energies

MDPI

Article

Gas Production Prediction Model of Volcanic Reservoir Based on Data-Driven Method

Haijie Zhang [1], Junwei Pu [1], Li Zhang [1], Hengjian Deng [1], Jihao Yu [2], Yingming Xie [3], Xiaochang Tong [4], Xiangjie Man [4] and Zhonghua Liu [2,4,*]

[1] Chong Qing Shale Gas Exploration and Development, Co., Ltd., Chongqing 401121, China; zhang_hj@petrochina.com.cn (H.Z.); pujunwei@petrochina.com.cn (J.P.); li_zhang@petrochina.com.cn (L.Z.); dlts_denghj@petrochina.com.cn (H.D.)
[2] School of Petroleum and Natural Gas Engineering, Chongqing University of Science and Technology, Chongqing 401331, China; 2022201054@cqust.edu.cn
[3] CNOOC EnerTech-Drilling & Production, Co., Beijing 100028, China; xieyingming56@163.com
[4] Pancasia Holding Co., Ltd., Chongqing 400000, China; cqlnytxc@163.com (X.T.); cqlnymxj@163.com (X.M.)
* Correspondence: liuzhonghua@cqust.edu.cn

Abstract: Based on on-site construction experience, considering the time-varying characteristics of gas well quantity, production time, effective reservoir thickness, controlled reserves, reserve abundance, formation pressure, and the energy storage coefficient, a data-driven method was used to establish a natural gas production prediction model based on differential simulation theory. The calculation results showed that the average error between the actual production and predicted production was 12.49%, and the model determination coefficient was 0.99, indicating that the model can effectively predict natural gas production. Additionally, we observed that the influence of factors such as reserve abundance, the number of wells in operation, controlled reserves, the previous year's gas production, formation pressure, the energy storage coefficient, effective matrix thickness, and annual production time on the annual gas production increases progressively as the F-values decrease. These insights are pivotal to a more profound understanding of gas production dynamics in volcanic reservoirs and are instrumental in optimizing stimulation treatments and enhancing resource recovery in such reservoirs and other unconventional hydrocarbon formations.

Keywords: production prediction model; volcanic reservoir; data-driven method; data nondimensionalization; dimension recovery

Citation: Zhang, H.; Pu, J.; Zhang, L.; Deng, H.; Yu, J.; Xie, Y.; Tong, X.; Man, X.; Liu, Z. Gas Production Prediction Model of Volcanic Reservoir Based on Data-Driven Method. *Energies* **2024**, *17*, 5461. https://doi.org/10.3390/en17215461

Academic Editors: Franco Berruti and Daoyi Zhu

Received: 3 September 2024
Revised: 30 October 2024
Accepted: 30 October 2024
Published: 31 October 2024

1. Introduction

Hydrocarbon production from unconventional reservoirs requires the integration of different technologies, including long lateral horizontal drilling and multi-stage, multi-cluster hydraulic fracture systems that activate natural fracture networks in unconventional formations [1,2]. Dark box approaches, which involve using data analytics techniques, have recently gained significant attention in many areas [3–5]. A variety of data analytics methods, such as machine learning [6], linear regression [7], and neural networks [8], are used for predicting gas and oil production.

Many scholars [9–13] from all over the world have applied linear regression methods in predicting gas and oil production. Zhou et al. [14] conducted a multiple regression model for 173 wells in the Marcellus Formation, predicting the one-year cumulative gas production as a function of the proppant mass, fracture fluid volume, number of stages, treatment rate, vertical depth, and lateral length. Grujic et al. [15] developed a predictive model for 172 wells in North America, forecasting oil, gas, and water production as a function of the volume, petrophysics, temperature, pressure, and geographical and completion parameters. Zhong et al. [16] predicted oil production in 476 wells in the Wolfcamp Formation as a function of the well and completion designs, using a multiple regression method and

comparing it with other methods. Lolon et al. [17] predicted the cumulative oil production as a function of the stage of cementing, percentage of ceramic proppant, stage spacing, proppant intensity, water cut, fracturing fluid, and maximum treatment rate, by comparing some machine learning methods. Khanal et al. [18] forecasted the gas rate, cumulative gas, and condensate-to-gas ratio (CGR) using linear regression analysis for 335 simulations and 46 wells in the Eagle Ford Formation. Xue et al. [19] predicted the dynamic shale gas production rate as a function of hydraulic fracturing and geological properties using a multi-objective random forest regression method. Their sensitivity analysis revealed that the most influential parameters were the geological properties, including initial pressure and formation thickness. Johan et al. [20] employed a genetic algorithm to optimize completion designs, using several predictor variables, including the depth, lateral length, azimuth, total fluid, fluid intensity, total proppant, proppant intensity, and additional engineering features to account for the influence of neighboring wells and depletion.

The overall goal of this study is to develop a time variation characteristic prediction model that can accurately predict gas production in volcanic gas reservoirs. Specifically, this study aims to address the limitations of existing prediction models in processing the dynamic production data of unconventional gas reservoirs, which typically require a large amount of geological and permeability data. Traditional prediction methods require extensive numerical simulations. To overcome these challenges, we derived a new prediction model based on differential simulation theory and fitted the model using dimensionless gas production data obtained from volcanic gas reservoirs. In addition, we also calculated the error, coefficient of determination (R^2), and F-value between actual and predicted gas production to verify the effectiveness and accuracy of the model. Through this study, we hope to provide a new and more effective tool for predicting gas production in unconventional gas reservoirs, thereby providing a scientific basis for a development strategy and production parameter adjustments in oil and gas fields.

2. Methodology

2.1. Workflow

Data-driven prediction, a method grounded in using existing data for analyzing and modeling to forecast future trends, outcomes, or events, is utilized in this paper to develop a production forecast model with time-varying characteristics. The methodology is outlined in a flow chart, depicted in Figure 1, and involves four sequential steps: initial data preprocessing, which entails the collection and normalization of gas field data to neutralize dimensional influences, followed by two stages of accumulation to reduce historical data randomness and prepare the data for modeling; the subsequent establishment of a multiple linear regression model based on a differential equation; parameter estimation using the least squares method to ensure predictive accuracy; and finally, data dimension recovery following two reduction processes to facilitate the computation of projected gas production. This comprehensive approach ensures the creation of a precise and reliable prediction model, essential in the accurate forecasting of gas field production.

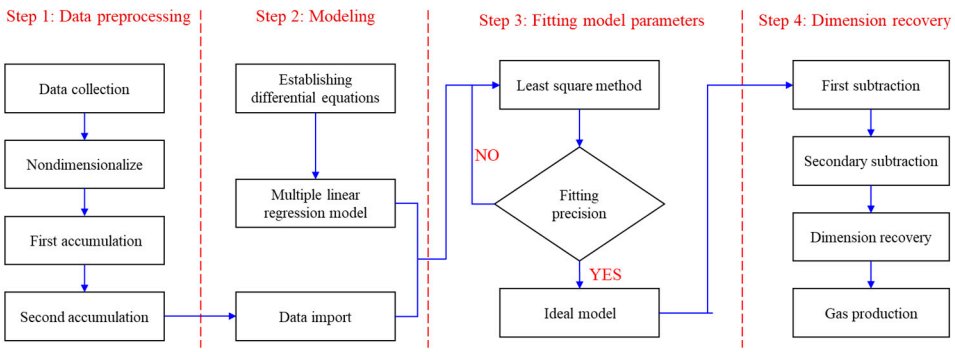

Figure 1. Flow chart of data-driven prediction.

2.2. Data Preprocessing

2.2.1. Data Collection

The indices for gas production data in volcanic gas reservoirs are bifurcated into two primary categories: the prediction index and the main control factor index. The prediction index is $Q(t)$ (gas production), and the main control factor index mainly contains U_1(t) (number of wells in operation), $U_2(t)$ (production time), $U_3(t)$ (effective thickness of reservoir), $U_4(t)$ (control reserves), $U_5(t)$ (reserve abundance), $U_6(t)$ (formation pressure), and $U_7(t)$ (energy storage factor). Then, all the data indexes are collected in an Excel sheet.

2.2.2. Data Nondimensionalization

To standardize the data and eliminate the discrepancies arising from different physical units among various parameters, each data index should be normalized using the following equation:

$$x^{(0)} = \frac{x}{\sum_1^n x} \tag{1}$$

where $x^{(0)}$ is the dimensionless value; x is the raw data of the gas field data; and n is the number of data.

2.2.3. Data Accumulation

To mitigate the randomness inherent to historical data and enhance the stability and reliability of the analysis, the dimensionless data are subjected to a first accumulation process. This is typically accomplished using the following equation:

$$x_i^{(1)} = \sum_1^i x^{(0)} \tag{2}$$

where $x_i^{(1)}$ is the dimensionless value after the first accumulation; x is the value before normalization; and n is the number of data.

Then, the dimensionless data after the first accumulation can be accumulated for the second time to form a fitting sample library as the following equation:

$$x_i^{(2)} = \sum_1^i x^{(1)} \tag{3}$$

2.3. Modeling

Differential simulation theory, a specialized data-driven approach, posits that the accumulated time series data exhibit exponential variation characteristics, a property that can be mathematically substantiated [21]. Leveraging this theory, a yield data-driven model can be formulated to capture the underlying trends and dynamics of the data. The model development typically involves the following steps:

$$\left[Q^{(2)}(t)\right]' = aQ^{(2)}(t) + BU^{(2)}(t) \tag{4}$$

where $\left[Q^{(2)}(t)\right]'$ represents the derivative with respect to time t; $U^{(2)}(t) = \begin{pmatrix} U_1^{(2)}(t) \\ U_2^{(2)}(t) \\ \cdots \\ U_3^{(2)}(t) \end{pmatrix}$; a and B are the non-identified parameters, and B is shown as follows:

$$B = (B_1, B_2, B_3, \ldots, B_7)$$

The above output data-driven Equation (4) is discretized by first-order approximation over time as follows:

$$Q_{k+1}^{(2)} = aQ^{(2)}(t) + BU^{(2)}(t) \tag{5}$$

Then, this Equation can be converted to the following equation:

$$Q_{k+1}^{(2)} = aQ_k^{(2)} + B_1U_{1k+1}^{(2)} + B_2U_{2k+1}^{(2)} + \ldots + B_7U_{7k+1}^{(2)} \quad (6)$$

The derived parameters a and B, along with the flowchart detailing the least squares fitting process, are depicted in Figure 2. The dimensionless data, following the second accumulation, were fed into Equation (6) for comprehensive multiple linear regression analysis. Employing the least squares method, we calculated the sum of squared residuals between the actual observed values and those predicted by the model. This approach was instrumental in determining the optimal parameter values, thereby enhancing the model's predictive accuracy and reliability.

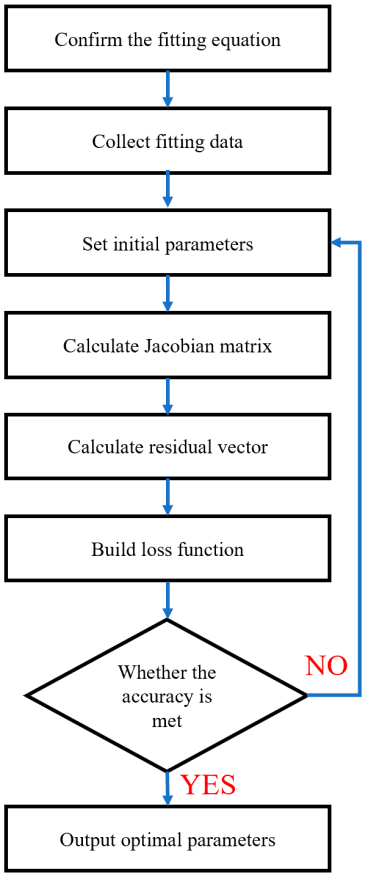

Figure 2. Flow chart of the least squares fitting.

2.4. Dimension Recovery

The predicted gas production is derived after two subtractions, utilizing the fitted equation. Specifically, the model is applied with the output data from year $k+1$ and the main control factors for year $k+1$. Thereafter, the dimensionless production data for year $k+1$, after two accumulations, are computed. Ultimately, the predicted gas production is ascertained through a series of two subtractions and dimension recovery processes, thereby yielding a precise forecast.

The first subtraction can be calculated as follows:

$$Q_{k+1}^{(1)} = Q_{k+1}^{(2)} - Q_k^{(2)} \quad (7)$$

where $Q_{k+1}^{(1)}$ is the production dimensionless data in $k+1$ after the first accumulation; $Q_{k+1}^{(2)}$ is the production dimensionless data in k after the second accumulation; and $Q_k^{(2)}$ is the production dimensionless data in k after the second accumulation.

The second subtraction can be calculated as follows:

$$Q_{k+1}^{(0)} = Q_{k+1}^{(1)} - Q_k^{(1)} \tag{8}$$

where $Q_{k+1}^{(0)}$ is the production dimensionless data in $k+1$; $Q_k^{(1)}$ is the production dimensionless data in k after the first accumulation.

The predicted gas production can be obtained as follows:

$$Q_{k+1} = Q_{k+1}^{(0)} \sum_1^n Q_n \tag{9}$$

where Q_{k+1} is the production data in $k+1$; $\sum_1^n Q_n$ is the sum of the gas production data.

3. Field Example

3.1. Data Preprocessing

The lithology of the targeted volcanic gas reservoir predominantly comprises acid tuff breccias, rhyolite, and andesite. The internal interlayering within the gas-bearing strata of the main rock mass is underdeveloped, with an interlayer density of merely 0.03 m/m, and the individual gas layers exhibit substantial thickness. A tuffaceous breccia septum, approximately 13 m thick, is present in the upper section of the gas layer, while a more substantial septum, about 170 m thick, is observed in the lower part. The fractures within the lower septum are relatively well developed, constituting 50.5% of the septum's total thickness. The porosity within the reservoir fluctuates between 7.1% and 22.2%, with an average value of 14.4%. The permeability varies widely, from 0.005 mD to 836.000 mD, averaging at 0.844 mD, indicating a heterogeneous subsurface environment.

The data indices for the targeted volcanic rock well area were categorized into two distinct groups: the prediction index, which represents gas production, and the main control factor index, encompassing the number of wells in operation, production time, effective thickness of the reservoir, controlled reserves, reserve abundance, formation pressure, and energy storage factor. Subsequently, all index data were meticulously organized in annual chronological order, as presented in Table 1.

Table 1. Annual natural gas production and main control factor index table.

Time /Year	$Q(t)$ /$\times 10^8$ m^3	$U_1(t)$ /Well	$U_2(t)$ /$\times 10^3$ h	$U_3(t)$ /m	$U_4(t)$ /$\times 10^8$ m^3	$U_5(t)$ /$\times 10^8$ m^3/km^2	$U_6(t)$ /MPa	$U_7(t)$ /-
2008	0.05	3	0.89	91.30	18.4	15.51	46.61	9.43
2009	1.15	8	26.65	94.64	22.6	15.41	45.05	9.46
2010	1.94	11	50.49	86.99	25.2	15.25	43.49	8.81
2011	1.96	13	57.51	80.92	32.3	15.09	41.92	8.31
2012	1.77	15	68.45	81.56	36.0	14.94	40.36	7.83
2013	2.27	17	85.78	86.77	43.3	14.75	38.80	9.01
2014	2.35	15	86.99	91.13	69.3	14.56	37.30	9.83
2015	2.71	17	96.56	89.47	77.1	14.33	35.90	9.61
2016	2.5	20	98.74	89.93	82.9	14.12	34.50	9.95
2017	2.66	21	107.23	86.47	90.1	13.90	33.10	9.38
2018	2.39	21	97.42	87.00	95.7	13.70	31.80	9.50

In accordance with Equation (1), the original data presented in Table 1 were subjected to dimensionless processing, and the outcomes are detailed in Table 2. The distribution of these dimensionless data points is graphically represented in Figure 3. The maximum values of the dimensionless data range from 0.089 to 0.144, indicating a variation of 38.19%. Conversely, the minimum values span from 0.001 to 0.077, reflecting a substantial change of 98.70%. It is evident that there are considerable disparities among the various data indices, highlighting the importance of normalization in comparative analyses.

Table 2. Data after dimensionless processing.

Time/Year	$Q^0(t)$	$U_1^0(t)$	$U_2^0(t)$	$U_3^0(t)$	$U_4^0(t)$	$U_5^0(t)$	$U_6^0(t)$	$U_7^0(t)$
2008	0.002	0.017	0.001	0.087	0.027	0.089	0.101	0.085
2009	0.047	0.044	0.030	0.090	0.033	0.088	0.098	0.086
2010	0.079	0.061	0.057	0.083	0.036	0.087	0.095	0.080
2011	0.080	0.072	0.064	0.077	0.047	0.086	0.091	0.075
2012	0.072	0.083	0.077	0.077	0.052	0.085	0.088	0.071
2013	0.093	0.094	0.096	0.082	0.062	0.084	0.084	0.081
2014	0.096	0.083	0.097	0.087	0.100	0.083	0.081	0.089
2015	0.111	0.094	0.108	0.085	0.111	0.082	0.078	0.087
2016	0.102	0.110	0.110	0.085	0.120	0.081	0.075	0.090
2017	0.109	0.116	0.120	0.082	0.130	0.079	0.072	0.085
2018	0.098	0.116	0.109	0.083	0.138	0.078	0.069	0.086
2019	0.110	0.110	0.131	0.083	0.144	0.077	0.066	0.086

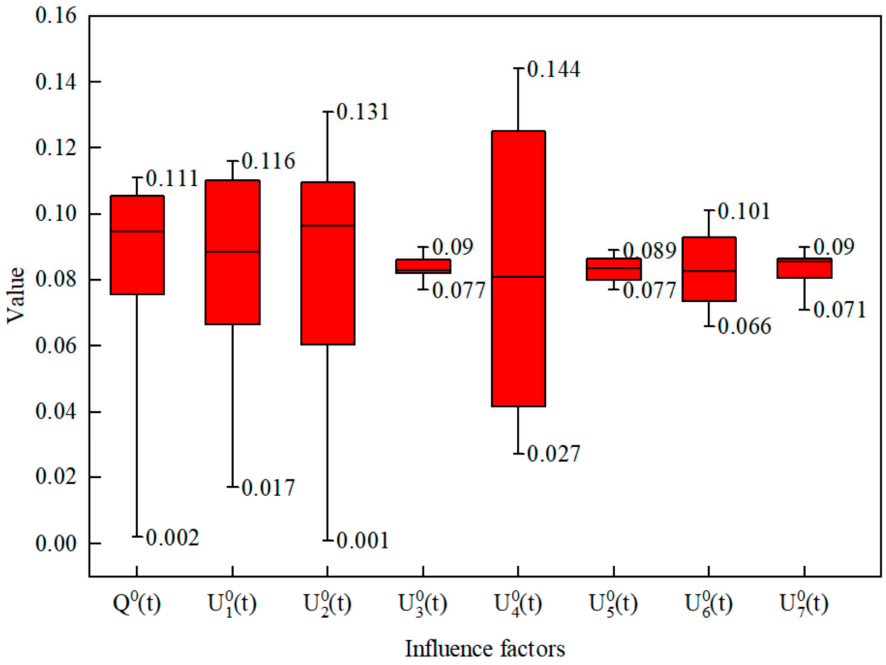

Figure 3. The distribution of the dimensionless data.

Utilizing Equation (2), the original data were subjected to the first accumulation, with the outcomes documented in Table 3. The distribution of the dimensionless data after the first accumulation is depicted in Figure 4. Notably, the maximum values of the dimensionless data after this initial accumulation uniformly reach 1, signifying no variation. The minimum values, on the other hand, range from 0.001 to 0.101, representing a reduction of 99.01%. This observation underscores that the variability in the dimensionless data following the first accumulation is markedly diminished compared to the initial dimensionless data set, thereby illustrating the effectiveness of the accumulation process in stabilizing data fluctuations.

In advance of the modeling phase, the dimensionless data following the first accumulation were further accumulated using Equation (3), with the results presented in Table 4. The distribution of the dimensionless data subsequent to the second accumulation is illustrated in Figure 5. The maximum values of the dimensionless data after this second accumulation range from 0.089 to 0.144, demonstrating a variation of 38.19%. The minimum values span from 0.001 to 0.077, indicating a substantial change of 98.70%. These refined data values, exhibiting reduced variability, are now suitable for inclusion in the modeling process.

Table 3. Data after the first cumulative processing.

Time/Year	$Q^1(t)$	$U_1^1(t)$	$U_2^1(t)$	$U_3^1(t)$	$U_4^1(t)$	$U_5^1(t)$	$U_6^1(t)$	$U_7^1(t)$
2008	0.002	0.017	0.001	0.087	0.027	0.089	0.101	0.085
2009	0.049	0.061	0.031	0.177	0.059	0.177	0.200	0.171
2010	0.128	0.122	0.087	0.259	0.096	0.264	0.294	0.250
2011	0.209	0.193	0.152	0.336	0.142	0.350	0.385	0.326
2012	0.281	0.276	0.228	0.413	0.194	0.435	0.473	0.396
2013	0.374	0.370	0.324	0.496	0.257	0.519	0.558	0.478
2014	0.470	0.453	0.422	0.582	0.357	0.603	0.639	0.567
2015	0.581	0.547	0.530	0.667	0.468	0.684	0.717	0.654
2016	0.683	0.657	0.640	0.753	0.588	0.765	0.792	0.744
2017	0.792	0.773	0.760	0.835	0.718	0.844	0.864	0.829
2018	0.890	0.890	0.869	0.917	0.856	0.923	0.934	0.914
2019	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

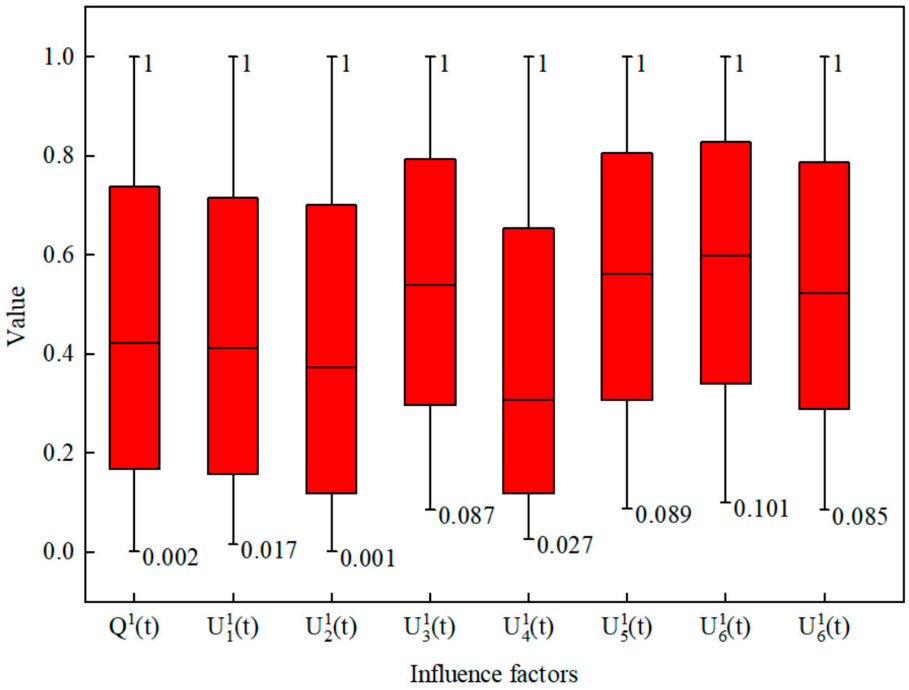

Figure 4. The distribution of the dimensionless data after the first accumulation.

Table 4. Data after the second cumulative processing.

Time/Year	$Q^2(t)$	$U_1^2(t)$	$U_2^2(t)$	$U_3^2(t)$	$U_4^2(t)$	$U_5^2(t)$	$U_6^2(t)$	$U_7^2(t)$
2008	0.002	0.017	0.001	0.087	0.027	0.089	0.101	0.085
2009	0.051	0.077	0.032	0.263	0.086	0.265	0.301	0.256
2010	0.180	0.199	0.119	0.522	0.181	0.529	0.595	0.507
2011	0.388	0.392	0.271	0.858	0.323	0.879	0.981	0.832
2012	0.669	0.669	0.499	1.272	0.518	1.314	1.454	1.229
2013	1.043	1.039	0.823	1.768	0.774	1.833	2.012	1.707
2014	1.514	1.492	1.245	2.350	1.131	2.436	2.651	2.273
2015	2.095	2.039	1.775	3.017	1.599	3.120	3.368	2.927
2016	2.778	2.696	2.415	3.770	2.186	3.886	4.160	3.671
2017	3.570	3.470	3.175	4.605	2.904	4.730	5.025	4.499
2018	4.460	4.359	4.044	5.522	3.759	5.653	5.958	5.414

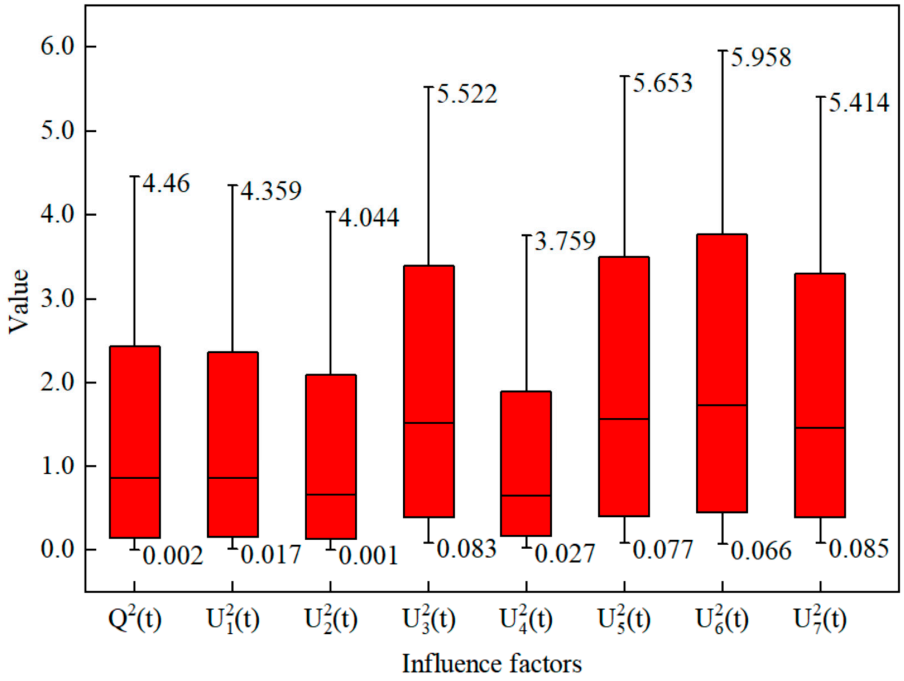

Figure 5. The distribution of the dimensionless data after the second accumulation.

3.2. Modeling

The dimensionless data after the second accumulation are input into Equation (6), then the a and B parameters are obtained after multiple linear regression simulation training by using the least squares method, and the results are shown below:

$$\begin{aligned} Q_{k+1}^{(2)} = {} & -0.514Q_k^{(2)} + 0.147U_1^{(2)}{}_{k+1} + 1.631U_2^{(2)}{}_{k+1} - 14.301U_3^{(2)}{}_{k+1} \\ & +0.266U_4^{(2)}{}_{k+1} + 0.441U_5^{(2)}{}_{k+1} + 4.995U_6^{(2)}{}_{k+1} \\ & +8.314U_7^{(2)}{}_{k+1} \end{aligned} \tag{10}$$

The gas production data and predicted values for the years 2008 to 2019, along with the projected production for 2020 to 2024, are graphically represented in Figure 6. A clear observation is that the predicted production aligns closely with the actual production across different years, indicating that the gas field reached a stable production phase by 2019. Additionally, Figure 7 illustrates the discrepancies between the actual and predicted gas production. The prediction error ranges from 0.46% to 30.47%, with an average error of 12.49%. It is evident that the error tends to diminish progressively with the extension of the mining period and the accumulation of production data, suggesting an improvement in the predictive accuracy over time.

3.3. Modeling Evaluation

3.3.1. Coefficient of Determination

The coefficient of determination (R^2) is a commonly used statistical indicator to evaluate the fit between the regression model and observed data [14–16]. It obviously indicates that the fitting model can explain the proportion of observed data variance, with a range of values between 0 and 1. The higher the R^2, the better the fit, and it can be calculated as follows:

$$\widetilde{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{11}$$

$$SST = \sum_{i}\left(y_i - \widetilde{y}\right)^2 \tag{12}$$

$$SSR = \sum_i (y_i - y_i')^2 \tag{13}$$

$$R^2 = 1 - \frac{SSR}{SST} \tag{14}$$

where $\widetilde{y}$ is the average of observed value; y_i' is the predicted value; y_i is the true value; SST is the sum of squares in the real data; and SSR is the sum of residual squares.

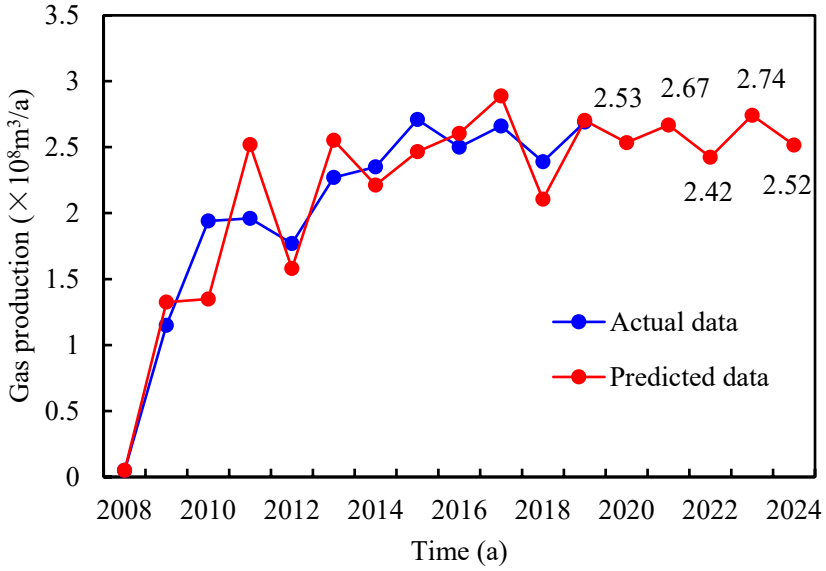

Figure 6. The gas production and predicted data.

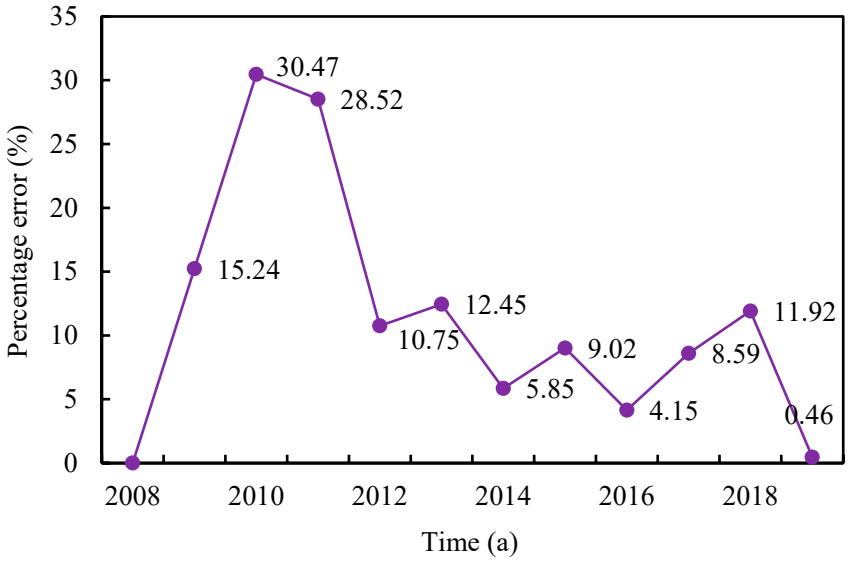

Figure 7. The error between gas production and predicted data.

Utilizing the gas production and predicted data from 2009 to 2019, the coefficient of determination was calculated to be 0.99, as per Equations (11)–(14). This high value for the coefficient of determination signifies that the gas production data are closely aligned with the predictions of the fitting model, thereby confirming the model's efficacy in capturing the production trends over the specified period.

3.3.2. Significance Testing

According to the relevant data of the established multiple linear regression fitting model, the significance of the regression equation is tested [22–25].

Hypothesis: $H_{0i}: \beta_i = 0, i\epsilon\{1, 2, \cdots, 12\}$, the sum of squared deviations are calculated as follows:

$$S_T^2 = \sum_i^n (y_i - \overline{y}) \tag{15}$$

$$S_R^2 = \sum_i^n (\hat{y} - \overline{y}) \tag{16}$$

$$S_E^2 = S_T^2 - S_R^2 \tag{17}$$

$$F = \frac{S_R^2/12}{S_R^2/12} \tag{18}$$

F-value is a statistical measure of analysis of variance used to test whether the regression equation is significant [26–30]. Taking a significance level of $F = 0.05$, the distribution table shows $F_{1-\alpha}(k, n-k-1) = F_{0.95}(729, 1974) = 1 < 19.249$; thus, the hypothesis should be rejected.

The *F*-values for $Q(t)$ (gas production), $U_1(t)$ (number of wells in operation), $U_2(t)$ (production time), $U_3(t)$ (effective thickness of reservoir), $U_4(t)$ (controlled reserves), $U_5(t)$ (reserve abundance), $U_6(t)$ (formation pressure), and $U_7(t)$ (energy storage factor) are graphically represented in Figure 8. A clear trend is evident, with the *F*-values arranged in descending order as follows: $U_2(t)$, $U_3(t)$, $U_7(t)$, $U_6(t)$, $Q(t)$, $U_4(t)$, $U_1(t)$, and $U_5(t)$. This ranking underscores the increasing influence of the annual production time, effective thickness of the reservoir, energy storage factor, formation pressure, gas production, controlled reserves, number of wells in operation, and reserve abundance on the annual gas production. The visualization of these *F*-values provides a quantitative assessment of the relative significance of each factor in the predictive model [31].

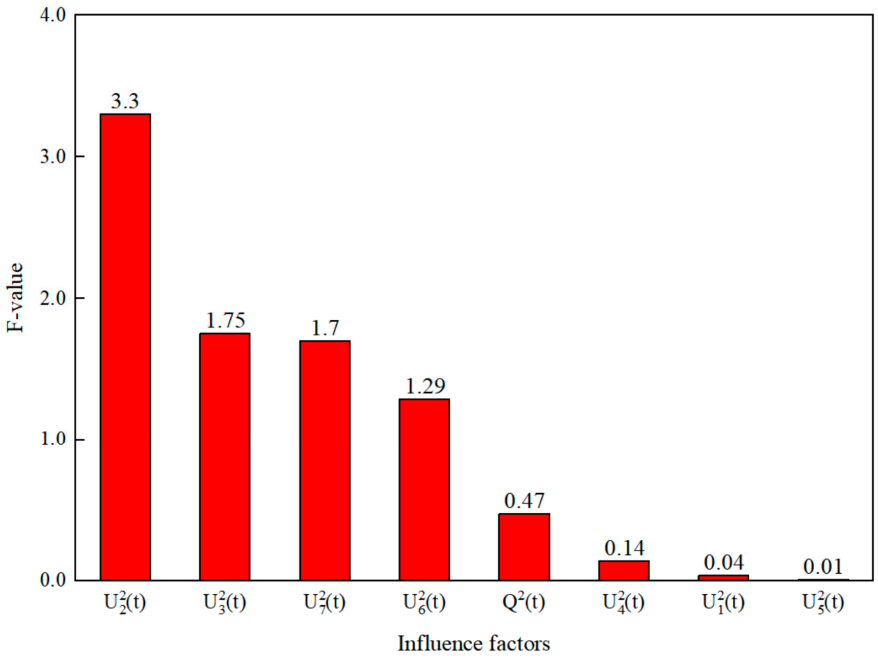

Figure 8. F-values of various influencing factors.

4. Conclusions

In this study, standardization ensures data consistency and comparability, enhancing data quality for analysis. Cumulative operations reduce random fluctuations, highlighting long-term trends for better model predictions. The application of two accumulations of dimensionless gas production data effectively mitigates the inherent randomness of historical data, thereby establishing a robust fitting sample library.

Based on on-site construction experience, considering the time-varying characteristics of the gas well quantity, production time, effective reservoir thickness, controlled reserves, reserve abundance, formation pressure, and energy storage coefficient, a data-driven method was used to establish a natural gas production prediction model based on differential simulation theory. The calculation results showed that the average error between the actual production and predicted production was 12.49%, and the model determination coefficient was 0.99, indicating that the model can effectively predict natural gas production.

The *F*-values in descending order are $U_2(t)$, $U_3(t)$, $U_7(t)$, $U_6(t)$, $Q(t)$, $U_4(t)$, $U_1(t)$, and $U_5(t)$, indicating that the influence of the reserve abundance, number of wells in operation, controlled reserves, previous year's gas production, formation pressure, energy storage coefficient, effective thickness of matrix, and annual production time on the annual gas production gradually increases.

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

Funding: This research was supported by the Science and Technology Research Program of Chongqing Municipal Education Commission (Grant No. KJZD-M202201502).

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: Authors Haijie Zhang, Junwei Pu, Li Zhang and Hengjian Deng were employed by the company Chong Qing Shale Gas Exploration and Development Company Limited; Author Yingming Xie was employed by the company CNOOC EnerTech-Drilling & Production Co.; Authors Xiaochang Tong, Xiangjie Man and Zhonghua Liu were employed by the company Pancasia Holding Company of Limited Liability. The 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.

References

1. Rahmanifard, H.; Gates, I. A Comprehensive review of data-driven approaches for forecasting production from unconventional reservoirs: Best practices and future directions. *Artif. Intell. Rev.* **2024**, *57*, 213. [CrossRef] [PubMed]
2. Liu, Z.; Bai, B.; Ding, Z.; Qu, H.; Zeng, S.; Da, X. Impact of cleanup additive on methane desorption on Longmaxi shale. *Fuel* **2021**, *300*, 121003. [CrossRef]
3. Syed, F.I.; Alnaqbi, S.; Muther, T.; Dahaghi, A.K.; Negahban, S. Smart shale gas production performance analysis using machine learning applications. *Pet. Res.* **2022**, *7*, 21–31. Available online: https://doaj.org/article/cb883d66b51743e89790c0462cd50902 (accessed on 22 August 2024). [CrossRef]
4. Mishra, S.; Lin, L. Application of data analytics for production optimization in unconventional reservoirs: A critical review. In Proceedings of the SPE/AAPG/SEG Unconventional Resources Technology Conference, Austin, TX, USA, 24–26 July 2017. [CrossRef]
5. Mohaghegh, S.D. Reservoir modeling of shale formations. *J. Nat. Gas Sci. Eng.* **2013**, *12*, 22–33. [CrossRef]
6. Li, D.; You, S.; Liao, Q.; Sheng, M.; Tian, S. Prediction of Shale Gas Production by Hydraulic Fracturing in Changning Area Using Machine Learning Algorithms. *Transp. Porous Med.* **2023**, *149*, 373–388. [CrossRef]
7. Mohammadmoradi, P.; Moradi, H.M.; Kantzas, A. Data-driven production forecasting of unconventional wells with apache spark. In Proceedings of the SPE Western Regional Meeting, Garden Grove, CA, USA, 22–26 April 2018. [CrossRef]
8. Enyioha, C.; Ertekin, T. Performance prediction for advanced well structures in unconventional oil and gas reservoirs using artificial intelligent expert systems. In Proceedings of the SPE Annual Technical Conference and Exhibition, San Antonio, TX, USA, 9–11 October 2017. [CrossRef]
9. Wang, S.; Chen, S. Insights to fracture stimulation design in unconventional reservoirs based on machine learning modeling. *J. Pet. Sci. Eng.* **2019**, *174*, 682–695. [CrossRef]
10. Han, D.; Jung, J.; Kwon, S. Comparative study on supervised learning models for productivity forecasting of shale reservoirs based on a data-driven approach. *Appl. Sci.* **2020**, *10*, 1267. [CrossRef]
11. Han, D.; Kwon, S.; Son, H.; Lee, J. Production forecasting for shale gas well in transient flow using machine learning and decline curve analysis. In Proceedings of the Asia Pacific Unconventional Resources Technology Conference, Brisbane, Australia, 18–19 November 2019. [CrossRef]

12. Liao, L.; Zeng, Y.; Liang, Y.; Zhang, H. Data mining: A novel strategy for production forecast in tight hydrocarbon resource in Canada by random forest analysis. In Proceedings of the International Petroleum Technology Conference, Dhahran, Kingdom of Saudi Arabia, 13–15 January 2020. [CrossRef]
13. Bhattacharyya, S.; Vyas, A. Application of machine learning in predicting oil rate decline for Bakken shale oil wells. *Sci. Rep.* **2022**, *12*, 16154. [CrossRef]
14. Zhou, Q.; Dilmore, R.; Kleit, A.; Wang, J.Y. Evaluating gas production performances in marcellus using data mining technologies. In Proceedings of the Unconventional Resources Technology Conference, Denver, CO, USA, 25–27 August 2014; pp. 20–36. [CrossRef]
15. Grujic, O.; Da Silva, C.; Caers, J. Functional approach to data mining, forecasting, and uncertainty quantification in unconventional reservoirs. In Proceedings of the SPE Annual Technical Conference and Exhibition, Houston, Texas, USA, 28–30 September 2015; pp. 1704–1720. [CrossRef]
16. Lolon, E.; Hamidieh, K.; Weijers, L.; Mayerhofer, M.; Melcher, H.; Oduba, O. Evaluating the relationship between well parameters and production using multivariate statistical models: A middle Bakken and three forks case history. In Proceedings of the SPE Hydraulic Fracturing Technology Conference, The Woodlands, TX, USA, 9–11 February 2016. [CrossRef]
17. Khanal, A.; Khoshghadam, M.; Lee, W.J.; Nikolaou, M. New forecasting method for liquid rich shale gas condensate reservoirs with data driven approach using principal component analysis. *J. Nat. Gas Sci. Eng.* **2017**, *38*, 621–637. [CrossRef]
18. Xue, L.; Liu, Y.; Xiong, Y.; Liu, Y.; Cui, X.; Lei, G. A data-driven shale gas production forecasting method based on the multi-objective random forest regression. *J. Pet. Sci. Eng.* **2021**, *196*, 107801. [CrossRef]
19. Johan, D.C.; Shukla, P.; Shrivastava, K.; Koley, M. Data-Driven Completion Optimization for Unconventional Assets. In Proceedings of the Unconventional Resources Technology Conference, Denver, CO, USA, 13–15 June 2023; p. 3504. [CrossRef]
20. Ghadami, A.; Epureanu, B.I. Data-driven prediction in dynamical systems: Recent developments. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2022**, *380*, 1429–1442. [CrossRef] [PubMed]
21. Mohaghegh, S.D. *Shale Analytics: Data-Driven Analytics in Unconventional Resources*, 1st ed.; Springer: Cham, Switerland, 2017.
22. Ma, H.; Zhao, W.; Zhao, Y.; He, Y. A data-driven oil production prediction method based on the gradient boosting decision tree regression. *Comput. Model. Eng. Sci.* **2023**, *134*, 1773–1790. [CrossRef]
23. De Oliveira Werneck, R.; Prates, R.; Moura, R.; Goncalves, M.M.; Castro, M.; Soriano-Vargas, A.; Junior, P.R.; 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]
24. Gryzlov, A.; Safonov, S.; Arsalan, M. Intelligent Production Monitoring with Continuous Deep Learning Models. *SPE J.* **2022**, *27*, 1304–1320. [CrossRef]
25. Chi, L.; Su, H.; Zio, E.; Qadrdan, M.; Li, X.; Zhang, L.; Fan, L.; Zhou, J.; Yang, Z.; Zhang, J. Data-driven reliability assessment method of Integrated Energy Systems based on probabilistic deep learning and Gaussian mixture Model-Hidden Markov Model. *Renew. Energy* **2021**, *174*, 952–970. [CrossRef]
26. Chen, D.; Huang, C.; Wei, M. Shale Gas Production Prediction Based on PCA-PSO-LSTM Combination Model. *J. Circuits Syst. Comput.* **2024**, *33*, 2450176. [CrossRef]
27. He, Y.W.; He, Z.Y.; Tang, Y.; Xu, Y.J.; Long, J.C.; Sepehrnoori, K. Shale gas production evaluation framework based on data-driven models. *Pet. Sci.* **2023**, *20*, 1659–1675. [CrossRef]
28. Meng, J.; Zhou, Y.J.; Ye, T.R.; Xiao, Y.T.; Lu, Y.Q.; Zheng, A.W.; Liang, B. Hybrid data-driven framework for shale gas production performance analysis via game theory, machine learning, and optimization approaches. *Pet. Sci.* **2023**, *20*, 277–294. [CrossRef]
29. Wang, Y.; Wang, Z.; Qian, S.; Qiu, X.; Shen, W.; Zhang, X.; Lyu, B.; Cui, J. Data-driven framework for prediction and optimization of gas turbine blade film cooling. *Phys. Fluids* **2024**, *36*, 035160. [CrossRef]
30. Brilliant, L.S.; Dulkarnaev, M.R.; Danko, M.Y.; Lisheva, A.O.; Nabiev, D.K.; Khutornaya, A.I.; Malkov, I.N. Oil production management based on neural network optimization of well operation at the pilot project site of the Vatyeganskoe field (Territorial Production Enterprise Povkhneftegaz). *Georesursy* **2022**, *24*, 3–15. [CrossRef]
31. Bahaloo, S.; Mehrizadeh, M.; Najafi-Marghmaleki, A. Review of application of artificial intelligence techniques in petroleum operations. *Pet. Res.* **2023**, *8*, 167–182. [CrossRef]