

Article

A Production Prediction Method for Shale Gas Wells Based on Multiple Regression

Wente Niu ^{1,2,3}, Jialiang Lu ^{1,2,3} and Yuping Sun ^{3,*}

¹ School of Engineering Science, University of Chinese Academy of Sciences, Beijing 101400, China; niuwente19@mails.ucas.edu.cn (W.N.); jllu69@petrochina.com.cn (J.L.)

² Institute of Porous Flow and Fluid Mechanics, Chinese Academy of Sciences, Langfang 065000, China

³ Research Institute of Petroleum Exploration and Development, Beijing 100089, China

* Correspondence: sunyuping01@petrochina.com.cn

Abstract: The estimated ultimate recovery (EUR) of a single shale gas well is one of the important evaluation indicators for the scale and benefit development of shale gas, which is affected by many factors such as geological and engineering, so its accurate prediction is difficult. In order to realize the accurate prediction of ultimate recovery, this study considered 172 shale gas wells in the Weiyuan block as samples and selected 19 geological and engineering factors that affect the ultimate recovery of shale gas wells. Furthermore, eight key controlling factors were selected by means of the Pearson correlation coefficient and maximum mutual information coefficient comprehensive evaluation method. The data were divided into training and testing samples. Different numbers of training samples were selected and seven schemes were designed. Based on the key controlling factors, the ultimate recovery prediction model for shale gas wells in this block was established through multiple regression methods. The effectiveness of the prediction model was verified by analyzing the testing samples. The result shows that with the increase of the size of training samples, the error of the ultimate recovery predicted by the model gradually decreases gradually. When predicting the single gas well, the average absolute error of ultimate recovery is less than 20% if the number of the training gas well is more than 80. When analyzing the development potential of similar blocks without drilling, the error of the sum of ultimate recovery is less than 10% if the size of the training gas well reaches 60.

Keywords: shale gas well; multiple regression; Weiyuan block; key controlling factors; production prediction



Citation: Niu, W.; Lu, J.; Sun, Y. A Production Prediction Method for Shale Gas Wells Based on Multiple Regression. *Energies* **2021**, *14*, 1461. <https://doi.org/10.3390/en14051461>

Academic Editor: Reza Rezaee

Received: 3 January 2021

Accepted: 2 March 2021

Published: 7 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In China, there is abundant shale gas of different types that are distributed widely [1–4]. For shale gas with enormous resources, the estimated ultimate recovery (EUR) of a single well is important to accurately estimate the potential of shale gas and to achieve scale and efficient development [5]. The nano-scale pore characteristics, the multi-shift mechanism, and the working system of “non-fixed pressure and non-fixed production” of the shale gas reservoir lead to the complex flow characteristics of shale gas [6–9]. These complexities bring great uncertainty to the EUR evaluation of shale gas wells. Therefore, it is instructive for exploring shale gas efficiently to figure out the key controlling factors of shale gas wells and obtain the EUR prediction model.

The EUR of shale gas wells is affected by many factors, and the factors influencing the EUR of shale gas wells in different blocks are different. Therefore, an increasing number of studies have been performed on the EUR in shale gas wells. Lei et al. [10] and Jia et al. [11] analyzed the key factors affecting shale gas well production and further proposed the technical direction of improving single well EUR. Xiao [12] and Geng et al. [13] screened the main controlling factors of shale gas well production and established the prediction model by using the grey correlation analysis method. Ma et al. [14] used the Pearson correlation coefficient (Pearson) and maximum mutual information coefficient (MIC) analysis method

to analyze the key factors controlling the productivity of shale gas wells in the early stage. The key factors controlling the EUR of shale gas wells are determined, which plays an important role in increasing the output of shale gas wells and guiding the development of shale gas.

The multiple regression method is a type of mathematical statistics method, which is based on the fundamental regression principle. It can screen out the significant factor, which influences the variation of a dependent variable so that it can be applied to the forecasting production rate of oil and gas fields [15]. Many Chinese scholars have researched the application of the multiple regression method in oil and gas fields. Tang et al. [16] predicted the production capacity of oil fields by mean of the multiple regression method, tested, and made an accurate estimation to the prediction. They thought the prediction has great instructive meanings and suggested that the multiple regression method should be used for the oil and gas production capacity prediction. Hu et al. [17] built a productivity prediction model for the Changqing oil field by using the multiple regression method. The prediction accuracy for the daily oil production per well reaches 85%. Wu et al. [18] built an initial productivity prediction model for the main area of Jiaoshiba after fracturing by means of the multiple regression method and the prediction is certainly accurate. Li et al. [19] built the productivity analogy prediction model for the oil well areas of Huangjinba by using the multiple regression method.

So far, however, the application of multiple regression methods to shale gas wells has been less researched. To this end, this paper studied the application of the multiple regression method to the productivity prediction of shale gas wells by considering 172 shale gas wells in the Weiyuan block as samples from the perspective of multiple linear regression.

2. Methods

EUR is influenced by many factors, some of which are in a linear relationship with EUR and others are in a nonlinear relationship with EUR. In order to determine the key controlling factors of EUR comprehensively, the Pearson correlation coefficient and maximum mutual information coefficient (Pearson–MIC) method were adopted in this study to measure the linear and nonlinear relationships between EUR and various factors respectively.

2.1. Pearson Correlation Analysis

Pearson correlation coefficient, one of the widely used relation measurement standards, can measure the linear relation between two random variables [20,21]. The calculation formula is as follows:

$$\rho = \frac{\text{Cov}(x, y)}{\sqrt{\text{Var}(x)\text{Var}(y)}} \quad (1)$$

where $\text{Cov}(x, y)$ is the covariance of x and y , $\text{Var}(x)$ is the variance of x , and $\text{Var}(y)$ is the variance of y .

The greater the absolute value of the correlation coefficient is, the stronger the correlation is. The closer the correlation coefficient is to 1 or -1 , the stronger the correlation is. The closer the correlation coefficient is to 0, the weaker the correlation is. Generally, the correlation strength of variables is judged according to the following value ranges, i.e., 0.8–1.0: extremely strong correlation; 0.6–0.8: strong correlation; 0.4–0.6: moderate correlation; 0.2–0.4: weak correlation; and 0.0–0.2: extremely weak correlation or no correlation.

2.2. Maximum Mutual Information Coefficient

Maximum mutual information coefficient (MIC) is a non-parametric exploration based on information, which is used to measure the strength of linearity or nonlinearity between two variables [22,23]. It can show the linear functional relationship between variables and find the nonlinear functional relationship (exponential and periodic). Moreover, it can show the functional relationship and the nonfunctional relationship. In this way, it has broad application [23].

MIC fundamental takes advantage of the mutual information concept, which can be illustrated in the following formula [24]:

$$I(x; y) = \int p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} dx dy \quad (2)$$

The calculation formula of MIC is as follows:

$$\text{MIC}(x; y) = \max_{a \cdot b < B} \frac{I(x, y)}{\log_2 \min(a, b)} \quad (3)$$

where: $p(x)$ is the probability of variable x , $p(y)$ is the probability of variable y , $p(x, y)$ is the joint probability of variable x and variable y , and a and b are the number of grids divided in the x and y directions, which is essentially a grid distribution and B is a variable.

The calculation result of the MIC method is between 0 and 1.0 indicates full uncorrelation, and 1 indicates complete correlation. It is generally deemed that the two variables have a strong correlation when MIC is bigger than 0.5 [25].

2.3. Pearson–MIC Comprehensive Evaluation Method

Pearson correlation coefficient is sensitive to linear relationships [25–28]. Compared to the Pearson correlation coefficient, MIC is more robust than the Pearson correlation coefficient, less susceptible to outlier values, and can be used to detect potential nonlinear relationships between variables [14]. Combined with the advantages of these two correlation analyses, the Pearson–MIC comprehensive evaluation method proposed by Ma et al. [14] is used to screen the key controlling factors affecting the EUR of shale gas wells.

2.4. Multiple Linear Regression Method

Linear regression is one of the most important mathematical models, and it is often used as the base of many other models [29]. Multiple linear regression is a very important method for multivariate statistical analysis. This method can be used to evaluate the relative importance of each independent variable to the dependent variable [30]. Multiple linear regression model can be expressed as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n, \quad (4)$$

where y is the dependent variable, $x_1 \dots x_n$ is the independent variable, and $\beta_0 \dots \beta_n$ is the unknown parameter.

Due to the importance of the multiple linear regression model, this method has been widely used in various industries, such as the economy, petroleum, and meteorological industries.

3. Factor Selection

Weiyuan shale gas field is located in the northern part of southern Sichuan, with an area of about 4024 km², showing a northern mountain and hilly terrain in the central and southern regions. The terrain is tilted from northwest to southeast [31]. The buried depth of the high-quality shale section of the Wufeng-Longmaxi Formation of Lower Silurian in the Weiyuan block is 1500–3700 m deep, and the burial depth increases from southwest to southeast. The pressure coefficient is 1.2–2.0, indicating that they are mostly overpressure gas reservoirs. The target horizon of the Weiyuan shale gas field in the study area is the L1₁ sub-member of Longmaxi Formation (Longmaxi Formation is vertically divided into the first and second members, which are referred to hereinafter as L1 and L2, respectively. L1 member is further subdivided into the first (L1₁) and second (L1₂) sub-members. The L1₁ sub-member is further divided into L1₁¹, L1₁², L1₁³, and L1₁⁴ sublayers), with reservoir characteristics of high total organic carbon (TOC), high porosity, and high gas content [32].

The production of shale gas wells is affected by many factors, including geological factors, engineering factors, and economic factors. Geological factors include reservoir thickness, TOC, gas content, maturity, porosity and permeability characteristics of the

matrix, pressure coefficient, fracture development, burial depth, contents of brittle minerals, water saturation, etc. Engineering factors include drilling length of a high-quality reservoir, number of fracturing segments, number of perforation clusters, segment distance, cluster distance, horizontal sections, sand contents, fracturing fluid volume, the amount of proppant, flowback rate, etc. [11,14,33–40]. Geological factors are uncontrollable, whereas engineering factors and economic factors are controllable. Engineering factors are affected by geological factors and economic factors. Geological factors and engineering factors can directly affect the productivity of shale gas wells [14].

In this study, the geologic and engineering parameters are collocated from 172 shale gas wells in Weiyuan. As shown in Table 1, and 19 geological and engineering factors are selected to analyze the key controlling factors of EUR in the Weiyuan block and build the EUR prediction model according to the availability and the efficiency of the statistics.

Table 1. Geological and engineering factors.

Geological factors	vertical depth	TOC	porosity	thickness of L1 ₁ ¹ sublayer
	gas saturation	pressure coefficient	contents of brittle minerals	
Engineering factors	fracturing segment length	fracturing section		fracturing fluids intensity
	intensity of sand	average discharge		30-day flowback rate
	drilling catching length into the class I reservoir			90-day flowback rate
	drilling catching length of L1 ₁ ¹ sublayer			180-day flowback rate
	flowback rate of the gas generation peak			360-day flowback rate

4. Results

4.1. The Selection of Key Controlling Factors

Table 2 shows the scatter plot and linear fitting between EUR and 19 factors, and the scatter plot with the goodness-of-fit greater than 0.1 is shown in Figure 1.

Table 2. The linear fitting between estimated ultimate recovers (EUR) and various factors (R²).

EUR	vertical depth 3.805×10^{-4}	TOC 0.015	porosity 0.002	thickness of L1 ₁ ¹ sublayer 0.189
	gas saturation 0.141	pressure coefficient 9.985×10^{-5}	contents of brittle minerals 0.071	fracturing segment length 0.254
	fracturing section 0.136	fracturing fluids intensity 0.009	intensity of sand 0.003	average discharge 6.775×10^{-4}
	drilling catching length into the class I reservoir 0.224		drilling catching length of L1 ₁ ¹ sublayer 0.288	30-day flowback rate 0.006
	90-day flowback rate 9.620×10^{-5}	180-day flowback rate 4.136×10^{-5}	360-day flowback rate 0.001	flowback rate of the gas generation peak 0.008

Table 3 shows the Pearson correlation coefficient between EUR and 19 factors, and the influencing factors with the correlation coefficient bigger than 0.35 are selected. Geological factors include the thickness of the L1₁¹ sublayer and gas saturation. Engineering factors include fracturing segment length, fracturing section, drilling catching length into the class I reservoir, and drilling catching length of L1₁¹ sublayer.

Table 4 shows the MIC correlation coefficient between EUR and 19 factors, and the influencing factors with the correlation coefficient greater than 0.55 are selected. The geological factor is the thickness of the L1₁¹ sublayer, and engineering factors include fracturing segment length, fracturing fluids intensity, drilling catching length into the class I reservoir, drilling catching length of L1₁¹ sublayer, and 360-day flowback rate.

Combined with the Pearson–MIC comprehensive evaluation method, factors selected as the key controlling factors to build a multilinear regression model are as follows: geological factors are the thickness of L1₁¹ sublayer and gas saturation, and engineering factors are

fracturing segment length, fracturing section, fracturing fluids intensity, drilling catching length into the class I reservoir, drilling catching length of $L1_1^1$ sublayer, and 360-day flowback rate.

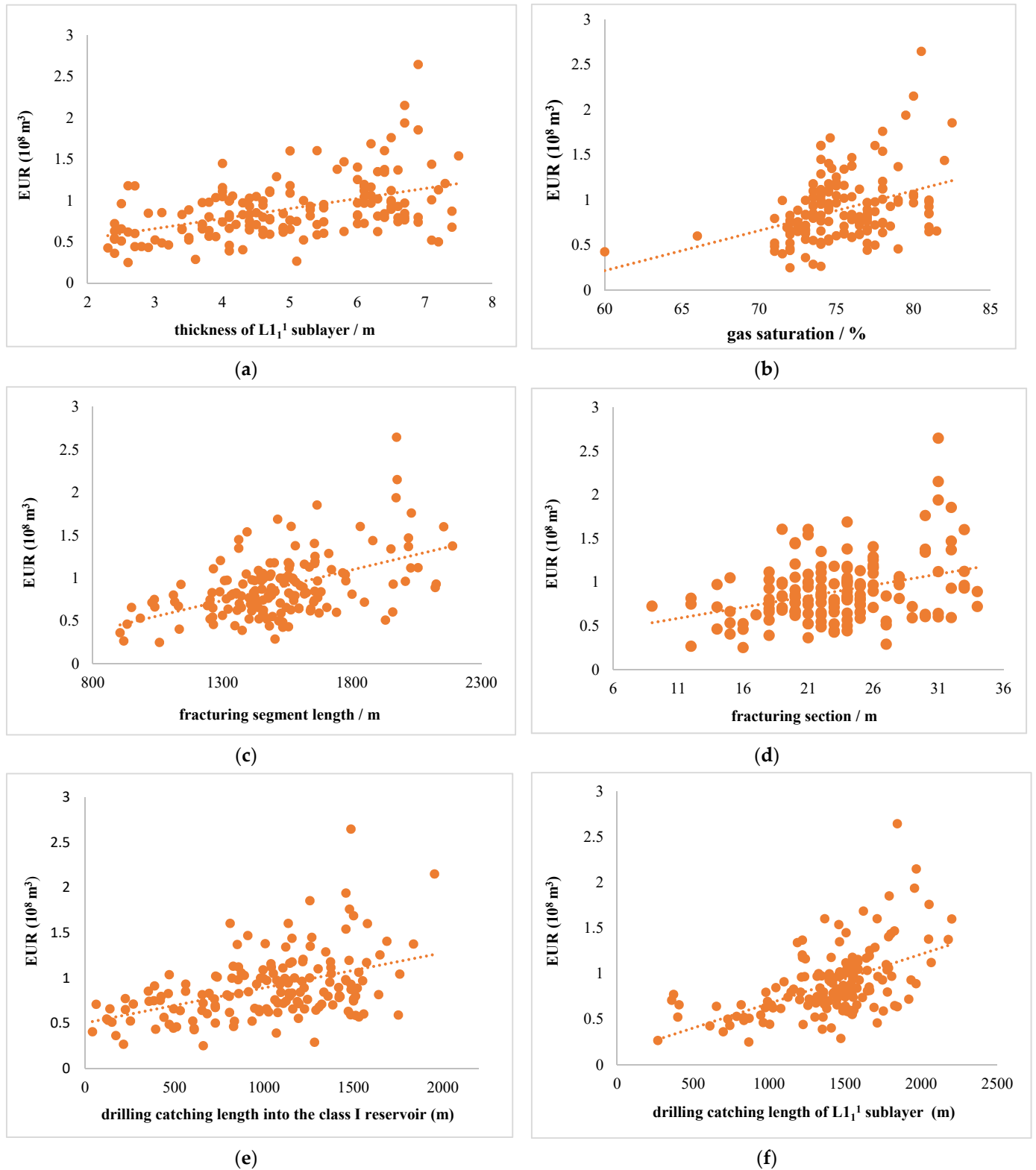


Figure 1. Scatter diagram of (a) EUR and thickness of $L1_1^1$ sublayer, (b) EUR and gas saturation, (c) EUR and fracturing segment length, (d) EUR and fracturing section, (e) EUR and drilling catching length into the class I reservoir, and (f) EUR and drilling catching length of $L1_1^1$ sublayer.

Table 3. Pearson correlation coefficient table.

	vertical depth −0.020	TOC −0.122	porosity 0.050	thickness of L1 ₁ ¹ sublayer 0.434
	gas saturation 0.375	pressure coefficient 0.010	contents of brittle minerals 0.266	fracturing segment length 0.504
EUR	fracturing section 0.368	fracturing fluids intensity −0.097	intensity of sand −0.051	average discharge −0.026
	drilling catching length into the class I reservoir 0.474		drilling catching length of L1 ₁ ¹ sublayer 0.536	30-day flowback rate 0.080
	90-day flowback rate 0.010	180-day flowback rate 0.015	360-day flowback rate −0.012	flowback rate of the gas generation peak −0.088

Table 4. Pearson correlation coefficient table.

	vertical depth 0.48	TOC 0.46	porosity 0.050	thickness of L1 ₁ ¹ sublayer 0.59
	gas saturation 0.45	pressure coefficient 0.48	contents of brittle minerals 0.55	fracturing segment length 0.6
EUR	fracturing section 0.5	fracturing fluids intensity 0.57	intensity of sand 0.43	average discharge 0.46
	drilling catching length into the class I reservoir 0.63		drilling catching length of L1 ₁ ¹ sublayer 0.71	30-day flowback rate 0.52
	90-day flowback rate 0.53	180-day flowback rate 0.52	360-day flowback rate 0.58	flowback rate of the gas generation peak 0.49

4.2. EUR Prediction of Shale Gas Wells Based on Multiple Regression

By considering 172 shale gas wells in the Weiyuan block as samples and EUR as the final evaluation target, the paper designed seven schemes to verify the effect of training sample size on the results of the regression model, and the details of the schemes are shown in Table 5. Among them, 40 testing samples are the same data of gas wells.

Table 5. Schemes.

Scheme	1	2	3	4	5	6	7
Training set	10	20	40	60	80	100	120
Test set	40	40	40	40	40	40	40
Sum	50	60	80	100	120	140	160

To carry out multi-factor analysis, a nondimensional treatment shall be performed to solve the problem that various indicators cannot be integrated because each indicator has its own nature and measurement unit, which are not comprehensive [41,42]. Therefore, the nondimensional treatment is used to deal with various factors by applying the extremum method to eliminate the influence of different dimensions (Formula 4). EUR prediction model can be built up based on the multiple regression of the processed data.

$$x'_i = \frac{\max x_i - x_i}{\max x_i - \min x_i} \quad (5)$$

where $\max x_i$ is the maximum value of the sample data and $\min x_i$ is the minimum value of the sample data.

From the Pearson–MIC comprehensive evaluation method, it can be derived that EUR has a nonlinear relationship with fracturing fluid intensity and 360-day flowback rate. After curve fitting, it is found that the secondary correlation between EUR and fracturing fluid intensity is better. The relationship is: $y = 0.109x - 0.002x^2 - 0.467$; the cubic correlation between EUR and 360-day flowback rate is better. The relationship is:

$y = -3.418x + 5.295x^2 - 2.391x^3 + 1.5$. Multiple linear regression was performed through a linear transformation of fracturing fluid intensity and 360-day flowback rate.

Considering that the relationship between various factors is relatively complicated and there may be multicollinearity, this paper adopts the stepwise regression method to solve this problem. The basic idea of this method is to gradually introduce new variables. If the partial regression square sum of the new variables is significant after testing, it indicates that it can be introduced. At this time, the new variables are considered as independent explanatory variables and cannot be represented by other explanatory variables (approximately) linearly. Otherwise, it means that the new variable is not independent and should not be introduced [43].

The multiple linear regression models and goodness-of-fit (R^2) of the seven schemes are shown in Table 6. For the test results of each scheme, the average relative error and the average absolute error of the true value, and the model predictions for each testing sample are calculated respectively. The evaluation results are shown in Figure 2 and Table 7.

Table 6. Multiple linear regression model results.

Scheme	Sample Size	Regression Model	R^2
1	50	$y = -6.329 * \text{gas saturation} - 0.548 * \text{fracturing segment length} + 3.363$	0.923
2	60	$y = -2.587 * \text{gas saturation} - 1.263 * \text{fracturing segment length} + 2.361$	0.873
3	80	$y = -0.881 * \text{fracturing segment length} - 0.538 * \text{thickness of } L1_1^1 \text{ sublayer} - 0.646 * \text{drilling catching length into the class I reservoir} + 1.899$	0.640
4	100	$y = -0.742 * \text{fracturing segment length} - 0.655 * \text{thickness of } L1_1^1 \text{ sublayer} - 0.600 * \text{drilling catching length into the class I reservoir} + 1.792$	0.647
5	120	$y = -0.771 * \text{fracturing segment length} - 0.608 * \text{thickness of } L1_1^1 \text{ sublayer} - 0.528 * \text{drilling catching length into the class I reservoir} + 1.711$	0.616
6	140	$y = -0.611 * \text{fracturing segment length} - 0.625 * \text{thickness of } L1_1^1 \text{ sublayer} - 0.469 * \text{drilling catching length into the class I reservoir} + 1.605$	0.542
7	160	$y = -0.638 * \text{fracturing segment length} - 0.465 * \text{thickness of } L1_1^1 \text{ sublayer} - 0.687 * \text{drilling catching length of } L1_1^1 \text{ sublayer} + 1.613$	0.491

* represents the product of two variables.

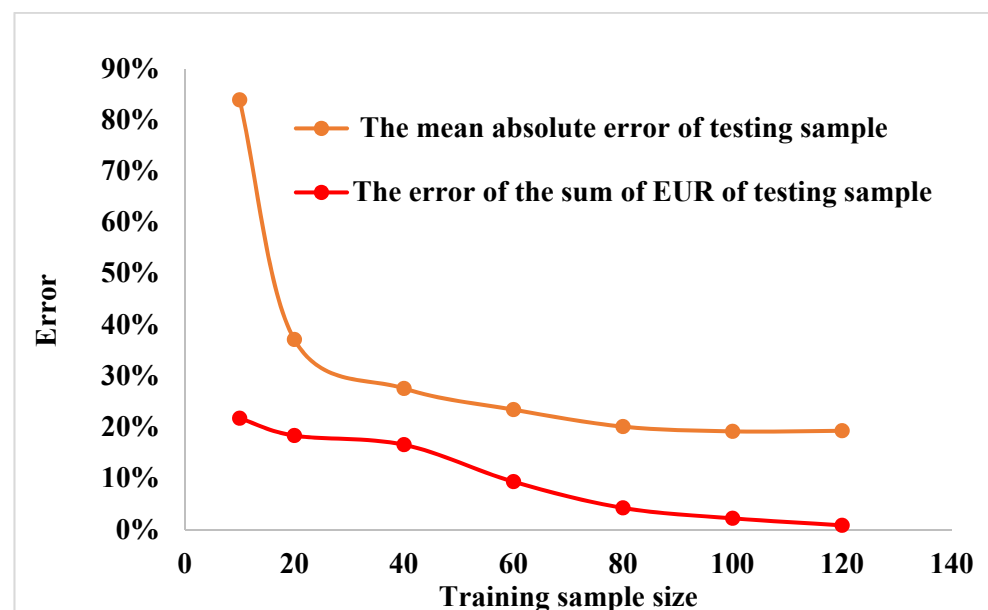


Figure 2. Model prediction results.

Table 7. The prediction results of different schemes.

	Scheme 1	Scheme 2	Scheme 3	Scheme 4	Scheme 5	Scheme 6	Scheme 7
Data sample size	50	60	80	100	120	140	160
Training sample	10	20	40	60	80	100	120
Testing sample	40	40	40	40	40	40	40
The mean absolute error of testing sample	83.87%	41.63%	27.55%	23.41%	20.10%	19.18%	19.29%
The error of the sum of EUR of testing sample	21.72%	18.35%	16.56%	9.34%	4.23%	2.20%	0.83%

5. Discussion

By observing the model obtained by means of multiple linear regression, it can be observed that when the training sample sizes were 10 wells and 20 wells, the relationship between EUR and each key controlling factor could not be explored fully due to the small amount of data; therefore, only two factors including gas saturation and fracturing segment length were included in the regression model. When the number of training samples increased to 40 gas wells, the regression model included three key controlling factors, namely, fracturing segment length, the thickness of the $L1_1^1$ sublayer, and drilling catching length into the class I reservoir. When the training sample size is increased to 120 gas wells, the significant impact of the drilling catching length of the $L1_1^1$ sublayer on EUR exceeds that of drilling catching length into the class I reservoir.

The prediction results of the testing samples show that as the number of training samples increases, the error of the prediction results can be significantly reduced, but the increase in the sample size makes the goodness-of-fit of the regression model worse. When the training sample size is less than 40 gas wells, the average absolute error and average relative error of the test samples are both large, and the ideal prediction effect cannot be achieved. As the training sample size increases to more than 80 gas wells, the average absolute error of the test samples can be reduced to less than 20%.

The error of the sum of EUR of 40 gas wells in the seven schemes of testing samples also gradually decreases with the increase of the training sample size. When the training sample size is 60 gas wells, the error is less than 10%, and when the sample size is 120 gas wells, the error is only 0.83%.

All in all, the EUR prediction model for shale gas wells in this area is established based on the multiple linear regression method, which verifies the effectiveness of the multiple regression method in shale gas well EUR prediction. This method can be operated simply and is suitable for on-site production prediction of shale gas well EUR. However, the application and promotion of the shale gas well EUR prediction model established by the above method in other blocks may still have certain restrictions. This paper only takes the Weiyuan block as an example to analyze the effectiveness of the multiple regression method, and specific blocks need specific analysis. The research results show that to reduce the error of the EUR prediction model, more production gas well data are needed. How to obtain a more accurate EUR prediction based on few data on gas wells is still the direction of continuous improvement.

6. Conclusions

(1) The geological and engineering parameters of 172 shale gas wells in the Weiyuan block are comprehensively analyzed. The key controlling factors affecting EUR in this block are determined by means of the Pearson–MIC comprehensive correlation evaluation method. The result shows that the main geological factors that affect the EUR of this block include the thickness of the $L1_1^1$ sublayer and gas saturation, and the engineering factors include the fracturing segment length, fracturing section, fracturing fluid intensity, drilling catching length into the class I reservoir, drilling catching length of $L1_1^1$ sublayer and 360-day flowback rate;

(2) The data of 172 actual production wells in the Weiyuan block are selected as samples, and seven different training sample schemes are designed. The multiple linear regression models are established based on the selected key controlling factors. The result

shows that the number of training samples for establishing the model has a great influence on the accuracy of prediction results. The more the training samples are, the smaller the error of predicted EUR is. When the training sample size is greater than 80 wells, the EUR prediction error is less than 20%, and the prediction applied to a single well EUR has good accuracy;

(3) By introducing the multiple linear regression method in the EUR prediction, the error of the sum of the EUR of the testing samples (40 gas wells) is less than 10% when the training sample size reaches 60 gas wells, and as the training sample size gradually increases, the sum of EUR of the testing samples will gradually decrease. This result shows that this method can be used as a good criterion for block standard well evaluation and applied to the development potential analysis of similar blocks without drilling.

In summary, this analysis method based on data mining provides a new idea for the EUR prediction of shale gas wells and improves the efficiency of EUR prediction for shale gas wells. Therefore, it is recommended to apply the multiple regression method to the EUR prediction of shale gas wells.

Author Contributions: Conceptualization, W.N.; methodology, Y.S.; validation, W.N., Y.S., and J.L.; formal analysis, W.N.; investigation, W.N.; resources, Y.S. and J.L.; data curation, W.N., Y.S., and J.L.; writing—original draft preparation, W.N.; writing—review and editing, W.N., Y.S., and J.L.; visualization, W.N.; supervision, J.L.; funding acquisition, J.L. and Y.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Science and Technology Major Project of the Ministry of Science and Technology of China Project (NO. 2016ZX05037006-006).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data is confidential, so the details are not disclosed.

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

References

1. Zou, C.N.; Pan, S.Q.; Jing, Z.H.; Gao, J.; Yang, Z.; Wu, S.; Zhao, Q. Shale oil and gas revolution and its impact. *Acta Pet. Sin.* **2020**, *41*, 1–12.
2. Zou, C.; Yang, Z.; Cui, J.; Zhu, R.; Hou, L.; Tao, S.; Yuan, X.; Wu, S.; Lin, S.; Wang, L.; et al. Formation mechanism, geological characteristics and development strategy of nonmarine shale oil in China. *Pet. Explor. Dev.* **2013**, *40*, 14–26. [[CrossRef](#)]
3. Li, J.-Z.; Dong, D.-Z.; Chen, G.-S.; Wang, S.-Q.; Cheng, K.-M. Prospects and strategic position of shale gas resources in China. *Nat. Gas Ind.* **2009**, *29*, 11–16.
4. Zou, C.; Zhao, Q.; Dong, D.; Yang, Z.; Qiu, Z.; Liang, F.; Wang, N.; Huang, Y.; Duan, A.; Zhang, Q.; et al. Geological characteristics, main challenges and future prospect of shale gas. *Nat. Gas Geosci.* **2017**, *28*, 1781–1796. [[CrossRef](#)]
5. Song, H.J.; Su, Y.H.; Xiong, X.L.; Liu, Y.; Zhong, S.-C.; Wang, J.-J. EUR evaluation workflow and influence factors for shale gas well. *Nat. Gas Geosci.* **2019**, *30*, 1531–1538.
6. Liu, X.H.; Yang, X.F.; Chen, Y.; Wu, J.; Feng, X. Seepage behavior and well testing in horizontal shale gas wells under multi-stage fracking. *Nat. Gas Ind.* **2013**, *33*, 77–81.
7. Pang, W.; Du, J.; Zhang, T. Production data analysis of shale gas wells with abrupt gas rate or pressure change. In Proceedings of the SPE Middle East Oil and Gas Show and Conference, Sanabis, Bahrain, 18–21 March 2019.
8. Wang, K.; Li, H.; Wang, J.; Jiang, B.; Bu, C.; Zhang, Q.; Luo, W. Predicting production and estimated ultimate recoveries for shale gas wells: A new methodology approach. *Appl. Energy.* **2017**, *206*, 1416–1431. [[CrossRef](#)]
9. Zhao, Y.L.; Liang, H.B.; Jing, C.; Shang, S.; Li, C. A new method for quick EUR evaluation of shale gas wells. *J. Southwest Pet. Univ.* **2019**, *41*, 124–131.
10. Lei, D.F.; Wang, L.; Zhang, X.W.; Yu, R.Z.; Guo, W. Research on key factors affecting shale gas well production. In Proceedings of the 2016 Natural Gas Academic Annual Conference, Yinchuan, China, 28 September 2016.
11. Jia, C.Y.; Jia, A.L.; He, D.B.; Wei, Y.; Qi, Y.; Wang, J. Key factors influencing shale gas horizontal well production. *Nat. Gas Ind.* **2017**, *37*, 80–88.
12. Xiao, H. Production evaluation method based on grey correlation analysis for shale gas horizontal wells in Weiyuan block. *Well Test.* **2018**, *27*, 73–78.

13. Geng, X.Y.; He, C.; Wan, Y.J. Productivity evaluation and prediction of shale gas horizontal wells based on grey correlation method. *J. Math. Pract. Theory* **2020**, *50*, 100–106.
14. Ma, W.L.; Li, Z.P.; Gao, C.; Sun, Y.P.; Zhang, J.P.; Deng, S.Z. Pearson-MIC analysis method for the initial production key controlling factors of shale gas wells. *China Sci. Pap.* **2018**, *13*, 1765–1771.
15. Wang, M.S.; Chen, X.G. Application of multiple linear regression method in oilfield production prediction. *Oil Gas Field Surf. Eng.* **2004**, *23*, 25–26.
16. Tang, S.A.; Zheng, Z.Z.; Zhu, X.B. Application of multiple regression analysis in forecasting oil and gas production capacity. *Sci. Technol. Innov. Herald.* **2008**, *21*, 169–171.
17. Hu, G.X.; Gong, F.H. Application of multiple regression analysis in productivity prediction of low permeability reservoirs. *Oil Gas Field Surf. Eng.* **2012**, *29*, 23–25.
18. Wu, L.Z.; Wang, S.; You, Y. Discussion on the method of production prediction after shale gas pressure. *J. Jiangnan Pet. Univ. Staff Work.* **2016**, *29*, 18–20.
19. Li, L.; Wang, J.J.; Liu, C.; Yuan, Y. Analysis and prediction model establishment of main controlling factors of shale gas production capacity. In Proceedings of the 2017 National Natural Gas Academic Annual Conference proceedings, Hangzhou, China, 19 October 2017.
20. Lena, P.D.; Margara, L. Optimal global alignment of signals by maximization of Pearson correlation. *Inf. Process. Lett.* **2010**, *110*, 679–686. [[CrossRef](#)]
21. Pearson, K. Mathematical contributions to the theory of evolution III. Regression, heredity, and panmixia. *Philos. Trans. R Soc. Lond. Ser. B Biol. Sci.* **1896**, 187–253. [[CrossRef](#)]
22. Reshef, D.N.; Reshef, Y.A.; Finucane, H.K.; Grossman, S.R.; McVean, G.; Turnbaugh, P.J.; Lander, E.S.; Mitzenmacher, M.; Sabeti, P.C. Detecting novel associations in large data sets. *Science* **2011**, *334*, 1518–1524. [[CrossRef](#)]
23. Tang, D.M.; Wang, M.W.; Zheng, W.F.; Wang, H. RapidMic: Rapid computation of the maximal information coefficient. *Evol. Bioinform.* **2014**, *10*, 11–16. [[CrossRef](#)]
24. Liang, T.; Zhang, Q.; Liu, X.; Lou, C.; Liu, X.; Wang, H. Time-frequency maximal information coefficient method and its application to functional corticomuscular coupling. *Trans. Neural Syst. Rehabil. Eng.* **2020**. [[CrossRef](#)]
25. Zhang, Y.H.; Li, Y.J.; Zhang, T. Maximum information entropy method for detecting multiple correlation. *J. Electron. Inf. Technol.* **2015**, *37*, 123–129.
26. Cui, N. Study on Climate-Malaria Sensitive Population Identification and Vulnerability Regionalization Based on MIC Method. Master's Thesis, Xi'an University of Science and Technology, Xi'an, China, 2013.
27. Cheng, X. Research on Farmer Household Credit Evaluation based on Support vector Machine. Ph.D. Thesis, Shanxi University of Finance and Economics, Taiyuan, China, 2015.
28. Liu, H.M.; Yang, D. Application of maximum information coefficient in differential expression miRNA recognition in valvular heart disease. *China Sci. Pap.* **2017**, *12*, 707–711.
29. Zhou, Z.H. *Machine Learning*; Tsinghua University Press: Beijing, China, 2016.
30. Leng, J.F.; Gao, X.; Zhu, J.P. Application of multivariate linear regression statistical prediction model. *Stat. Decis.* **2016**, 82–85. [[CrossRef](#)]
31. Wu, X.Z. Key Technologies in the Efficient Development of the Wei-yuan Shale Gas Reservoir, Sichuan Basin. *Pet. Drill. Technol.* **2019**, *47*, 1–9.
32. Chen, X.; Xu, J.L.; Li, Q.; Xiao, J.; Zhong, S. An analysis of main controlling factors of production for horizontal shale gas well in Weiyuan Block. *J. Southwest Pet. Univ.* **2020**, *42*, 63–74.
33. Blasingame, T. Reservoir engineering aspects of unconventional reservoirs. In Proceedings of the 11th Annual Ryder Scott Reserves Conference, Houston, TX, USA, 2 September 2015.
34. Zheng, A.W.; Liang, B.; Shu, Z.G.; Zhang, B.Q.; Li, J.Q.; Lu, Y.Q.; Li, L.; Shu, Z.H. Analysis of influencing factors of shale gas productivity based on large data technology: A case of Jiaoshiha block in Fuling Gas Field, Sichuan Basin. *Nat. Gas Geosci.* **2020**, *31*, 542–551.
35. Fang, D.Z.; Zeng, H.; Wang, N.; Zhang, Y. Study on high production factors of high-pressure shale gas from Haynesville Shale Gas development data. *Oil Drill. Prod. Technol.* **2015**, *37*, 58–62.
36. Nie, H.K.; Tang, X.; Bian, R.K. Controlling factors for shale gas accumulation and prediction of potential development area in shale gas reservoir of South China. *Acta Pet. Sin.* **2009**, *30*, 484–491.
37. Chong, K.K.; Grieser, B.; Jaripatke, O.; Passman, A. A completions roadmap to shale-play development: A review of successful approaches toward shale-play stimulation in the last two decades. In Proceedings of the International Oil and Gas Conference and Exhibition in China, Beijing, China, 8–10 June 2010.
38. Li, A.; Shi, W.R.; Yuan, Z.H.; Zhang, Y.; Zhao, H.; Ren, Y.; Shi, Y. Main Factors Controlling Marine Shale Gas Accumulation in Jiaoshiha Shale Gas Field of Fuling Area. *Unconv. Oil Gas.* **2016**, *3*, 27–34.
39. Mallick, M.A.; Chalpurkar, M.P. Factors controlling shale gas production: Geological perspective. In Proceedings of the Abu Dhabi International Petroleum Exhibition and Conference, Abu Dhabi, United Arab Emirates, 10–13 November 2014.
40. Liang, B. New way of predicting initial productivity of shale gas based on orthogonal test analysis theory in fuling jiaoshiha work-zone. *J. Jiangnan Pet. Univ. Staff Work.* **2017**, *30*, 8–12.
41. Han, S.J. Comparison of data dimensionless methods in SPSS cluster analysis. *Sci. Mosain.* **2008**, *03*, 229–231.

-
42. Ma, L.P. Standardization of statistical data—Dimensionless method—Study and application of modern statistical analysis Methods (III). *Beijing Stat.* **2000**, *3*, 34–35.
 43. You, S.B.; Yan, Y. Stepwise regression analysis and its application. *Stat. Decis.* **2017**, *14*, 31–35.