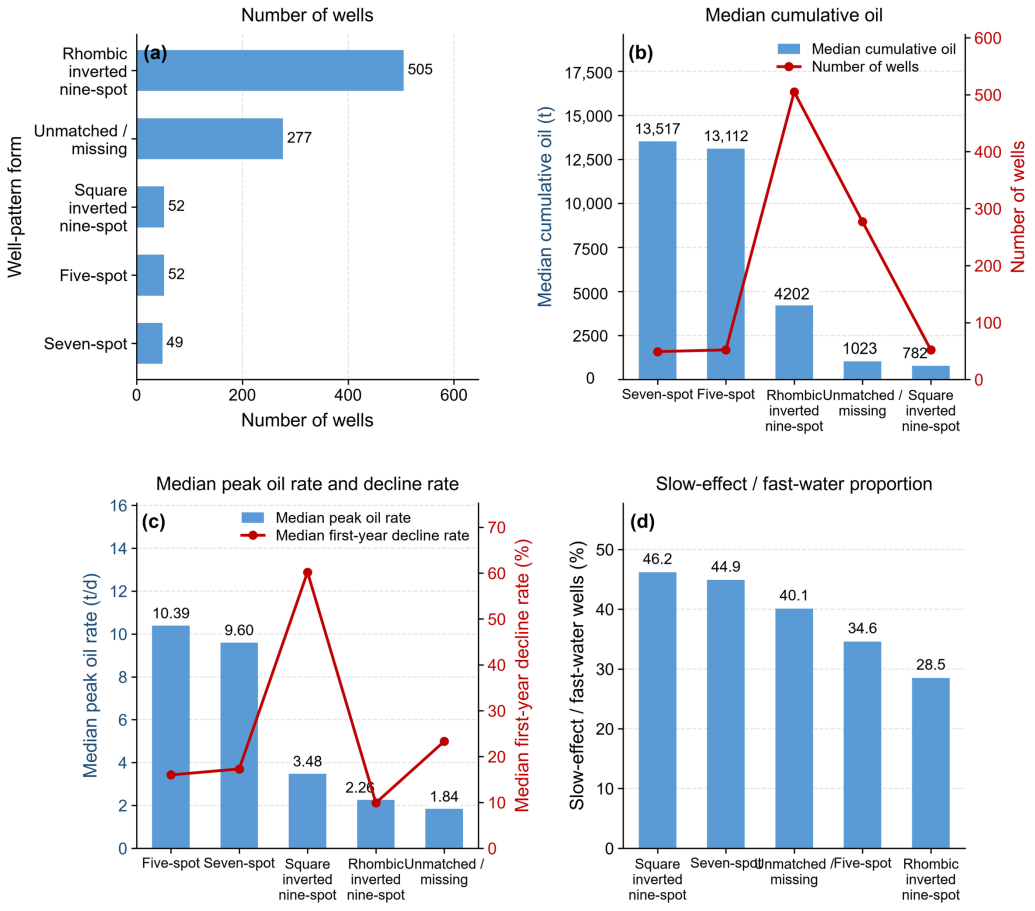


Figure 7. Effect of well-pattern form on Y284 development performance. (**a**) Pattern-form composition, dominated by the rhombic inverted nine-spot. (**b**) Median cumulative oil, where minority five-/seven-spot patterns outperform the dominant rhombic inverted nine-spot. (**c**) Median peak oil rate and first-year decline rate, with five-/seven-spot wells combining higher peak rates and lower decline. (**d**) Proportion of slow-effect/fast-water wells, lowest for the rhombic inverted nine-spot, indicating weaker injection–production connectivity.

The well-pattern results suggest that Y284 optimization should focus on the matching degree between pattern geometry and reservoir connectivity. Where spacing is too large, or the pattern is incomplete, injected water may fail to provide effective pressure support. Where connectivity is too strong, rapid water breakthrough may dominate. Within the model-ready subset, this explains why row spacing, well pattern, and development mode appear as important variables in the machine-learning interpretation; the result cannot be extrapolated to pattern classes that are not represented in the training sample.

3.5. Model Performance and Model Selection

The four tree-based algorithms were trained on the same pre-decision feature set and on the same Y284 model-ready sample ($n = 338$). Their performance was quantified with R2, RMSE, and MAE computed from the out-of-fold predictions of a well-type-stratified five-fold cross-validation and repeated over 20 random seeds, and the same runs were repeated under the original random split and under a well-type holdout so that the three protocols can be compared directly (Section 3.8). Regularisation and early stopping were applied throughout: the tree depth was limited to 3–5, the minimum leaf size to 5–10, the learning rate to 0.03–0.05, and the row and column subsampling to 0.7–0.9. Table 6 reports the metrics and states which model was accepted for interpretation. Figure 8 compares the predicted and observed five-year cumulative oil of the accepted model across the training and out-of-fold test wells, and Figure 9 shows its residual diagnostics on the same out-of-fold predictions. The accepted model is the one that satisfies statistical accuracy, validation stability, and reservoir-engineering plausibility simultaneously; a higher test R2 alone is not sufficient for acceptance.

Table 6. Performance comparison of random forest, XGBoost, LightGBM and CatBoost for five-year cumulative oil (target unit: 10^4 t), trained on the Y284 model-ready sample ($n = 338$; 104 horizontal and 234 directional wells). Metrics are out-of-fold statistics from a well-type-stratified five-fold cross-validation; the test R2 is reported as the mean ± standard deviation over 20 random seeds, and RMSE and MAE are means over the same seeds. The training R2 is that of the model fitted on the training folds. The group-aware results of the previous version are reported separately in Section 3.8. RMSE is also reported for the training folds, so that the inflation caused by predicting unseen wells can be read directly from the table; it is smallest for the accepted model.

Model	R2 (Train)	R2 (Test)	RMSE (Train)	RMSE (Test)	MAE (Test)	Accepted
Random forest	0.704	0.546 ± 0.012	0.148	0.184	0.122	Yes
CatBoost	0.734	0.541 ± 0.043	0.138	0.184	0.122	No
LightGBM	0.736	0.537 ± 0.036	0.138	0.185	0.124	No
XGBoost	0.740	0.513 ± 0.044	0.137	0.190	0.126	No

Note. A multilayer-perceptron reference (R2 = 0.424 ± 0.078) is reported in the Supplementary Materials but is not included in the acceptance decision, because its out-of-fold spread is an order of magnitude larger. No post-hoc or diagnostic variable (decline type, water cut, response-to-injection indicator) enters the feature set, so the train–test gap reported here reflects the capacity of the models relative to the sample size rather than leakage.

Two entries in Table 6 require explicit comment. First, the random forest and CatBoost are indistinguishable on out-of-fold RMSE and MAE: both give 0.184 in RMSE and 0.122 in MAE at the precision of the table, and the seed-resolved means differ by 0.0009 in RMSE, which is less than half of one seed-to-seed standard deviation of the forest. The acceptance decision therefore cannot rest on those two figures alone. Models are ranked by out-of-fold RMSE, and where two models tie at the reported precision, the ranking is decided by the stability of that metric across seeds and by the size of the train–test gap, which are properties a screening tool must have. On this criterion, the random forest is preferable: its out-of-fold R2 reproduces to ±0.012 over 20 seeds, whereas CatBoost, LightGBM and XGBoost vary

by ±0.043, ±0.036 and ±0.044, and their training R2 varies by ±0.047 to ±0.065 against ±0.002 for the forest. Its train–test inflation is also the smallest of the four models, in both R2 (0.704 against 0.546) and RMSE (0.148 against 0.184). A ranking of candidate wells that changes when the random seed changes is of little operational use, which is why the most stable model was retained even though a boosting model reaches a marginally higher training score.

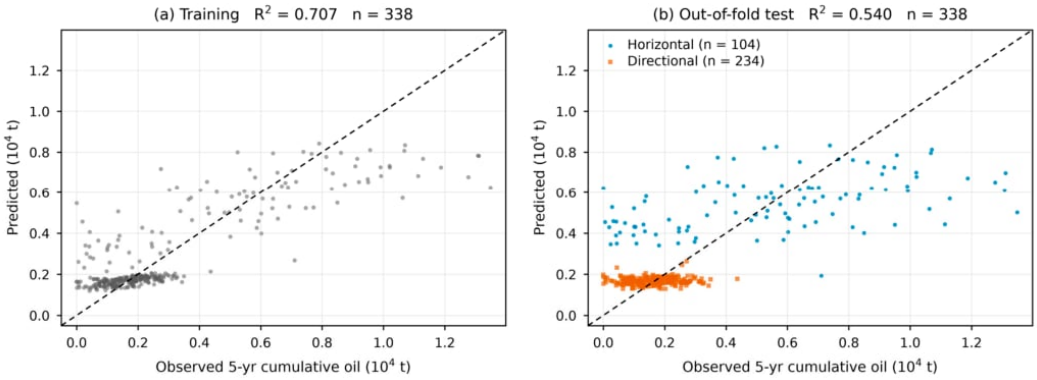


Figure 8. Predicted versus observed five-year cumulative oil for the accepted regularised random-forest model (Y284 model-ready sample, $n = 338$). (**a**) Training and (**b**) out-of-fold test wells of the well-type-stratified five-fold cross-validation; test wells are coloured by well type, and the 1:1 line is a reference. The panels show the first of the 20 seeds summarised in Table 6, which reports 20-seed means (0.704 training; 0.546 ± 0.012 out-of-fold).

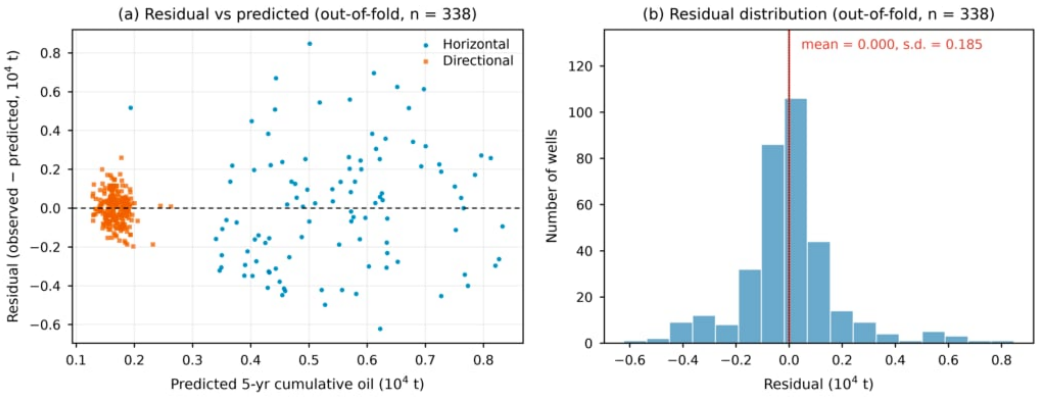


Figure 9. Residual diagnostics of the accepted model on the out-of-fold predictions of the well-type-stratified cross-validation ($n = 338$). (**a**) Residual versus predicted value with the zero line, the test wells coloured by well type, and (**b**) residual distribution with the mean marked, in order to check homoscedasticity and systematic bias along the production range.

The train–test gap is reported in full rather than left implicit. For the accepted forest, the in-fold RMSE is 0.148 against 0.184 out-of-fold, an inflation of 24%, and the corresponding R2 values are 0.704 and 0.546. Three checks indicate that this gap reflects the sample size rather than memorisation. (i) Regularisation was applied deliberately, and its effect is measurable: the unregularised configuration of the same forest reaches R2 (train) = 0.938 with an in-fold RMSE of 0.068, but its out-of-fold R2 remains 0.543 (RMSE 0.184), so the extra capacity is spent entirely on the training wells and buys no predictive accuracy; the accepted configuration removes that capacity at no cost in out-of-fold performance. (ii) The learning curve is monotone and does not diverge: when the training set is reduced from 270 to 108 wells, the out-of-fold RMSE rises from 0.182 to 0.199 while the in-fold RMSE rises from 0.148 to 0.166, so the gap stays between 0.03 and 0.04 and additional data improve

generalisation instead of being absorbed by the model. (iii) The residual diagnostics in Figure 9 show no systematic structure along the production range, and the 20-seed spread of the accepted model (±0.012 in R2, i.e., 2.2% of its mean) is small compared with the gap itself. The gap is therefore reported as a property of a 338-well, twelve-predictor problem rather than as evidence against the model, and the paper continues to present the framework as a screening and ranking tool rather than as a high-accuracy predictor.

3.6. Dominant Controls Identified by SHAP

The explainable machine-learning results indicate that development effectiveness in the model-ready sample is dominated by engineering communication intensity, with reservoir conditions and injection-production geometry as secondary controls. Aggregated by physical factor group, engineering communication intensity accounts for 69.9% of the total mean absolute SHAP value, geological storage capacity for 21.6%, and injection-production control for 8.6%. Well type contributes no additional information once the stimulation-intensity variables are included, and the marginal contribution of permeability, porosity, and oil saturation is not statistically distinguishable from zero at this sample size. At the level of individual variables, total stage count and pump rate rank first and second. The mean absolute SHAP values of all variables are reported in Table 7 and visualized in Figure 10.

Table 7. Contribution and importance of the accepted variables. Permutation importance is the decrease in out-of-fold R2 after shuffling a feature, averaged over the cross-validation folds (XGBoost reported; the LightGBM, random-forest and full multi-algorithm values are given in the Supplementary Materials). Mean absolute SHAP values are those of the regularised gradient-boosting model (XGBoost). The contribution direction records the sign of the mean SHAP value where it is identifiable; entries marked n.i. have no identifiable marginal contribution in this sample.

Rank	Feature	Mean|SHAP|(10^4 t)	Permutation Importance (ΔR2)	Contribution Direction	Physical Category
1	Total stages (count)	0.086	0.215	positive	Engineering communication intensity
2	Pump rate (m^3/min)	0.066	0.211	positive	Engineering communication intensity
3	Reservoir thickness (m)	0.028	0.046	positive	Geological storage capacity
4	Well spacing (m)	0.018	0.044	positive	Injection-production control
5	Fluid volume per stage (m^3)	0.021	0.040	positive	Engineering communication intensity
6	Total fluid volume (m^3)	0.023	0.030	positive	Engineering communication intensity
7	Row spacing (m)	0.007	0.024	positive	Injection-production control
8	Porosity (%)	0.019	0.018	n.i.	Geological storage capacity
9	Well type	0.000	0.000	n.i.	Well type
10	Permeability (mD)	0.007	−0.002	n.i.	Geological storage capacity
11	Oil saturation (%)	0.009	−0.004	n.i.	Geological storage capacity
12	Proppant volume per stage (m^3)	0.008	−0.006	n.i.	Engineering communication intensity

Note. Importance is computed on the out-of-fold predictions of the well-type-stratified five-fold cross-validation (n = 338), averaged over 10 repeats and over the folds. The column labelled n.i. (not identifiable) means that the permutation importance is within ±0.011 of zero and the mean absolute SHAP value is below 0.01, so no marginal contribution can be resolved at this sample size; oil saturation in particular shows no negative contribution, and the negative entry of the previous version was a confounding artefact caused by the association between well type and reservoir quality (Section 3.10). Because several predictors are strongly correlated, the split of importance between them should be read as a joint attribution rather than as a unique causal ranking.

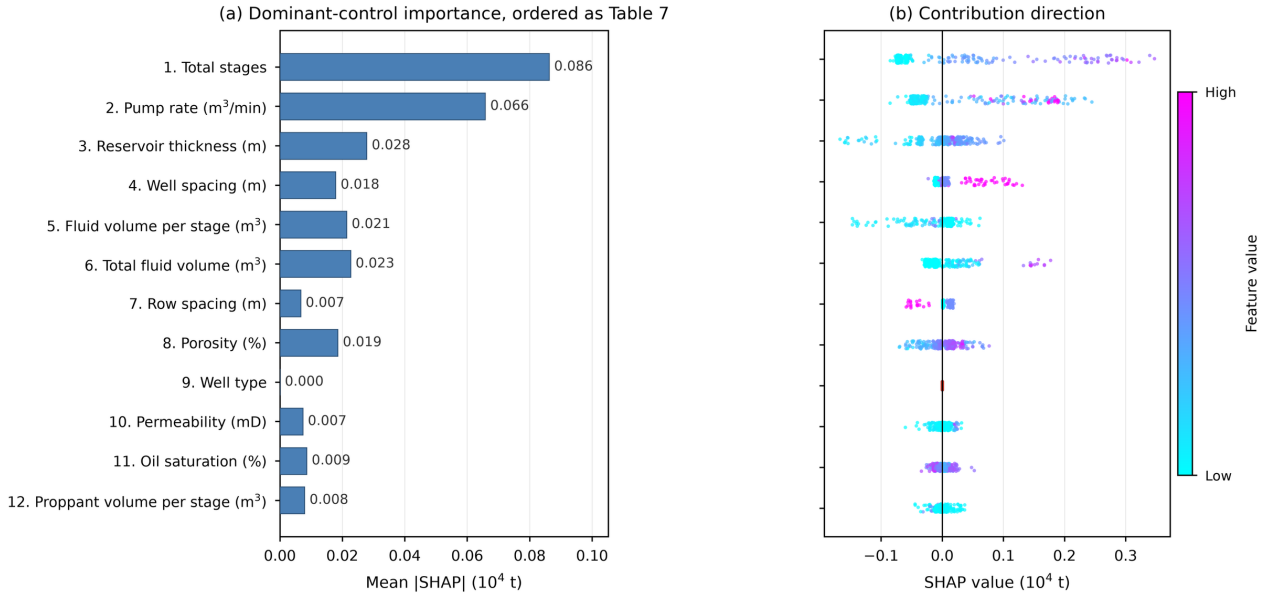


Figure 10. Dominant controlling factors of the Y284 model-ready sample ($n = 338$); the SHAP values are those of the regularised gradient-boosting model fitted on all 338 wells and are listed in Table 7. (**a**) Ranking of dominant-control importance, given as the mean absolute SHAP value in target units (10^4 t) and ordered from the largest to the smallest value; the predictors and their values are the same as in Table 7, whose rank column is defined by permutation importance, so the two orders coincide at the top two ranks and may differ further down. (**b**) Contribution-direction scatter plot, in which each point is one well of the model-ready sample, the horizontal position is its SHAP value and the colour is its feature value (dark blue, low; crimson, high); well type is drawn with the discrete colours indicated in the legend. Total stage count and pump rate are the two leading controls, followed by reservoir thickness, the injected-fluid volumes, porosity and well spacing; the remaining five predictors are marginal.

This result is consistent with the preceding reservoir-engineering diagnosis. Pump rate and total stage count describe how effectively the fracturing treatment establishes communication with the reservoir, whereas porosity and oil-layer thickness describe the storage and seepage basis for productivity. Treatment intensity should nevertheless be interpreted together with the engineering constraints of large-volume hydraulic fracturing, including fracture-induced seismicity and pre-existing-fault reactivation [47]. Injection-production control, represented by well pattern, well spacing and row spacing, plays a smaller but non-negligible role. SHAP results should therefore be interpreted by combining mean absolute SHAP values with permutation importance, dependence plots and interaction checks, and by reading strongly correlated predictors as a joint attribution rather than as a unique causal ranking. In particular, pump rate and total stage count are treated as candidate nonlinear controls rather than simple larger-is-better variables; this reading is supported by the numerical importance ranking in Table 7 and by the dependence and interaction plots in Figures 11 and 12.

The SHAP explanation follows the additive decomposition in Equation (9), where phi0 is the expected model output, $\varphi i,j$ is the contribution of feature j for well i, and I_j is the mean absolute SHAP importance used for ranking. Positive or negative $\varphi i,j$ values are interpreted as model contributions within the accepted feature scope; they are not treated as independent causal effects.

$$f(x_i) = \varphi_0 + \sum_{j=1}^{p} \varphi_{i,j},\ I_j = \left(\frac{1}{n}\right)\sum_{i=1}^{n}\left|\varphi_{i,j}\right| \tag{9}$$

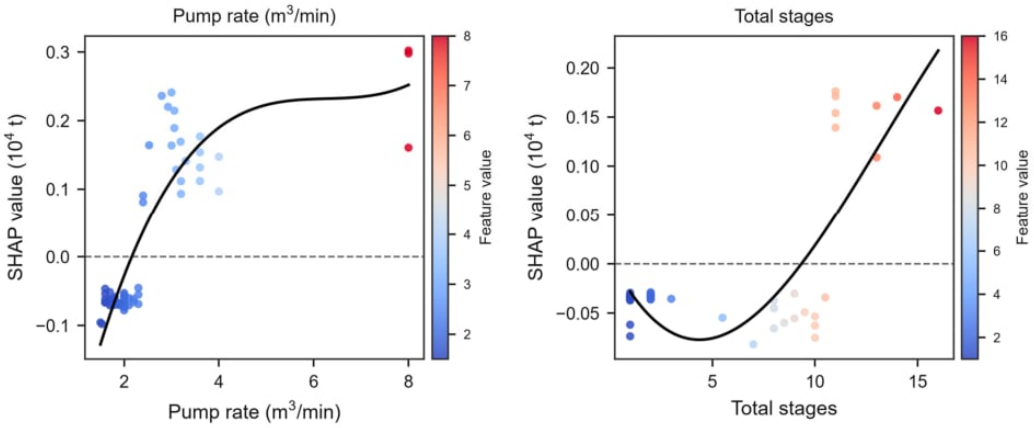


Figure 11. SHAP dependence plots for pump rate and total stage count, the two leading variables of the engineering-communication group ($n = 338$). A cubic trend line is added to each panel in order to test whether the two variables act as nonlinear controls and to locate the inflection zones that can be converted into field thresholds.

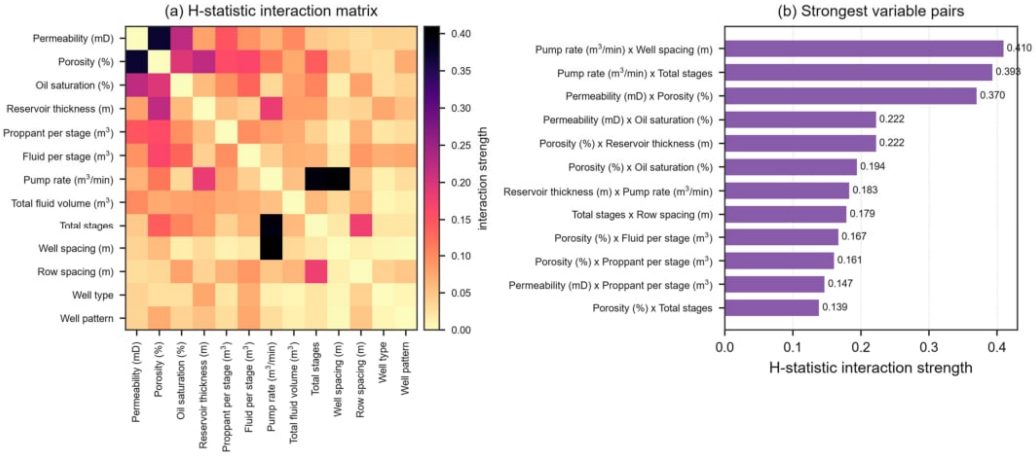


Figure 12. Interaction analysis of the accepted model. (**a**) Friedman-Popescu H-statistic interaction matrix and (**b**) the twelve strongest variable pairs. The strongest interaction links pump rate with well spacing (H = 0.410) and with total stage count (H = 0.393), indicating that the effect of the stimulation rate depends on the injection-production geometry rather than acting as an independent control.

3.7. Reservoir-Development Matching Framework

Based on the dynamic diagnosis and SHAP interpretation, a matching framework was established for Y284. The framework starts from reservoir quality and pore-throat structure, then evaluates whether the energy-supplementation mode and well-pattern geometry can provide effective pressure support. It then checks whether decline intensity, water response, and stimulation scale indicate pressure shortage, weak connectivity, preferential water breakthrough, or excessive early productivity loss.

The policy-mapping step is consistent with established well-placement, control-optimization and injector-producer connectivity studies, which show that geometry and operating controls must be evaluated together rather than optimized in isolation [48–54].

For wells with favorable reservoir conditions and effective injection response, the development policy should maintain stable pressure support and avoid aggressive drawdown. For wells with favorable reservoir quality but weak oil response, injection-production connectivity and pattern completeness should be reviewed. For wells with fast water response and limited oil increase, the priority should shift to water control, profile control, and differentiated injection allocation. For wells with large spacing or incomplete pattern control,

pattern correction, infill screening, or producer-injector relationship adjustment should be considered. In all four cases, the stimulation design should be reviewed against the local stress state before the treatment scale is increased [55]. The four cases summarized above are reported as field-readable matching rules in Table 8, and the overall framework that links reservoir condition, energy supplementation, well-pattern/well-type configuration and development-policy recommendation is given in Table 9.

Table 8. Field-readable matching rules for Y284 development-policy optimization.

Reservoir-Development Condition	Main Diagnosis	Recommended Policy
Favorable reservoir + effective oil response + controlled water cut	Geological quality can be converted into production under stable pressure support	Maintain pressure support and avoid excessive drawdown
Favorable reservoir + weak oil response	Static reservoir potential is not fully converted because connectivity or pattern control is insufficient	Review injector-producer relation, pattern completeness and injection allocation
Fast water response + limited oil increment	Injected water advances rapidly but sweep or displacement efficiency is weak	Implement water control, profile control and differentiated injection
Large spacing or incomplete pattern + low cumulative oil	Pressure support and controlled drainage volume are insufficient	Evaluate pattern correction, infill screening or producer-injector relationship adjustment
High early decline after strong stimulation	Early productivity is high but sustained pressure support or fracture effectiveness is insufficient	Optimize production system and review stimulation-parameter matching

Table 9. Y284 reservoir-development matching framework. The framework links reservoir condition, energy supplementation, well-pattern/well-type configuration and development-policy recommendation.

Constraint	Main Content	Application Direction
Reservoir Constraints	Differences between Class II and Class III reservoirs affect the energy-recharge response and well-pattern adaptability	Develop strategies by reservoir type
Energy-Recharge Constraints	Responses of natural energy, near-wellbore water injection, peripheral water injection and well-pattern water injection differ and remain to be verified	Determine the direction for improving energy recharge
Well-Pattern Constraints	Well type, well-pattern form, well spacing and row spacing affect water breakthrough, fluid breakthrough and production decline	Optimize the well-pattern and well-type combination
Dynamic Constraints	Rising water cut, decline intensity and pressure maintenance determine the priority of adjustments	Differentiated production stabilization and water control

3.8. Validation-Strategy Sensitivity and the Origin of the Negative Group-Aware R2

The validation protocol used in the previous version grouped the folds by well type, so that the model was trained on one well type and tested on the other. Because the two populations differ by a factor of 3.3 in the target (median five-year normalized cumulative production 0.5591 for the horizontal wells against 0.1746 for the directional wells), such a split constitutes an extrapolation to a target range that the training fold never contains;

since R2 is normalized by the within-group variance, the systematic offset drives the metric strongly negative. Re-running the original protocol with the original, unregularised hyper-parameters reproduces the group-aware values of the previous version of the performance table exactly: −25.87 for LightGBM and −11.25 to −12.42 for the other three algorithms, matching the range quoted there.

Alternative explanations for the negative values were tested and excluded. The share of the total target variance attributable to well type is only 5.6%, far too small to explain values of the order of −10, because the compression of the within-group variance originates from the horizontal-well population rather than from the directional one. The directional wells are not internally homogeneous either: the coefficient of variation of the target within that group is 0.465, and the 5th–95th percentile span covers a factor of 6.4. No single omitted variable raises the within-group R2 above 1.9%, and the engineering variables are not collinear enough to destabilize the fit. The negative values are therefore a property of the validation design rather than a failure of the model.

The protocol has accordingly been replaced by a scheme whose folds respect the structure of the field data. Under a well-type-stratified five-fold cross-validation, in which every fold preserves the horizontal/directional proportion, the out-of-fold R2 of the four tree ensembles is 0.51–0.55 (Table 6); a within-type variant, in which folds are built separately inside each well type, gives the same result, whereas the original random split and the well-type holdout are retained for comparison only (Figure 13b). Regularisation and early stopping were applied in all cases, which reduces the train–test gap of XGBoost from 0.54 to 0.22 while raising its out-of-fold R2 from 0.46 to 0.51 (Figure 13a). Cross-well-type extrapolation is explicitly outside the intended scope of the model, and the earlier statement that grouped validation is the only effective test of transferability has been removed.

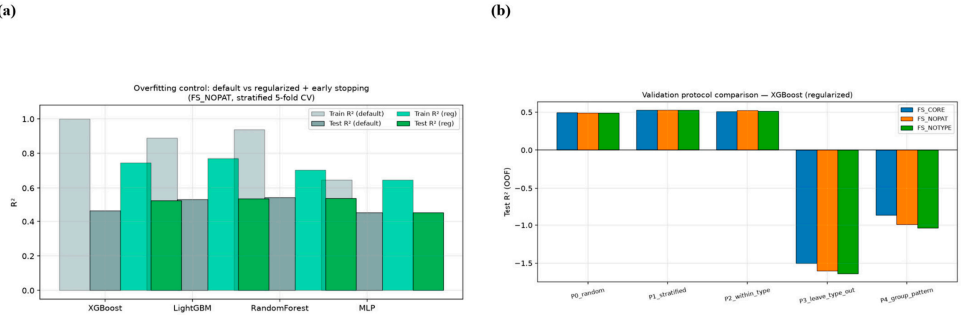


Figure 13. Validation-strategy sensitivity of the accepted model family. (**a**) Training and out-of-fold test R2 of the four algorithms under default and regularised hyper-parameters, showing the reduction of the train–test gap and the simultaneous improvement of the out-of-fold score; (**b**) out-of-fold test R2 of the regularised XGBoost model under five validation protocols for three feature sets, illustrating that only the protocols whose folds respect the well-type composition return a positive score, whereas holding out an entire well type (P3) or an entire well pattern (P4) does not.

3.9. Cross-Comparison with Conventional Reservoir-Analysis Methods

Full-scale numerical simulation was not available for this dataset, and the reasons are stated explicitly rather than left implicit. The primary input of a tight-oil production simulator is the hydraulic-fracture geometry—half-length, conductivity, and stimulated reservoir volume—whereas this dataset contains only treatment parameters (pump rate, stage count, fluid and proppant volume) and no microseismic or net-pressure inversion, so a fracture model cannot be calibrated. No compositional or geomechanical model has been calibrated for the block, so the dominant uncertainty of a simulation would come from its inputs rather than from the physics. Finally, the information required by a dynamic simulator—well history, pressure and fracture diagnostics—is precisely the post-development information

that the present framework deliberately excludes, so a simulator would answer a different question from the one addressed here.

The model was therefore cross-checked against three conventional methods on the same well set: Arps decline extrapolation fitted to the first 36 months of each well and used to forecast the 60-month cumulative; an analogy-based estimate using the ten nearest neighbouring wells of the same well type; and a volumetric/material-balance screening. The results are summarized in Table 10 and Figure 14. Arps extrapolation is the most accurate of the three (MAE 0.0291, R2 = 0.977), followed by the random-forest model (MAE 0.1206, R2 = 0.555), with the analogy method performing similarly to the machine-learning model (MAE 0.1263, R2 = 0.465).

Table 10. Cross-comparison of conventional reservoir-analysis methods and the machine-learning model on the Y284 wells. Arps extrapolation is fitted on the first 36 months of each well; the analogy method uses the ten nearest neighbouring wells of the same well type; the machine-learning entry is the out-of-fold result of the accepted model. The volumetric screening is a physical-plausibility check that rests on three assumptions (oil density, formation volume factor, and drainage-area definition) and is therefore excluded from the accuracy ranking.

Method	Information Required	*n*	MAE	RMSE	R2
Arps decline extrapolation (36-month fit)	Own 36-month production history	296	0.0291	0.0430	0.977
Arps extrapolation, increment of months 37–60 only	Own 36-month production history	296	0.0291	0.0430	0.765
Analogy method (k = 10 nearest neighbours)	Five-year history of 10 neighbouring wells	338	0.1263	0.1992	0.465
Machine-learning model (random forest, out-of-fold)	Pre-development decision variables only	338	0.1206	0.1817	0.555
Volumetric/material-balance screening	Static properties and pattern geometry	4 case wells	plausibility check only	—	—

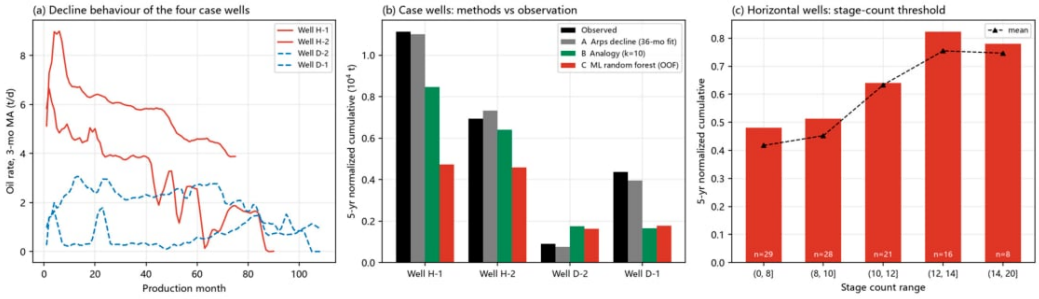


Figure 14. Case-well group and cross-comparison of conventional and machine-learning estimates. (**a**) Decline behaviour of the four selected wells; (**b**) observed five-year normalized cumulative production against the estimates of Arps decline extrapolation (36-month fit), the analogy method (k = 10 nearest neighbours) and the random-forest model (out-of-fold); (**c**) median five-year normalized cumulative production of the horizontal wells by stage-count range.

The comparison is presented as a division of roles rather than as a ranking. The decline method requires 36 months of the well's own production history and the analogy method requires five years of production from neighbouring wells; neither is available at the deployment-decision point at which the planning question is posed. The machine-learning model uses only pre-development decision variables and is the only one of the four available at that stage. In the opposite direction, the decline parameters produced by the conventional

analysis—decline type, first-year decline rate, response lag and breakthrough lag—are used as inputs to the engineering-diagnosis stage of the workflow, so the two families of methods are complementary rather than competing. The volumetric screening gives implied five-year recovery factors of 1.2–5.3%, inside the expected range for tight oil, and is used only as a physical-plausibility check.

Two properties of the comparison are reported deliberately. First, the apparent accuracy of the decline method is partly an artefact of the metric: the 36-month cumulative production is already contained in its input, so the quantity actually forecast is the 24-month increment, and when the comparison is restricted to that increment, the relative error of the decline method rises to 31.4%. The machine-learning model is unaffected by this argument because it never observes the well's own history. Second, the machine-learning model systematically under-predicts the best-performing horizontal wells—all 26 wells with an observed value above 0.8 have a negative residual, the largest being −0.877—which is the expected shrinkage towards the conditional mean at small sample size. The framework is therefore positioned as a screening and ranking tool rather than as an accurate single-well predictor.

3.10. Well-Type Deployment Bias Examined by Propensity-Score Analysis

The well-type statistics of Section 3.4 show an apparent productivity advantage of the horizontal wells, which could in principle be confounded by reservoir-quality selection. The premise was tested directly. In this dataset, the horizontal wells are not preferentially placed in the better-quality reservoir: their mean porosity is 9.30% against 11.80% for the directional wells (standardized mean difference −2.66), so the difference runs opposite to the usual expectation, while permeability, oil saturation, and net-pay thickness are statistically indistinguishable between the two groups ($|\mathrm{SMD}| \leq 0.09$). Well type is therefore assigned to materially different porosity ranges, and not in the direction of preferentially selecting higher-porosity zones.

A propensity-score model was then fitted with well type as the treatment and the four pre-treatment reservoir properties as covariates; post-treatment engineering variables such as stage count and fluid volume were deliberately excluded, because they are themselves determined by well type (10.5 stages for the horizontal wells against 1.4 for the directional wells, with no overlap) and controlling for them would absorb the effect under study. Porosity alone separates the two populations almost completely, so the propensity-score model attains an AUC of 0.970 and 135 of 338 wells (40%) fall outside the common support region; after nearest-neighbour matching the largest remaining |SMD| is 0.37–0.48, four to five times the conventional 0.1 threshold, and permeability deteriorates from +0.07 to −0.48 (Table 11). Propensity-score matching is therefore not feasible on this dataset, and this is reported as a result rather than replaced by an average treatment effect that the data cannot support (Figure 15).

Table 11. Covariate balance and common support of the propensity-score analysis for well type (horizontal wells = treatment, directional wells = control), computed on the 338 wells for which all four reservoir covariates are available. The standardized mean difference (SMD) is reported before and after nearest-neighbour matching with a caliper of 0.2.

Covariate	Mean, Horizontal (n = 104)	Mean, Directional (n = 233)	|SMD| Before Matching	|SMD| After Matching
Permeability (mD)	0.495	0.439	0.074	0.483
Porosity (%)	9.30	11.80	2.658	0.200
Oil saturation (%)	55.53	55.53	0.001	0.452
Net-pay thickness (m)	20.52	21.16	0.087	0.367
Propensity-score AUC = 0.970; wells outside the common support = 135 of 338 (40%); matched pairs = 69				

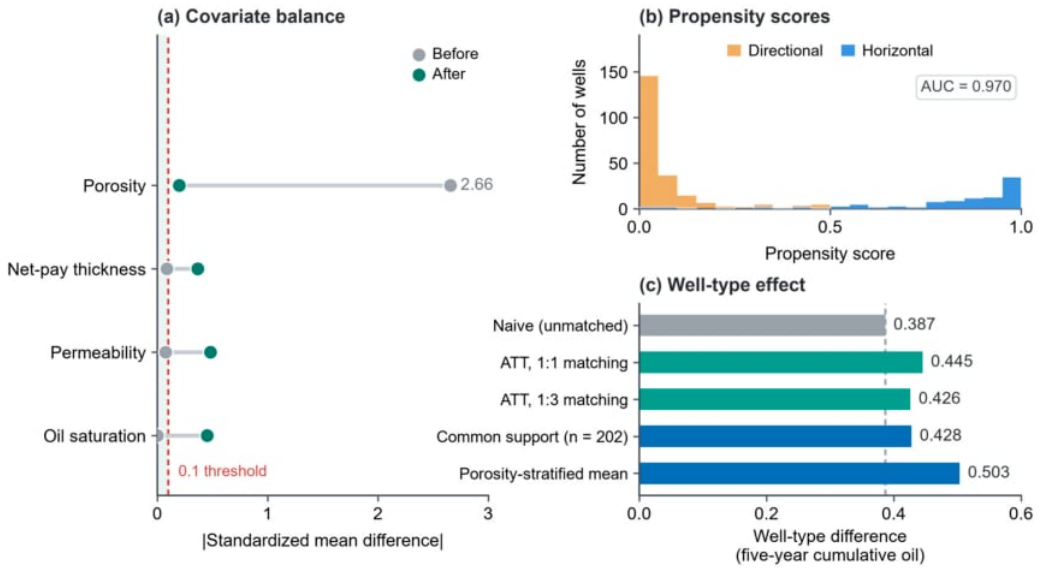


Figure 15. Propensity-score diagnostics for the well-type comparison (horizontal wells = treatment, directional wells = control; the 338 wells with a complete set of reservoir covariates). (**a**) Covariate balance before and after nearest-neighbour matching, expressed as the absolute standardized mean difference, with the conventional 0.1 threshold marked; (**b**) distribution of the estimated propensity score for the two well types, showing the limited overlap; (**c**) well-type difference in five-year normalized cumulative production estimated five ways—the unmatched naive difference, the average treatment effect on the treated (ATT) after 1:1 and 1:3 matching, the difference restricted to the common support region ($n = 202$) and the mean of the three porosity strata that contain both well types. Because matching does not restore balance (panel **a**), the two ATT values are reported as diagnostics only and are not used as the effect estimate; the accompanying text relies on the two conservative comparisons in the lower part of panel (**c**).

Two more conservative comparisons were used instead (Figure 15). Restricting the sample to the common support region ($n = 202$) leaves a well-type difference of +0.43 in normalized five-year cumulative production. Stratifying the sample by quintiles of porosity gives a positive difference in every stratum that contains both well types (+0.35, +0.46 and +0.71; stratum mean +0.50), including the balanced middle stratum (37 horizontal against 30 directional wells, difference +0.46). The horizontal-well advantage is therefore not an artefact of selective placement in high-porosity zones. Because well type is nevertheless not randomly assigned, the difference is not interpreted as a causal effect: the revised conclusion states that production is governed jointly by the well-type/well-pattern configuration and by the associated stimulation intensity, and that the two cannot be separated in this dataset. It is confirmed here that well type and well pattern are exactly collinear—all 234 directional wells are completed on the rhombic inverted nine-spot pattern, whereas the 104 horizontal wells are distributed over the five-spot (52), seven-spot (49) and nine-spot (3) patterns with no overlap—so well pattern has been removed from the feature set and the "well-pattern effect" is reported as part of the joint configuration rather than as an independent control.

3.11. Worked Example: From Data Input to Policy Output

To demonstrate the complete workflow, four representative wells were selected by an explicit and reproducible rule: complete pre-development decision variables, at least 60 months of production history (325 of the 338 wells qualify), and an established decline fit. The four wells form two controlled contrasts—two horizontal wells with comparable reservoir quality but different stimulation intensity (wells H-1, 15 stages, and H-2, 7 stages) and two directional wells with effectively identical single-stage completion (about

1.8–1.9 m^3/min) but different net pay (wells D-1, 26.4 m, and D-2, 16.3 m). To comply with the field-data confidentiality agreement, the wells are referred to by anonymized labels. The complete four-stage workflow for these wells is summarized in Figure 16.

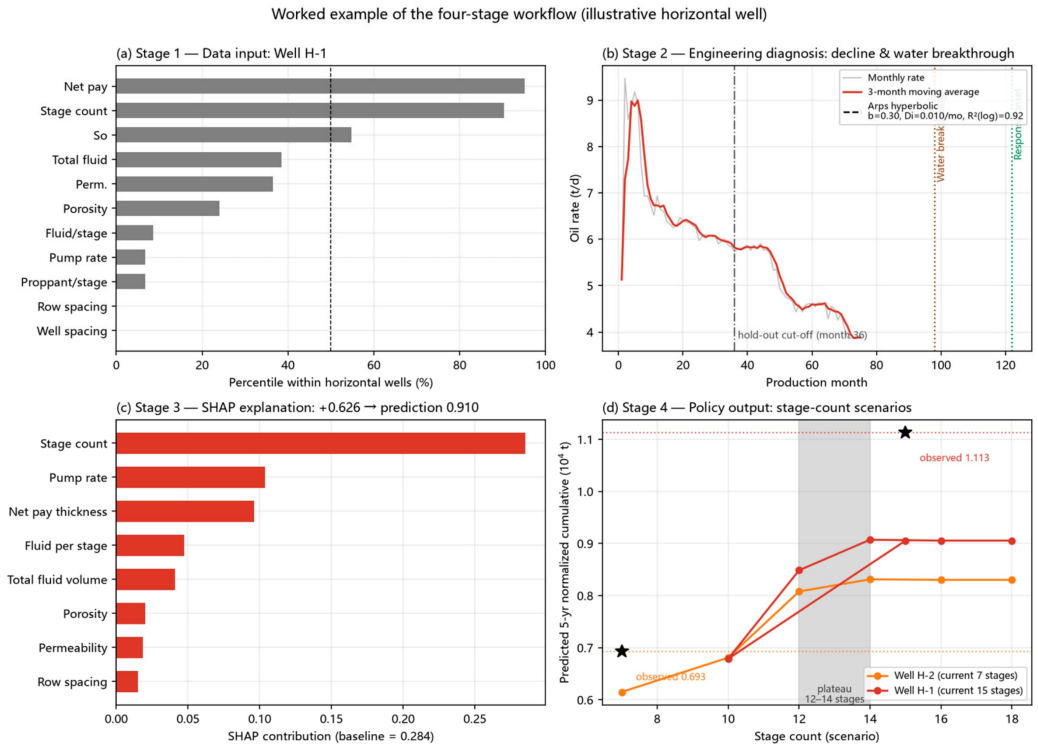


Figure 16. Worked example of the four-stage workflow, illustrated on well H-1. (**a**) Stage 1, data input: percentile rank of each pre-development decision variable within the horizontal-well population. The vertical dashed line at the 50th percentile marks the median benchmark of horizontal wells; (**b**) Stage 2, engineering diagnosis: monthly oil rate, three-month moving average, fitted Arps decline, water-breakthrough and waterflood-response markers, and the 36-month cut-off used in the cross-comparison; (**c**) Stage 3, SHAP decomposition from the model baseline; (**d**) Stage 4, policy output: stage-count scenarios for wells H-2 and H-1 with the observed values marked, the shaded band indicating the 12–14 stage plateau. The horizontal dotted lines mark the observed 5-year normalized cumulative production of Wells H-1 (1.113) and H-2 (0.693), providing a reference against the predicted scenario curves at varying stage counts.

Step 1—Data input. Well H-1 is used as the illustrative case. Its net pay (29.1 m) and stage count (15) lie at the 99th percentile of the horizontal-well population, whereas its pump rate (2.394 m^3/min) lies at the 10th percentile, and its porosity and permeability are below the field median. The input vector alone therefore already isolates the candidate limiting parameter.

Step 2—Engineering diagnosis. Applying the decline-fitting protocol of Section 2.2 together with water-breakthrough and waterflood-response identification gives the following. H-1 shows a slow decline (first-year decline 19.4%, R2 of the log-rate fit 0.969) with late water breakthrough (month 97) and a late response, that is, a healthy energy balance. H-2 shows a first-year decline of 39.5% with breakthrough in the first producing month, a slow response, and a fast water breakthrough. D-1 shows a fast response (7 months), a slow decline and the thickest pay of the group, and is the best-performing directional well. D-2 shows a first-year decline of 72.3%, water breakthrough within 6 months, and the thinnest pay, leaving no remaining engineering freedom.

Step 3—SHAP explanation. The accepted model, whose baseline output is 0.2838, attributes the prediction for H-1—which lies 0.626 above the baseline—mainly to stage

count (+0.285) and pump rate (+0.104) and only marginally to porosity (+0.020). For both directional wells, the negative contributions are dominated by stage count = 1 (−0.069 to −0.073) and by a pump rate of about 1.8–1.9 m^3/min (−0.044 to −0.048). The SHAP decomposition is thus consistent with the independent reservoir-engineering diagnosis of Step 2, which is the consistency check that the workflow is designed to enforce.

Step 4—Policy output. Binning the 104 horizontal wells by stage count gives median five-year normalized cumulative production of 0.481, 0.514, 0.640, 0.823 and 0.780 for the ranges (0, 8], (8, 10], (10, 12], (12, 14] and (14, 20] stages, that is, a clear plateau at 12–14 stages (Table 12). Correspondingly, the model predicts that raising H-2 from 7 to 12–14 stages would increase its five-year cumulative production from 0.615 to 0.808–0.831 (+31%), whereas adding stages to H-1, already at 15, produces no further gain. Because pump rate takes only a small number of discrete design values in this dataset (19 horizontal wells share a single value of 8.0 m^3/min), it is reported as a supporting condition rather than as an independent policy lever, and the stage count is used as the primary threshold. A block-wide application of the 12-stage plateau—the well-by-well list of horizontal wells lying below it together with a screening-level gain estimate—is reported in the Supplementary Materials rather than in the main text, because it is an inference from the model rather than a direct observation, and it is there explicitly labelled as a screening-level upper bound rather than as a production commitment.

Table 12. Stage-count policy thresholds for the horizontal wells of the block: median five-year normalized cumulative production by stage-count range.

Stage-Count Range	Median Five-Year Normalized Cumulative Production	Interpretation
(0, 8]	0.481	Below the plateau
(8, 10]	0.514	Below the plateau
(10, 12]	0.640	Below the plateau
(12, 14]	0.823	Plateau—recommended threshold for new designs
(14, 20]	0.780	No further gain; added stages fall into the interference zone

4. Discussion

4.1. Reservoir-Engineering Meaning of the Y284 Matching Problem

The Y284 case shows that development performance in a tight-oil block is essentially a matching problem among reservoir quality, pressure support, well-pattern geometry, and stimulation intensity. Static reservoir properties define the potential storage and seepage basis, but the realized five-year cumulative oil depends on whether the engineering system can effectively connect and maintain that potential. This explains why favorable porosity, permeability, or oil-layer thickness alone cannot guarantee high cumulative production. A well located in a favorable interval may still perform poorly if the injection-production relationship is incomplete, if row spacing is too large to maintain pressure, or if injected water advances through preferential pathways without sufficient oil displacement.

The dynamic diagnosis further indicates that Y284 cannot be optimized simply by increasing injection intensity. The coexistence of rapid water response and differentiated oil response suggests that some well groups have strong hydraulic communication but weak effective sweep. In this situation, water injection may quickly influence production wells but fail to build balanced pressure support across the target interval. Therefore, the field implication is to shift from volume-oriented injection evaluation to connectivity-oriented and sweep-oriented evaluation. Injection allocation, profile control, producer-injector

relation, and pattern completeness should be interpreted together with decline behavior and water-cut evolution.

The proposed screening logic is also rooted in waterflood reservoir engineering: areal pattern, pressure support, breakthrough timing, and well-test evidence should be interpreted jointly before an injection increase, spacing adjustment, or selective water-control action is recommended [56–59].

The interaction among pump rate, well spacing and stage count has a physical basis rather than a purely statistical one. Increasing the pump rate raises the net pressure and the induced-fracture complexity, which enlarges the stimulated reservoir volume and therefore the effective drainage area of a well. If the well spacing is not enlarged at the same time, the stimulated volumes of adjacent wells overlap, and the resulting inter-well interference converts part of the intended gain into accelerated pressure depletion rather than additional recovery. The stage count controls how the same total fluid and proppant volume are distributed along the lateral: too few stages concentrate the treatment into a small number of clusters and leave the remainder of the lateral effectively unstimulated, whereas beyond a certain stage density the incremental clusters fall into the interference zone of their neighbours and add little new drainage. The three parameters must therefore be matched to one another and to the well spacing, and the stress-coupled stimulation design of comparable tight reservoirs supports this reading; it also explains why they appear as a group in the SHAP decomposition rather than as independent controls. The field data are consistent with this chain: binning the 104 horizontal wells by stage count gives a clear plateau at 12–14 stages in the median five-year normalized cumulative production (Section 3.11), which is the empirical signature of a stage count that has reached the interference limit. Pump rate, by contrast, is not monotonic in these data—it takes a small number of discrete design values and is coupled with the commissioning batch and the well pattern—so it is reported as a supporting condition rather than as an independent policy lever. In this block, the spacing effect cannot be separated from the well-type effect, because well type and well pattern are exactly collinear (Section 3.10); this is stated as a limitation rather than resolved by the model.

4.2. Engineering-Constrained Modeling and Validation Requirements

A key methodological issue in five-year cumulative-oil prediction is the risk of data leakage [60–62]. Variables such as annual oil, produced liquid, water cut, decline type, injection-effect type, and water-response type are highly informative because they partly summarize the same production process used to define the target. If these variables are used without time-window control, the model may reproduce the answer rather than provide an operational prediction. The revised feature design therefore separates the pre-decision feature set from the post-hoc diagnostic feature set. The former is suitable for planning and early evaluation, whereas the latter is suitable for explaining why a well or well group has already performed well or poorly. This separation is the leakage-aware step of the engineering-constrained workflow defined in the Introduction, and it is applied consistently throughout the paper.

This distinction also affects how model performance should be reported. A high random-split R2 obtained with diagnostic variables should not be interpreted as evidence of pre-development predictive ability. For field deployment, model evaluation should include train, validation, and independent test metrics, repeated random seeds, predicted-versus-observed plots, residual distributions, and group-aware validation. The cross-validation scheme must therefore respect the structure of the field data. Because neighboring wells, wells in the same pattern and wells developed in similar periods may share geological and operational conditions, the folds used here preserve the horizontal/directional proportion

(Section 3.8); a holdout that removes an entire well type, by contrast, forces the model to extrapolate to a population it has never seen and is therefore reported separately rather than used as the acceptance criterion.

4.3. Well-Type and Well-Pattern Effects Under Sample-Selection Bias

The well-type statistics show an apparent productivity advantage of horizontal wells, but this result should be interpreted as an association rather than direct causality. Horizontal wells may have been preferentially deployed in better reservoir intervals, with longer controlled sections or stronger stimulation designs. Directional wells form the dominant sample population and may cover a broader range of reservoir conditions and development stages. Therefore, direct comparison among well types can be biased by reservoir selection, completion scale, and production-start time. This possibility has now been tested directly (Section 3.10). In this dataset, the horizontal wells are not preferentially placed in the better-quality reservoir—their mean porosity is lower, not higher—and a porosity-stratified comparison gives a positive well-type difference in every stratum containing both well types. Propensity-score matching itself is not feasible here, because porosity separates the two groups almost completely. Well type is nevertheless not randomly assigned, so the difference is reported as an association rather than as a causal effect.

The same caution applies to well-pattern conclusions. In the full-field population, the dominant rhombic inverted nine-spot pattern does not necessarily deliver the best cumulative-production response in the current statistics, while some smaller pattern categories show higher median production. However, small sample size and local reservoir differences may amplify apparent pattern effects. The engineering value of the present result is therefore not to declare one pattern universally superior, but to identify the conditions under which a pattern fails to match reservoir connectivity. Large row spacing, incomplete injection-production control, and excessive hydraulic communication can all weaken the conversion from reservoir quality to effective production.

4.4. From SHAP Interpretation to Development-Policy Optimization

The SHAP results provide a bridge between multivariable model output and reservoir-engineering interpretation. Across the model-ready sample, engineering communication intensity is repeatedly identified as the dominant factor group, with pump rate and total stage count the most important single variables. These variables correspond to three physical categories: geological storage capacity, injection-production control volume, and engineering communication intensity. This grouping is more useful for field application than a simple ranked list, because it helps engineers judge whether a low-efficiency well is mainly reservoir-limited, pressure-support-limited, pattern-limited, or stimulation-limited.

Policy recommendations should therefore be mapped from combined diagnostic states rather than from one variable. For favorable reservoir intervals with effective oil response and controlled water cut, the appropriate policy is stable pressure maintenance and controlled drawdown. For favorable reservoir intervals with weak oil response, the priority is to improve injection-production connectivity and pattern completeness. For wells with fast water breakthrough but limited oil increase, water-control and profile-control measures should be evaluated before increasing injection volume. For wells with large spacing or incomplete pattern control, infill screening, pattern adjustment, or producer-injector relationship correction should be considered. These rules turn the machine-learning explanation into a practical screening framework.

4.5. Limitations and Future Improvement

Several limitations should be emphasized. First, although the model-ready sample contains 338 wells of the block with valid five-year labels, the sample is small; the framework is therefore presented as a block-scale screening and ranking tool for planning-stage policy selection, and its results should be extrapolated with caution. The negative scores obtained when an entire well type is held out are a property of that validation design rather than a failure of the model (Section 3.8), and cross-well-type extrapolation lies outside the intended scope of the study. Second, some variables with strong engineering meaning, such as detailed pore-throat distribution, fracture geometry, pressure-transient indicators, and operating-system changes, still require broader and more consistent coverage. Third, SHAP values represent model-based associations. They can support engineering diagnosis and policy screening, but they do not prove causal effects unless combined with controlled field comparison, numerical simulation, or field adjustment verification. Fourth, the model-ready sample contains 104 horizontal and 234 directional wells, and well type and pattern form are strongly associated in this block because every directional well is completed on the rhombic inverted nine-spot pattern. The SHAP contributions of well type and well pattern therefore cannot be separated from each other, and the identified controls are valid only for the well-type and pattern classes represented in the sample.

Future work should strengthen the closed loop from data to decision. The first step is to lock the Y284 training database and report complete model-performance metrics under leakage-controlled and group-aware validation. The second step is to add EUR or other full-life-cycle indicators as targets, so that short-term productivity and long-term economic value can be evaluated together. The third step, already implemented in Section 3.6, is to translate the inflection zones of the SHAP dependence and interaction plots for the most important variables into field-level policy thresholds. Finally, representative well groups should be selected for before-and-after prediction or field verification, so that the proposed matching framework can evolve from an interpretation tool into an operational development-policy optimization tool.

5. Conclusions

The Y284 block was reconstructed as a tight-oil case study whose machine-learning stage is supported by a model-ready sample of 338 wells drawn from the block itself and covering both well types and the dominant pattern class. With $n = 338$ wells and nine candidate model families, the sample is small; the framework is therefore presented as a block-scale screening and ranking tool for planning-stage policy selection, and its results should be extrapolated with caution.

Reservoir-engineering diagnosis shows that Y284 has staged production and injection behavior, increasing water-cut pressure, dominant slow-decline wells, and a mismatch between fast water response and differentiated oil response. This explains why favorable reservoir quality does not automatically translate into favorable development effect.

Well-type analysis shows that horizontal wells have the strongest productivity and cumulative-production advantage, whereas directional wells form the main sample population. Well-pattern analysis shows that, in the full-field population, the dominant rhombic inverted nine-spot pattern is not necessarily the most effective pattern, whereas the model-ready sample covers the rhombic inverted nine-spot pattern as well; this highlights the need for spacing and connectivity optimization and limits pattern-level extrapolation. Because well type and well pattern are exactly collinear in this block—all directional wells are completed on the rhombic inverted nine-spot pattern, whereas the horizontal wells are distributed over three other patterns with no overlap—the two descriptors cannot be separated, and the corresponding conclusion is phrased in terms of the joint well-type/well-

pattern configuration together with the associated stimulation intensity rather than as a well-pattern effect.

Grouped by physical factor, the SHAP interpretation identifies engineering communication intensity (69.9% of total importance, led by total stage count and pump rate) as the dominant control, followed by geological storage capacity (21.6%) and injection-production control (8.6%). The results confirm that five-year cumulative oil is governed by the matching between reservoir-engineering factors rather than by a single geological or engineering variable.

A Y284 development-policy matching framework is proposed. It supports differentiated decisions for pressure maintenance, injection-production adjustment, water control, pattern optimization, stimulation-parameter review, and low-efficiency-well screening. The model is intended as a block-scale screening and ranking tool; its thresholds require recalibration for other blocks, and cross-block extrapolation is not claimed. Before field deployment, the framework should be calibrated using leakage-controlled model runs, well-type-stratified validation, and representative well-group verification.

Author Contributions: Conceptualization, H.Y. and X.A.; methodology, G.H. and H.Y.; software and data curation, Y.L.; validation, X.A. and Y.X.; formal analysis, X.F. and X.Y.; investigation, H.L. and J.W.; writing—original draft preparation, H.Y.; writing—review and editing, G.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the CNPC Changqing Oilfield Scientific Project "Re-evaluation of Development Performance and Study on Rational Development Modes for Chang 6 Tight Oil in Longdong Area" (No. 2025D4YP04).

Data Availability Statement: The field datasets used in this study are subject to confidentiality restrictions. Aggregated data, derived indicators, and non-confidential model outputs may be available from the corresponding author upon reasonable request and approval by the data owner.

Acknowledgments: The authors thank the field engineers and data-management staff who supported production, injection, geological, and engineering data collection.

Conflicts of Interest: Authors Huanying Yang, Xiaoping An, Youwei Xu, Hongchang Li, Jing Wang, Xiaohui Fan and Xueying Yu were employed by the PetroChina Changqing 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 PetroChina Changqing Oilfield Company had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

Abbreviation	Definition
DCA	Decline-curve analysis
EUR	Estimated ultimate recovery
FS-P	Pre-decision feature set
FS-D	Post-hoc diagnostic feature set
MAE	Mean absolute error
RMSE	Root mean square error
SHAP	Shapley additive explanations
MLP	Multilayer perceptron
PSM	Propensity-score matching
SMD	Standardized mean difference
SRV	Stimulated reservoir volume
CV	Cross-validation

References

1. Arps, J.J. Analysis of decline curves. *Trans. AIME* **1945**, *160*, 228–247. [CrossRef]
2. Duong, A.N. Rate-decline analysis for fracture-dominated shale reservoirs. *SPE Reserv. Eval. Eng.* **2011**, *14*, 377–387. [CrossRef]
3. 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. [CrossRef]
4. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
5. Fetkovich, M.J. Decline curve analysis using type curves. *J. Pet. Technol.* **1980**, *32*, 1065–1077. [CrossRef]
6. Ilk, D.; Rushing, J.A.; Perego, A.D.; Blasingame, T.A. Exponential vs. hyperbolic decline in tight gas sands: Understanding the origin and implications for reserve estimates using Arps decline curves. In Proceedings of the SPE Annual Technical Conference and Exhibition, Denver, CO, USA, 21–24 September 2008; SPE 116731. [CrossRef]
7. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
8. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [CrossRef]
9. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. LightGBM: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3146–3154.
10. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased boosting with categorical features. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 6638–6648.
11. Liu, C.; Feng, Q.; Zhou, W.; Li, S.; Zhang, X. Infill Well Location Optimization Method Based on Recoverable Potential Evaluation of Remaining Oil. *Energies* **2024**, *17*, 3492. [CrossRef]
12. Zhan, S.; Li, X.; Wang, Y.; Liu, J. Optimization of well pattern and spacing in low-permeability reservoirs considering pressure sensitivity. *Geofluids* **2022**, *2022*, 8215223. [CrossRef]
13. Chen, Z. *Reservoir Simulation: Mathematical Techniques in Oil Recovery*; SIAM: Philadelphia, PA, USA, 2007. [CrossRef]
14. Fang, J.; Jia, C.; Ran, Q.; Hui, G.; Zhou, X. The Study of Fracture in Volcanic Reservoir. In Proceedings of the International Oil and Gas Conference and Exhibition in China, Beijing, China, 8–10 June 2010; p. SPE-131813-MS. [CrossRef]
15. Zhu, Y.; Wang, J.; Li, M.; Zhang, Q. Machine-learning-assisted injection-production optimization for low-permeability reservoirs. *Energy Fuels* **2026**, *40*, 3525–3537. [CrossRef]
16. Hui, G.; Chen, S.; He, Y.; Wang, H.; Gu, F. Machine learning-based production forecast for shale gas in unconventional reservoirs via integration of geological and operational factors. *J. Nat. Gas. Sci. Eng.* **2021**, *94*, 104045. [CrossRef]
17. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4765–4774. [CrossRef] [PubMed]
18. Lundberg, S.M.; Erion, G.G.; Chen, H.; DeGrave, A.J.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [CrossRef] [PubMed]
19. Shapley, L.S. A value for n-person games. In *Contributions to the Theory of Games II*; Kuhn, H.W., Tucker, A.W., Eds.; Princeton University Press: Princeton, NJ, USA, 1953; pp. 307–317.
20. Hui, G.; Ren, Y.; Bi, J.; Wang, M.; Liu, C. Artificial Intelligence Applications and Challenges in Oil and Gas Exploration and Development. *Adv. Geo-Energy Res.* **2025**, *17*, 179–183. [CrossRef]
21. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
22. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: New York, NY, USA, 2009. [CrossRef]
23. Molnar, C. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, 2nd ed.; 2022. Available online: https://christophm.github.io/interpretable-ml-book/ (accessed on 1 October 2026).
24. Yu, Z.; Liu, X.; Chen, L.; Wang, P. CatBoost-SHAP-based productivity prediction and control-factor interpretation for shale-oil wells. *J. Earth Sci.* **2026**. [CrossRef]
25. Tadjer, A.; Hong, A.; Bratvold, R.B. Machine learning based decline curve analysis for short-term oil production forecast. *Energy Explor. Exploit.* **2021**, *39*, 1747–1769. [CrossRef]
26. Li, Y.; Wang, S.; Chen, Z.; Zhang, L. Development of decline curve analysis parameters for tight oil wells using a machine learning algorithm. *Geofluids* **2022**, *2022*, 8441075. [CrossRef]
27. Friedman, J.H. Stochastic gradient boosting. *Comput. Stat. Data Anal.* **2002**, *38*, 367–378. [CrossRef]
28. Biau, G.; Scornet, E. A random forest guided tour. *TEST* **2016**, *25*, 197–227. [CrossRef]
29. Strobl, C.; Boulesteix, A.L.; Zeileis, A.; Hothorn, T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinform.* **2007**, *8*, 25. [CrossRef] [PubMed]
30. Ribeiro, M.T.; Singh, S.; Guestrin, C. Why should I trust you? Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 1135–1144. [CrossRef]

31. Apley, D.W.; Zhu, J. Visualizing the effects of predictor variables in black box supervised learning models. *J. R. Stat. Soc. Ser. B* **2020**, *82*, 1059–1086. [CrossRef]
32. Hooker, G. Generalized functional ANOVA diagnostics for high-dimensional functions of dependent variables. *J. Comput. Graph. Stat.* **2007**, *16*, 709–732. [CrossRef]
33. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, Montreal, QC, Canada, 20–25 August 1995; pp. 1137–1143.
34. Varma, S.; Simon, R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinform.* **2006**, *7*, 91. [CrossRef] [PubMed]
35. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schroder, B.; Thuiller, W.; et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **2017**, *40*, 913–929. [CrossRef]
36. Bergmeir, C.; Hyndman, R.J.; Koo, B. A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Comput. Stat. Data Anal.* **2018**, *120*, 70–83. [CrossRef]
37. Varoquaux, G. Cross-validation failure: Small sample sizes lead to large error bars. *NeuroImage* **2018**, *180*, 68–77. [CrossRef] [PubMed]
38. Sterne, J.A.C.; White, I.R.; Carlin, J.B.; Spratt, M.; Royston, P.; Kenward, M.G.; Wood, A.M.; Carpenter, J.R. Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ* **2009**, *338*, b2393. [CrossRef] [PubMed]
39. van Buuren, S. *Flexible Imputation of Missing Data*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2018. [CrossRef]
40. Iglewicz, B.; Hoaglin, D.C. *How to Detect and Handle Outliers*; ASQ Quality Press: Milwaukee, WI, USA, 1993.
41. Li, H.; Misra, S.; He, J.; Elsworth, D. Hydrocarbon production dynamics forecasting using machine learning: A review. *Fuel* **2023**, *333*, 127067. [CrossRef]
42. Li, J.; Wang, F.; Zhu, W.; Ning, K.; Mu, L.; Yuan, G.; Hui, G. Physics-Constrained AI-Assisted Flowing Material Balance for Productivity Evaluation During High-Volume, Long-Duration Flowback in Ultra-Deep Wells. *Processes* **2026**, *14*, 2604. [CrossRef]
43. Li, Y.; Hui, G.; Zhang, Z.; Wang, M.; Li, J.; Pi, Z.; Ge, C.; Bao, P.; Zhang, Y.; Yang, X.; et al. Insights on Shale Gas Production Prediction via Combining Multimodal Machine Learning with Deterministic Geological Modeling. In Proceedings of the ADIPEC, Abu Dhabi, United Arab Emirates, 3–6 November 2025; p. D031S114R005.
44. Rosenbaum, P.R.; Rubin, D.B. The central role of the propensity score in observational studies for causal effects. *Biometrika* **1983**, *70*, 41–55. [CrossRef]
45. Stuart, E.A. Matching methods for causal inference: A review and a look forward. *Stat. Sci.* **2010**, *25*, 1–21. [CrossRef] [PubMed]
46. Pearl, J. *Causality: Models, Reasoning, and Inference*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2009. [CrossRef]
47. Hui, G.; Pi, Z.; Ge, C.; Li, Y.; Chen, Z.; Wu, D.; Lu, Y.; Ni, F.; Yang, X.; Zhang, Y.; et al. Managing Hydraulic Fracturing-Induced Seismicity in the Duvernay Shale: Triggering Mechanisms, Influencing Factors, and Practice-Oriented Mitigation Strategies. *Int. J. Coal Sci. Technol.* **2026**, *13*, 85. [CrossRef]
48. Onwunalu, J.E.; Durlofsky, L.J. A new well-pattern-optimization procedure for large-scale field development. *SPE J.* **2011**, *16*, 594–607. [CrossRef]
49. Zhang, L.; Sun, H.; Wang, Y.; Liu, K. Cooperative optimization of well spacing and fracturing parameters in tight reservoirs. *Pet. Sci.* **2026**, *23*, 5693–5705. [CrossRef]
50. Onwunalu, J.E.; Durlofsky, L.J. Application of a particle swarm optimization algorithm for determining optimum well location and type. *Comput. Geosci.* **2010**, *14*, 183–198. [CrossRef]
51. Jansen, J.D. Adjoint-based optimization of multi-phase flow through porous media: A review. *Comput. Fluids* **2011**, *46*, 40–51. [CrossRef]
52. Brouwer, D.R.; Jansen, J.D. Dynamic optimization of waterflooding with smart wells using optimal control theory. *SPE J.* **2004**, *9*, 391–402. [CrossRef]
53. Yousef, A.A.; Gentil, P.H.; Jensen, J.L.; Lake, L.W. A capacitance model to infer interwell connectivity from production and injection rate fluctuations. *SPE Reserv. Eval. Eng.* **2006**, *9*, 630–646. [CrossRef]
54. Cheng, C.; Chen, Z.-X.; Hui, G. Energy-Efficient Fracturing Based on Stress-Coupled Perforation. *Pet. Sci.* **2026**, *23*, 4873–4890. [CrossRef]
55. Dake, L.P. *Fundamentals of Reservoir Engineering*; Elsevier: Amsterdam, The Netherlands, 1978.
56. Willhite, G.P. *Waterflooding*; Society of Petroleum Engineers: Richardson, TX, USA, 1986.
57. Huang, W.; Wang, P.; Hui, G.; Kong, X.; Jia, Y.; Huang, L.; Bai, Y.; Pi, Z.; Li, Y.; Yao, F.; et al. Unconventional Fracture Networks Simulation and Shale Gas Production Prediction by Integration of Petrophysics, Geomechanics and Fracture Characterization. *Energies* **2024**, *17*, 5084. [CrossRef]
58. Craig, F.F. *The Reservoir Engineering Aspects of Waterflooding*; Society of Petroleum Engineers: Richardson, TX, USA, 1971.
59. Horne, R.N. *Modern Well Test Analysis: A Computer-Aided Approach*, 2nd ed.; Petroway: Palo Alto, CA, USA, 1995.

60. Ma, Y.; Ni, F.; Ma, J.; Qiu, W.; Hui, G. Dual-Prior-Constrained Temporal Inversion and FNO Surrogate-Model-Driven Coordinated Optimization of Drilling Engineering Parameters. *Processes* **2026**, *14*, 3036. [CrossRef]
61. Dwumah, S.; Mohammed, S. Multi-Output Machine Learning for Simultaneous Prediction of Oil, Gas, and Water Rates: Effect of Well Data Partitioning on Model Performance. In Proceedings of the SPE Nigeria Annual International Conference and Exhibition, Lagos, Nigeria, 3–5 August 2026; Paper SPE-234796-MS. [CrossRef]
62. Werneck, R.d.O.; Prates, R.; Moura, R.; Gonçalves, M.M.; Castro, M.; Soriano-Vargas, A.; Júnior, P.R.M.; Hossain, M.M.; Zampieri, M.F.; Ferreira, A.; et al. Data-driven deep-learning forecasting for oil production and pressure. *J. Pet. Sci. Eng.* **2022**, *210*, 109937. [CrossRef]

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.