

Article

Investigation of Flowback Behaviours in Hydraulically Fractured Shale Gas Well Based on Physical Driven Method

Wei Guo , Xiaowei Zhang *, Lixia Kang, Jinliang Gao and Yuyang Liu

PetroChina Research Institute of Petroleum Exploration & Development, Beijing 100083, China; guowei69@petrochina.com.cn (W.G.); kanglixia@petrochina.com.cn (L.K.); gaojinliang@petrochina.com.cn (J.G.); yuyangliu@petrochina.com.cn (Y.L.)

* Correspondence: zhangxw69@petrochina.com.cn; Tel.: +86-010-83599457

Abstract: Due to the complex microscope pore structure of shale, large-scale hydraulic fracturing is required to achieve effective development, resulting in a very complicated fracturing fluid flowback characteristics. The flowback volume is time-dependent, whereas other relevant parameters, such as the permeability, porosity, and fracture half-length, are static. Thus, it is very difficult to build an end-to-end model to predict the time-dependent flowback curves using static parameters from a machine learning perspective. In order to simplify the time-dependent flowback curve into simple parameters and serve as the target parameter of big data analysis and flowback influencing factor analysis, this paper abstracted the flowback curve into two characteristic parameters, the daily flowback volume coefficient and the flowback decreasing coefficient, based on the analytical solution of the seepage equation of multistage fractured horizontal Wells. Taking the dynamic flowback data of 214 shale gas horizontal wells in Weiyuan shale gas block as a study case, the characteristic parameters of the flowback curves were obtained by exponential curve fittings. The analysis results showed that there is a positive correlation between the characteristic parameters which present the characteristics of right-skewed distribution. The calculation formula of the characteristic flowback coefficient representing the flowback potential was established. The correlations between characteristic flowback coefficient and geological and engineering parameters of 214 horizontal wells were studied by spearman correlation coefficient analysis method. The results showed that the characteristic flowback coefficient has a negative correlation with the thickness $\times$ drilling length of the high-quality reservoir, the fracturing stage interval, the number of fracturing stages, and the brittle minerals content. Through the method established in this paper, the shale gas flowback curve containing complex flow mechanism can be abstracted into simple characteristic parameters and characteristic coefficients, and the relationship between static data and dynamic data is established, which can help to establish a machine learning method for predicting the flowback curve of shale gas horizontal wells.



Citation: Guo, W.; Zhang, X.; Kang, L.; Gao, J.; Liu, Y. Investigation of Flowback Behaviours in Hydraulically Fractured Shale Gas Well Based on Physical Driven Method. *Energies* **2022**, *15*, 325. <https://doi.org/10.3390/en15010325>

Academic Editor: Maxim Tyulenev

Received: 10 November 2021

Accepted: 23 December 2021

Published: 4 January 2022

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



Copyright: © 2022 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/>).

Keywords: shale gas; flowback; big-data analysis; horizontal well; fracturing fluids

1. Introduction

Multistage fracturing of horizontal wells is widely used in the exploration and development of shale gas. It typically takes tens of thousands of square meters of fracturing fluids and thousands of tons of sand to fracture a shale gas horizontal well, and the flowback of the fracturing fluids directly affects shale gas production. Hence, the flowback curve of fracturing fluids is a key issue in shale gas exploration [1]. The pores and fractures in shale gas reservoirs have different scales [2], and new multiscale pore-split systems are created by multistage fracturing. Thus, the shale gas flow mechanism is very complex, including pressure sensitivity, adsorption, diffusion, slip, imbibition, and seepage [3–6]. Currently, no consensus has been reached on this flow mechanism. Moreover, there is gas–water two-phase flow in the formation after large-scale fracturing. At present, there is no commercial software to realize the numerical simulation of flowback curve. It may take

several years to develop such software because, when developing such calculation programs, the solutions of the flow equations, unstructured meshing, equation discretization, and large-scale irregular sparse matrix are very complex. The gas flow equation for shale gas involves many flow mechanisms. The solution of this equations requires many known parameters which are difficult to obtain directly or measure experimentally. For example, shale gas is a multicomponent mixture containing hydrocarbons and nonhydrocarbon gases (CO_2 and N_2) [7], and its molecular free path at a high temperature and pressure cannot be obtained for unknown component proportions. Even if there is a perfect numerical simulation program, inaccurate parameters will affect the reliability of the results.

In recent years, with the development of big-data-related technology, researchers began to introduce machine learning into the research of oil and gas development and have mainly used it as a prediction tool. Kohli et al. [8] took well-log data as input parameters and trained the multilayer forward neural network by using the least square Levenberg–Marquardt optimization algorithm to predict the formation permeability; the predicted permeability is consistent with field data. Jia et al. [9] studied the water channel problem resulting from long-term water injection. A density peak clustering algorithm based on streamline clustering was used to quantify the flow area for water flooding in the oil reservoir and thereby effectively identify the invalid water injection circulation channels between the injection and production wells, as well as areas with development potential. Adibifard et al. [10] carried out a Chebyshev polynomial interpolation of pressure derivative data and input the results to an artificial neural network to estimate reservoir parameters. Ghaffarian et al. [11] processed the pseudo pressure of gas wells and used the pseudo pressure derivative data as the input of single and coupled multilayer perceptron network to identify the condensate gas reservoir model. Tian et al. [12] used recurrent neural network learning to train the data collected by permanent downhole pressure gauge (PDG) for the inversion of reservoir permeability and other parameters and production prediction. Hung et al. [13] introduced the application of Gaussian process regression (GPR), support vector machine (SVM), and random forest (RF) to predict CO_2 trapping efficiency in saline formations.

As the fracturing fluids flowback is influenced by many factors, big data technology has attracted increasing attention. Bai et al. [14] used exponential and harmonic functions to fit the water production in the flowback stage and produced the water stages of 32 wells and developed a prediction tool for accumulated water production based on the fitting results. Zhou et al. [15] used various binary and multivariate methods to analyse the data from 187 wells in the Marcellus block in the USA. The flowback rate during the first three weeks of data collection was found to increase with the thermal maturity and decrease with increasing strata thickness. Lin et al. [16] used a back propagation neural network to model the static data and flowback rate for 74 wells in Sichuan Province, China. The flowback rate and gas production in the first month were estimated using six parameters with relatively high weights in the model. Liu et al. [17] developed two neural networks with different structures to predict the flowback rate at specific time points and compared the predictions. In summary, current big data-based methods used to determine the flowback pattern focused on static or dynamic data of many wells for a given block, and a combined analysis of static and dynamic data has not been conducted. Some studies have been performed by using a fixed data point or a periodic average of dynamic data, which is a static data methodology in essence. All these lead to incomplete data utilization and limited application scope of the method.

In order to simplify the time-dependent flowback curve into simple parameters, as the target parameters of big data analysis and flowback influencing factor analysis, this paper combined the flowback dynamic data and static data of shale gas fracturing horizontal wells to study the flowback curve. The approaches and processes used in this study are shown in Figure 1. The first part of this paper briefly introduces the geological background of the research area and the relevant data collected. In the second part, the multi-stage fracturing physical model of shale gas horizontal wells was established. The flow equation in Laplace

space was obtained from the convolution formula, and the time-dependent flowback rate was obtained. In the third part, the flowback data of 214 horizontal wells in the Weiyuan block were fitted to obtain the characteristic parameters. The calculation formula of characteristic coefficient including characteristic parameters was established. Then, the correlation analysis was carried out by using a flowback characteristic coefficient, and the influencing factors of the flowback characteristic coefficient were comprehensively studied. Through the method established in this paper, the shale gas flowback curve containing complex flow mechanism can be abstracted into simple characteristic parameters and characteristic coefficients, and the relationship between static data and dynamic data is established, which can help to establish a machine learning method for predicting the flowback curve of shale gas horizontal wells.

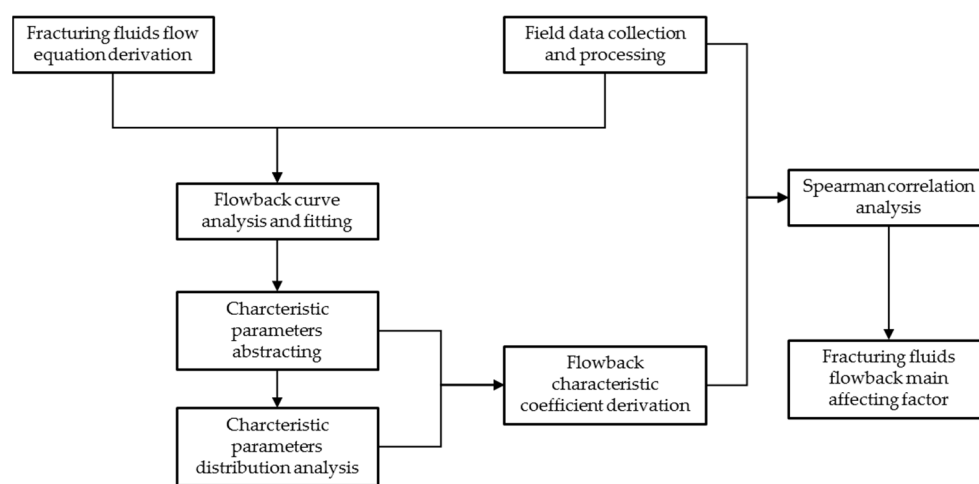


Figure 1. Schematic diagram of the approach used in the study 2. Research area background and analysis data.

Weiyuan shale gas block is located in the southwest of Sichuan Basin of China. The whole Weiyuan block has a monoclinic structure inclined to the southeast. The Longmaxi Formation (LF) is the main exploration target layer. The LF can be divided into two layers: L1 and L2. The first layer of L1 can be subdivided into two sublayers, namely, L1₁ and L1₂. The L1₁ sublayer can be further subdivided into four layers, namely, L1₁¹, L1₁², L1₁³ and L1₁⁴, from bottom to top. The middle and lower parts of L1₁¹ have the best reservoir quality.

Table 1 shows the data from the study area. A total of 20 different types of data were collected from 282 wells in Weiyuan shale gas block, including the fracturing lateral length, the number of fracturing stages, first-year average daily production rate, vertical depth, TOC, porosity, high-quality reservoir thickness, gas saturation, pressure coefficient, brittle mineral content, average fracturing stage interval, fracturing fluid intensity, proppant intensity, average hydraulic fracturing fluid displacement, drilling length in high-quality reservoir, EUR (Estimated Ultimate Recovery), 30 day flowback rate, 90 day flowback rate, 180 day flowback rate, 360 day flowback rate, and peak gas production flowback rate. The data volume ranges from 214 to 282. The minimum data volume is 214, and the corresponding data of these 214 wells is used in subsequent analysis. The 30 day flowback rate, 90 day flowback rate, 180 day flowback rate, 360 day flowback rate, peak gas production flowback rate, and first-year average daily production rate are taken from the daily production reports of the horizontal wells. The fracturing lateral length, number of fracturing stages, average fracturing stage interval, fracturing fluid intensity, proppant intensity, and average hydraulic fracturing fluid displacement are taken from the drilling and completion reports of horizontal wells. The vertical depth, TOC, porosity, high-quality reservoir thickness, gas saturation, pressure coefficient, brittle mineral content and drilling length in high-quality reservoirs are taken from well logging interpretation.

Table 1. Statistical tables of data for the study area. Copyright permission: The copyright of Table 1 belongs to the publisher and authors.

Data set	Data Volume	Minimum Value	Maximum Value	Average Value	Standard Deviation
30 day flowback rate	260	5%	77%	24%	0.12
90 day flowback rate	240	8%	103%	35%	0.16
180 day flowback rate	231	10%	116%	42%	0.18
360 day flowback rate	214	13%	129%	48%	0.20
Peak gas production flowback rate	243	3%	83%	23%	0.15
Fracturing lateral length (m)	281	502.00	2577.00	1515.87	308.47
Number of fracturing stages	281	3.00	36.00	22.63	5.56
First-year average daily production rate (10 ⁴ m ³ /d)	281	0.98	34.02	9.47	5.28
Vertical depth (m)	279	2200.00	3800.00	3008.24	352.49
TOC content (%)	282	3.30	7.80	5.39	0.76
Porosity (%)	282	5.00	8.90	7.11	0.96
High-quality reservoir thickness (m)	282	2.30	7.50	5.17	1.36
Gas saturation (%)	279	60.00	83.00	75.70	3.14
Pressure coefficient	282	1.35	2.05	1.74	0.19
Brittle mineral content (%)	282	62.50	96.00	78.67	7.79
Average fracturing stage interval (m)	281	43.10	479.33	69.94	27.81
Fracturing fluid Intensity (m ³ /m)	281	7.12	47.94	27.25	4.51
Proppant intensity (t/m)	281	0.29	3.00	1.63	0.38
Average hydraulic fracturing fluid displacement (m ³ /min)	281	6.19	15.00	11.80	1.33
Drilling length in high-quality reservoir (m)	206	40.60	2380.50	1011.98	452.00

2. Basic Theory

The time-dependent daily flowback volume curve for the Weiyuan shale gas block in southern Sichuan Province, China, shows that the daily flowback volume is large during the early stages of continuous production and decreases to a stable value during the late stage of production. The daily flowback volume during early-stage production is two to three orders of magnitude higher than that during the later stage, and the relationship between daily flowback and time is generally exponential. The flowback of a fracturing fluid after large-scale fracturing is essentially the seepage of fluid from the stimulated reservoir volume (SRV) area to the wellbore. The corresponding law can also be obtained from the seepage equation of fracturing fluid.

2.1. Bottomhole Flow Equation

The formula derivation in this paper considers single-phase liquid seepage, assumes constant flowing bottom hole pressure (FBHP) and continuous production, and is based on the multi-stage fracturing seepage equation of horizontal well [18]. The following dimensionless parameters are defined:

Dimensionless pressure:

$$p_D = \frac{2\pi kh[p_i - p(r, t)]}{qB\mu} \quad (1)$$

Dimensionless time:

$$t_D = \frac{kt}{\phi\mu C_t L^2} \quad (2)$$

Dimensionless distance:

$$x_D = \frac{x}{L} \quad (3)$$

$$y_D = \frac{y}{L} \quad (4)$$

where K denotes the permeability, m^2 ; ϕ denotes the porosity; μ denotes the viscosity, $\text{Pa}\cdot\text{s}$; C_t denotes the comprehensive compressibility coefficient, $1/\text{Pa}$; x_{fi} denotes the half-length of the i th fracture, m ; and $L = \sum_{i=1}^n x_{fi}$ denotes the sum over the half-lengths of n fractures, m .

For multistage fractured horizontal wells, the total flowback volume is the sum of the flowback volumes of all the individual fractures. The flowback volume can be related to the FBHP in Laplace space as follows:

$$\sum_{j=1}^n \bar{q}_{Dj} = \frac{1}{s} \quad (5)$$

$$\bar{P}_{WDi}(f(u)) = \sum_{j=1}^n s \bar{q}_{Dj} \bar{P}_{Dij}(f(u)) \quad (6)$$

where:

$$\bar{P}_{Dij}(f(u)) = \pi \int_0^\infty P_D(t_D) e^{-f(u)t_D} dt_D \quad (7)$$

$$P_D(t_D) = \int_0^{t_D} G_{xD}(\tau_D) G_{yD}(\tau_D) d\tau_D \quad (8)$$

The expressions for G_{xD} and G_{yD} vary with the external boundary conditions. For enclosed strata with rectangular boundaries, G_{xD} and G_{yD} are defined as follows [19]:

$$G_{xD}(\tau_D) = \frac{1}{x_{eD}} \left[1 + 2 \sum_{n=1}^{\infty} \exp\left(\frac{-n^2 \pi^2 (\tau_D)}{2x_{eD}^2}\right) \cos \frac{n\pi x_{wD}}{x_{eD}} \cos \frac{n\pi x_D}{x_{eD}} \right] \quad (9)$$

$$G_{yD}(\tau_D) = \frac{2}{y_{eD}} \left[1 + \frac{2y_{eD}}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} \exp\left(\frac{-n^2 \pi^2 (\tau_D)}{2y_{eD}^2}\right) \sin \frac{n\pi}{y_{eD}} \cos \frac{n\pi y_{wD}}{y_{eD}} \cos \frac{n\pi y_D}{y_{eD}} \right] \quad (10)$$

Equations (5) and (6) can be used to obtain the following linear equation:

$$\begin{bmatrix} f(u)\bar{P}_{D1,1} & f(u)\bar{P}_{D1,2} & f(s)\bar{P}_{D1,3} & \cdots & f(u)\bar{P}_{D1,n} \\ f(u)\bar{P}_{D2,1} & f(u)\bar{P}_{D2,2} & f(s)\bar{P}_{D2,3} & \cdots & f(u)\bar{P}_{D2,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f(u)\bar{P}_{Dk,1} & f(u)\bar{P}_{Dk,2} & f(u)\bar{P}_{Dk,3} & \cdots & f(u)\bar{P}_{Dk,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f(u)\bar{P}_{Dn,1} & f(u)\bar{P}_{Dn,2} & f(u)\bar{P}_{Dn,3} & \cdots & f(u)\bar{P}_{Dn,n} \\ u & u & u & \cdots & u \end{bmatrix} \begin{bmatrix} \bar{q}_{1D} \\ \bar{q}_{2D} \\ \vdots \\ \bar{q}_{kD} \\ \vdots \\ \bar{q}_{nD} \\ \bar{P}_{WD} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad (11)$$

The solution of Equation (11) without considering the wellbore storage and the skin factor yields the FBHP as the dimensionless pressure $\bar{P}_{WD}$. The dimensionless pressure $\bar{P}_{CWD}$ is obtained considering the skin factor and wellbore storage for multistage fracturing of horizontal wells in low permeability reservoirs as:

$$\bar{P}_{CWD} = \frac{f(u)\bar{P}_{WD}(f(u)) + S}{u\{1 + C_D f(u) S \bar{P}_{WD}(f(u)) + S\}} \quad (12)$$

where C is the wellbore storage constant, Pa/m^3 ; $C_D = \frac{C}{2\pi\phi C_t h L^2}$ denotes the dimensionless wellbore storage constant; S is the total skin factor of the fracture system; u is the Laplace

transform variable; and $f(u)$ is a function that characterizes the properties of the strata. For a homogeneous stratum, $f(u) = u$.

For production wells with a variable flowback volume, the dimensionless pressure can be expressed in the following convolution form:

$$P_D(x_D, y_D, z_D, t_D) = \int_0^{t_D} Q_D(t_D - \tau) \frac{\partial P_u(x_D, y_D, z_D, \tau)}{\partial \tau} d\tau \quad (13)$$

where $P_u(x_D, y_D, z_D, t)$ denotes the dimensionless pressure of multistage fractured horizontal wells per unit flowback volume; $Q_D(t_D) = \frac{q(t)}{q_I}$ denotes the dimensionless flowback volume; and q_I represents the unit flowback volume.

Using the properties of Laplace transform for convolutions yields the relationship between the pressure and flowback volume in Laplace space:

$$\bar{P}_D(x_D, y_D, z_D, u) = \bar{Q}_D(u) u \bar{P}_u(x_D, y_D, z_D, u) \quad (14)$$

Equation (14) can be used to obtain the flowback volume in Laplace space:

$$\bar{Q}_D(u) = \frac{\bar{P}_D(x_D, y_D, z_D, u)}{u \bar{P}_u(x_D, y_D, z_D, u)} \quad (15)$$

Equation (15) can also be expressed in terms of the FBHP as follows:

$$\bar{Q}_D(u) = \frac{\bar{P}_{wfD}(u)}{u \bar{P}_{wfU}(u)} \quad (16)$$

where $\bar{P}_{wfU}(u)$ is the $\bar{P}_{WD}$ obtained by solving Equation (11) for a unit flowback using G_{xD} and G_{yD} , which are both exponential functions of time. Therefore, the flowback can be simplified to $Q_I(t) = \alpha \exp(-\beta t)$. Thus, the time-dependent flowback curve is defined by two characteristic parameters, α and β . Figure 2 shows the flowback volume versus time for horizontal segments of a 1000-m horizontal well with 20 stages, an effective fracture half-length of 40 m, and an effective fracture permeability of 0.25 mD. Fitting the curve yields the following equation for the flowback volume:

$$q(t) = 109.23 \exp(-0.007t) \quad (17)$$

where t denotes the flowback time, (d); and $q(t)$ is the daily flowback volume, (m^3/d).

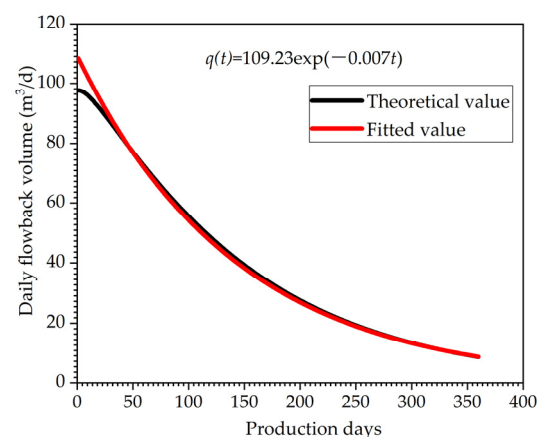


Figure 2. Theoretical flowback curve for multistage fracturing of horizontal wells.

Equation (9) accurately describes the characteristics of the flowback curve of horizontal wells, which indicates that α and β can be effectively used to parametrize the dynamic flowback curves. In addition, the flowback pattern is consistent with field measurements

(Figure 3). The field data in Figure 3 are from one of the 282 Wells in Table 1. The abscissa is production days, the ordinate is daily fracturing fluid flowback volume, and the blue dot represents the daily flowback volume of a specific day.

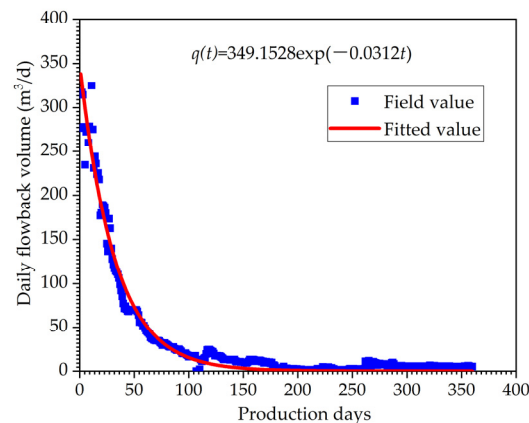


Figure 3. Field flowback curve for multistage fracturing of horizontal wells.

2.2. Fitting

The exponential expression for the flowback volume can be used to obtain the daily flowback rate R_l for a single well as follows:

$$R_l(t) = \frac{q}{Q_{total}} = \frac{\alpha}{Q_{total}} \exp(-\beta t) = \gamma \exp(-\beta t) \quad (18)$$

where q denotes the daily flowback volume, m^3/d ; Q_{total} denotes the total quantity of injected fracturing fluid, m^3 ; α is a daily flowback volume coefficient, m^3/d ; β is a decreasing flowback coefficient, $1/\text{d}$, which is a fitting parameter; $\gamma = \alpha/Q_{total}$ is the daily flowback rate coefficient, $1/\text{d}$, which is also a fitting parameter; and t denotes the production time (d).

The curves for the daily flowback rate and daily flowback volume have the same shape but different magnitudes. For a total daily flowback volume that is recorded only once a day, the range of t in Equation (18) is $\{t \geq 1, t \in \mathbb{Z}\}$. The cumulative flowback rate $\hat{S}_N$ for the previous N days can be then calculated as:

$$\hat{S}_N = \sum_{t=1}^N \gamma \exp(-\beta t) \frac{\gamma}{e^\beta - 1} \left(1 - \frac{1}{e^{N\beta}}\right) \quad (19)$$

The cumulative flowback rates on the 30th, 90th, 180th and 360th day of production for a shale gas well are used as the characteristic flowback rates, and the estimated values of α and β are obtained by solving Equation (20):

$$\begin{cases} S_{30} = \hat{S}_{30} \\ S_{90} = \hat{S}_{90} \\ S_{180} = \hat{S}_{180} \\ S_{360} = \hat{S}_{360} \end{cases} \quad (20)$$

The values on the left and right sides of Equation (20) correspond to measured data and the calculation results from Equation (19), respectively. As there are fewer unknowns than equations in Equation (20), exact solutions for γ and β cannot be obtained. Therefore, γ and β are estimated by minimizing the sum of squares of the differences between the two sides of Equation (20):

$$(\hat{\gamma}, \hat{\beta}) = \underset{(\gamma, \beta)}{\operatorname{argmin}} \sum_N (\hat{S}_N - S_N)^2 \quad (21)$$

where the values of N are 30, 90, 180, and 360.

2.3. Big Data Analysis

A large quantity of data is generated during the exploration and development of shale gas, i.e., from drilling to fracturing and from well closing in to flowback. Various factors affect the flowback, which cannot be solved using currently available mathematical equations. Big data can be used to solve this problem. In this study, the Spearman correlation coefficient and distribution estimation were used to perform a big data analysis.

2.3.1. Correlation Analysis

The Spearman correlation coefficient has two advantages. First, sortable variables are assumed, and a normal sample distribution of the samples is not required, which is appropriate for the dataset used in this study. Second, the Spearman correlation corresponds to a monotonous, rather than linear, correlation between two random variables X and Y and is therefore likely to reveal a nonlinear relationship between the variables. The Spearman correlation coefficient is defined as follows [18]:

$$\rho_{XY} = 1 - \frac{6 \sum_{i=1}^N [\text{rank}(x_i) - \text{rank}(y_i)]^2}{N(N^2 - 1)} \quad (22)$$

where N denotes the number of samples, and rank denotes the rank number of an observed value of X or Y .

The range of ρ_{XY} is $[-1, 1]$. A positive or negative ρ_{XY} indicates a positive or negative correlation between X and Y , respectively. The higher the absolute value of ρ_{XY} is, the stronger the correlation is. The significance of the Spearman correlation coefficient can be tested by the following hypotheses [20]:

$$H_0 : \rho_{XY} = 0 \quad (23)$$

$$H_1 : \rho_{XY} \neq 0 \quad (24)$$

If H_0 is satisfied, then:

$$\rho_{XY} \sim \text{Normal} \left(0, \frac{1}{n-1} \right) \quad (25)$$

Generally, if $p \leq 0.05$, H_0 is rejected, and ρ_{XY} is considered significant, whereas $p > 0.05$ indicates there is insufficient evidence to reject H_0 , and ρ_{XY} may be false.

2.3.2. Estimation of Distribution of Fitted Parameters

In order to test whether the study sample is statistically significant, it is necessary to analyse the probability distribution of parameter fitting results. Consider a candidate distribution for the set of fitted parameters with a probability density function $f(x/p)$ and a log-likelihood function $l(p/x)$. The optimal parameter $\hat{p}$ of this candidate distribution can be obtained using the following equation:

$$\hat{p} = \underset{p}{\operatorname{argmax}} l(p|x) \quad (26)$$

Let $\hat{f}(x|\hat{p})$ denote the optimal estimate of the probability density function $f(x|p)$. The quality of the estimate can be tested using the following equation:

$$SSE = \sum_{i=1}^{N_H} \left(\hat{f}(x_i|\hat{p}) - h(x_i) \right)^2 \quad (27)$$

where SSE denotes the sum of the squares of the errors; $h(x)$ denotes the probability density obtained from a histogram; and N_H denotes the number of bins of the histogram.

The optimal estimate minimizes SSE and is denoted as $\hat{\mathcal{P}}$. The correctness of the estimation can be evaluated by the Kolmogorov–Smirnov test. For a dataset $\mathcal{D}$ with a sample size n and a reference distribution $\mathcal{P}$, the Kolmogorov–Smirnov test is [20]: $H_0 : \mathcal{D}$ and $\mathcal{P}$ have the same distribution and $H_1 : \mathcal{D}$ and $\mathcal{P}$ have different distributions.

The tested statistics are as follows:

$$D_n = \sup_x |F_n(x) - F(x)| \quad (28)$$

where $F_n(x)$ is an empirical distribution function for the data, and $F(x)$ is the cumulative distribution function of the reference distribution $\mathcal{P}$. Let the fitted parameters corresponding to the dataset $\mathcal{D}$ and $\hat{\mathcal{P}}$ denote the optimal reference distribution for the reference distribution $\mathcal{P}$. When $p \leq 0.05$ in Equation (28), H_0 is rejected, that is, the estimated results may not conform to the distribution of the fitted parameters. For $p > 0.05$, H_0 cannot be rejected, that is, the estimated results may conform to the distribution of the fitted parameters.

3. Results and Discussion

The Weiyuan shale gas block was considered as a case study. Flowback, production, fracturing, drilling, and geological data were collected for 214 horizontal wells as shown in Table 1. The flowback data of each well was fitted to yield the following daily flowback rate:

$$R_l(t) = 0.0121 \exp(-0.0293t) \quad (29)$$

where 0.0121 and 0.0293 are the mean values of $\hat{\gamma}$ and $\hat{\beta}$. The mean R^2 for all the fitting results is 0.9212, and the mean value of the mean absolute error is 0.0157, indicating a good fit. The distributions of the fitted parameters, $\hat{\gamma}$ and $\hat{\beta}$, are asymmetrical. Therefore, it is inappropriate to use normal distributions for $\hat{\gamma}$ and $\hat{\beta}$. The theoretical domain of $\hat{\gamma}$ and $\hat{\beta}$ is $(0, +\infty)$. To obtain the distributions of $\hat{\gamma}$ and $\hat{\beta}$, seven candidate distributions with support sets of $(0, +\infty)$ or $(0, +\infty)$ were selected, and the optimal parameters of each distribution were calculated by considering scaling but not translation, as shown in Table 2.

Table 2. Estimated probability densities for $\hat{\gamma}$ and $\hat{\beta}$.

Distribution		SEE, $\hat{\gamma}$	SEE, $\hat{\beta}$
Contrast	Normal	200,347.01	86,414.97
	Chi-square	109,125.65	52,896.92
	Exponential	147,857.30	73,137.37
	F	96,919.53	52,771.35
Candidate	Gamma	109,125.86	52,896.91
	Rayleigh	126,688.09	54,432.71
	Rice	126,703.31	55,320.32
	Inverse Gaussian	127,636.96	64,920.44

Table 2 shows that the F distribution is the optimal distribution for $\hat{\gamma}$ and $\hat{\beta}$, and the estimated results and optimal parameters are shown in Figure 4.

The Kolmogorov–Smirnov test was conducted on the empirical distributions of $\hat{\gamma}$ and $\hat{\beta}$, and the optimal distributions had p values of 0.8209 and 0.4793, respectively, which were both above 0.05. Therefore, the distributions in Figure 3 can be used as empirical distributions of $\hat{\gamma}$ and $\hat{\beta}$. As can be seen from Figure 3, the fitting results of 214 wells are statistically significant and can be further used for big data analysis. For the Weiyuan shale gas block, this empirical distribution can be used to predict the most likely distribution range of γ and β of wells that are about to be put into production and further predict the flowback curve.

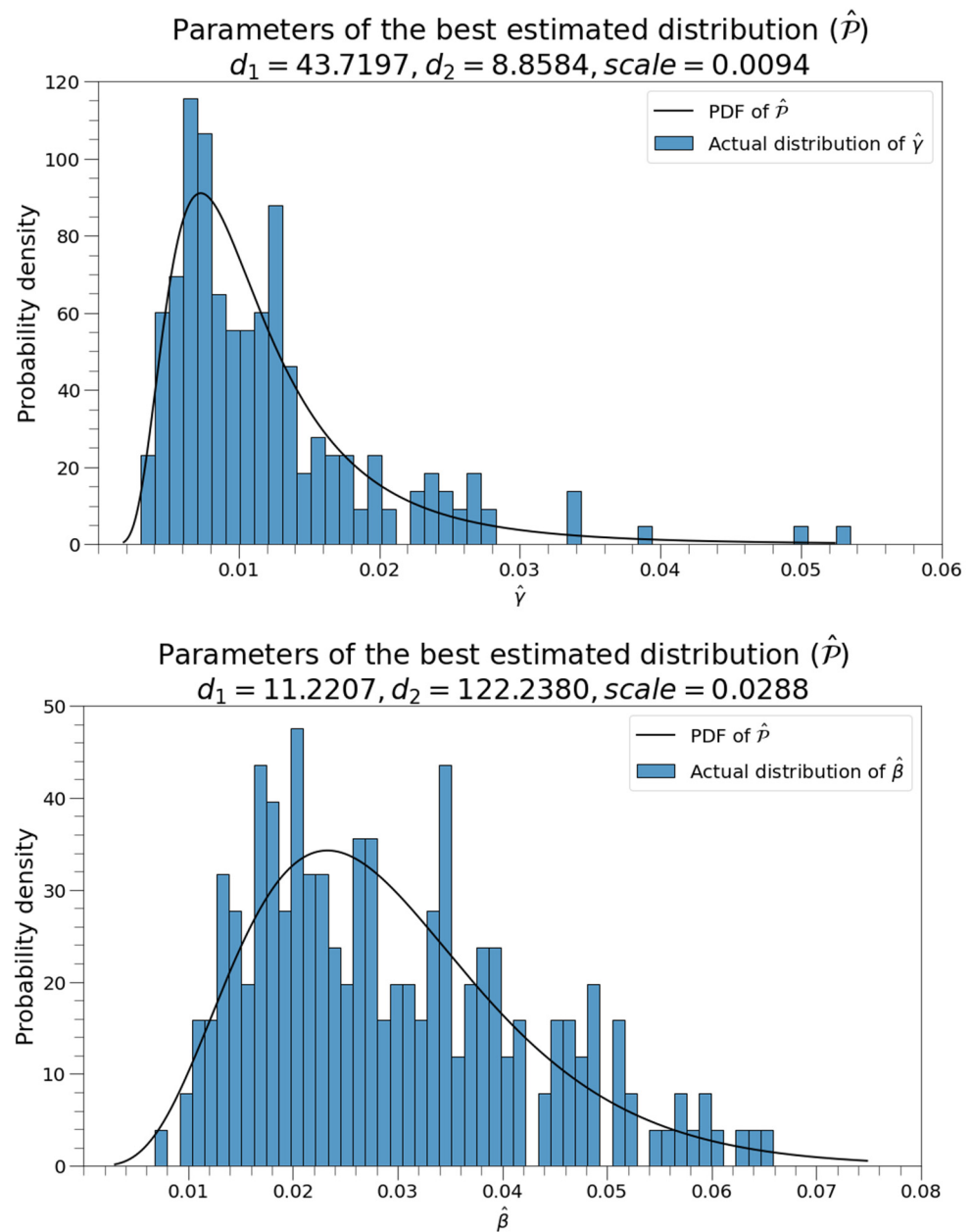


Figure 4. Best estimated distribution of $\hat{\gamma}$ and $\hat{\beta}$.

A physical interpretation of the results is that $\gamma e^{(-\beta)}$ corresponds to the flowback rate on the first day, which is also the peak daily flowback rate, and β corresponds to the speed at which the theoretical daily flowback decreases. Thus, the larger γ is, the higher the starting point of the flowback curve is; the larger β is, the lower the starting point of the flowback curve is, and the more rapidly the flowback decreases. Figure 5 shows the relationship between $\hat{\gamma}$ and $\hat{\beta}$.

Figure 5 shows a strong and significant correlation between $\hat{\gamma}$ and $\hat{\beta}$. Both parameters are related to the shape of the theoretical flowback curve and are therefore very likely to depend on each other. As the curve shape is jointly determined by two factors, the height of the starting point and the speed of decrease, neither factor can fully reflect the flowback volume. Hence, a variable composed of γ and β was used in this study to directly describe the flowback volume or flowback rate and is used in conjunction with static variables to perform a binary analysis. The fitting results are as follows: the minimum and maximum

values of $\hat{\beta}$ are 0.0068 and 0.0658, respectively, and the 0.01th and 99.99th percentiles of the estimated distribution are 0.0052 and 0.0686, respectively.

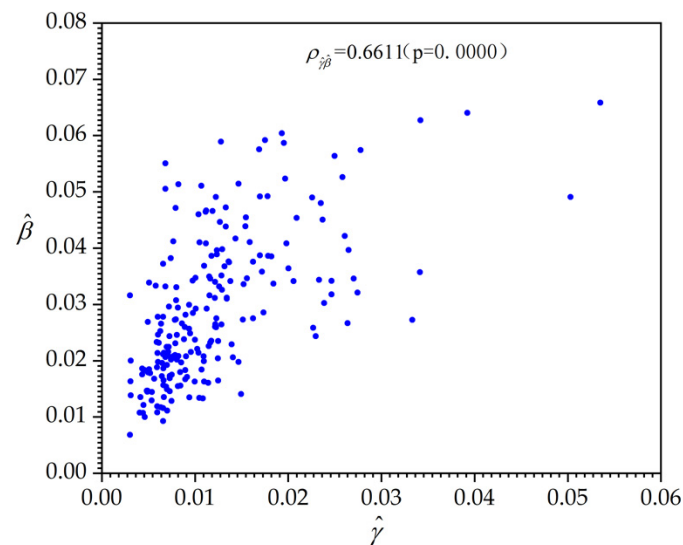


Figure 5. Correlation between $\hat{\gamma}$ and $\hat{\beta}$.

Figure 6 shows that the higher the value of β is for the theoretical curve, the more rapidly $f(N; \beta) = 1 - \frac{1}{e^{N\beta}}$ converges to 1. For $\beta = 0.0052$, $f(N; \beta)$ reaches 0.9945 at $N = 1000$, and for $\beta = 0.0686$, $f(N; \beta)$ reaches 0.9990 at $N = 100$, indicating that there is a minimal fracturing fluid volume after 1000 days of flowback for most shale gas wells. In Equation (19), $\gamma/(e^\beta - 1)$ corresponds to the approximate long-term cumulative flowback rate; thus, a flowback characteristic coefficient is defined, $\eta = \gamma/(e^\beta - 1)$, that reflects the flowback potential of a horizontal well.

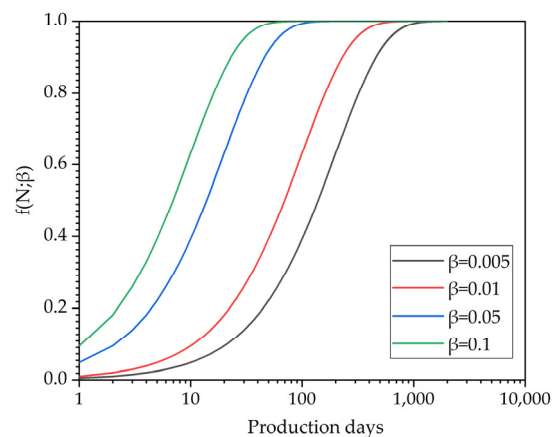


Figure 6. Sensitivity analysis of $f(N; \beta)$.

The flowback curve can be simplified by flowback characteristic coefficient. With η as the analysis object, the Spearman coefficient can be used to analyse the correlation between flowback rates and geological and engineering parameters and find the main affecting factors.

The correlation between η and the 14 static parameters in Table 1 was analysed using the Spearman correlation coefficient method. For comparison, the correlation between the first-year average daily production rate and 14 static parameters was also studied. Table 3 shows the correlations between these two variables and the static parameters, where the p value is given in parentheses. Static parameters with nonsignificant correlations with the two variables were excluded.

Table 3. Correlations between η , first-year average daily production rate, and static parameters.

Variable	η	First-Year Average Daily Production Rate
Fracturing lateral length	−0.1767 (0.0096)	0.4435 (0.000)
Number of fracturing stages	−0.2102 (0.0020)	0.4068 (0.000)
Fracturing stage intervals	0.1991 (0.0034)	−0.1036 (0.1308)
Gas saturation	−0.1061 (0.1216)	0.2536 (0.0002)
brittle mineral content	−0.1746 (0.0105)	0.3302 (0.0000)
Pressure gradient $\times$ vertical depth	0.1923 (0.0047)	−0.0335 (0.6265)
Thickness $\times$ drilling length of high-quality reservoir	−0.2928 (0.0002)	0.6400 (0.000)

Table 2 shows that the first-year average daily production rate is highly correlated with the thickness $\times$ reservoir drilling length of high-quality reservoir, and the correlation coefficient exceeds 0.5, indicating that the larger the thickness $\times$ drilling length of the high-quality reservoir, the higher the first-year average daily production rate, which is consistent with the results of Ma et al. [21]. The correlation between the static parameters and the first-year daily production rate can be sorted by the correlation coefficient as follows: thickness $\times$ drilling length of high-quality reservoir > fracturing lateral length > number of fracturing stages > brittle mineral content > gas saturation > fracturing stage interval > pressure coefficient $\times$ vertical depth. The drilling length of the high-quality reservoir, the fracturing lateral length, the number of fracturing stages, and the fracturing stage interval are engineering parameters. The better the drilling effect, the better fracturing effect and the higher the first-year average daily production rate. High-quality reservoir thickness, brittle mineral content, gas saturation and pressure coefficient $\times$ vertical depth are geological parameters. The better physical properties of shale gas reservoir, the higher first-year average daily production rate. This is why there is a good correlation between the first-year average daily production rate and various parameters.

Although the correlation coefficients between η and the static parameters are less than 0.5, relatively speaking, η and the thickness $\times$ drilling length of high-quality reservoir are best correlated, and the correlation symbol is opposite to that of the first-year daily production rate, which means that the higher the first-year average daily production rate, the lower the flowback rate. The correlation between static parameters and η is sorted by coefficient, which is basically consistent with that of first-year average production rate, but the positions of individual parameters are interchanged. The order according to the absolute value of correlation is thickness $\times$ drilling length of high-quality reservoir > number of fracturing stages > pressure coefficient $\times$ vertical depth > fracturing stage interval > fracturing lateral length > brittle mineral content > gas saturation. η is positively correlated with the fracturing stage interval, which means that the larger the fracturing stage interval, the worse the fracturing effect. The hydraulic fracturing process does not create complex fracture networks, and the fracturing fluids mainly concentrate near the main fracture and wellbore, resulting in high flowback rate. η is positively correlated with the pressure coefficient $\times$ vertical depth. The pressure coefficient $\times$ vertical depth represents the formation pressure. The higher the formation pressure, the stronger the liquid carrying capacity of the shale gas well and the higher the flowback rate. η is negatively correlated with the fracturing lateral length, the number of fracturing stages, and the content of brittle minerals. The higher the brittle minerals, the easier the reservoir is to be fractured; the longer the fracturing lateral length is and the more the number of fracturing stages is, the more complex the fracture network will be formed after hydraulic fracturing and the larger the SRV will be. The fracturing fluids are bound in complex micro fractures and cannot be discharged, resulting in a low flowback rate.

Table 4 shows the correlations between η and the whole static parameters in Table 1. Figure 6 shows the hot map of correlations between η and static parameters. From Table 1 and Figure 7, it can be seen that η is well correlated with the 30 day, 90 day, 180 day, 360 day, and peak gas production flowback rates. The longer the flowback time, the better the correlation between η and the flowback rate. The correlation between η and the 30 day, 90, 180, and 360 day flowback rate confirmed that it is reasonable and scientific to use η to characterize the flowback rate. η is positively correlated with vertical depth, TOC content, porosity, pressure coefficient, and average fracturing intervals, which means the higher these parameters are, the higher the flowback rates are. η is negatively correlated with high-quality reservoir thickness, gas saturation, brittle mineral content, fracturing fluid intensity, proppant intensity, average hydraulic fracturing fluid displacement, and drilling length in high-quality reservoirs, which means the higher these parameters are, the lower flowback rates are. η is negatively correlated with the first-year average daily production rate and EUR, but the correlation coefficients are low. It means that flowback rate is not merely correlated with gas production.

Table 4. Correlations between η and static parameters.

Variable	η
30 day flowback rate	0.7871 (0.0000)
90 day flowback rate	0.9294 (0.0000)
180 day flowback rate	0.9759 (0.0000)
360 day flowback rate	0.9950 (0.0000)
Peak gas production flowback rate	0.7352 (0.0000)
Fracturing lateral length (m)	−0.1767 (0.0096)
Number of fracturing stages	−0.2102 (0.0020)
First-year average daily production rate ($10^4\text{m}^3/\text{d}$)	−0.0544 (0.4282)
Vertical depth (m)	0.1954 (0.0041)
TOC content (%)	0.1154 (0.0922)
Porosity (%)	0.0613 (0.3722)
High-quality reservoir thickness (m)	−0.1478 (0.0037)
Gas saturation (%)	−0.1061 (0.1216)
Pressure coefficient	0.1849 (0.0067)
Brittle mineral content (%)	−0.1746 (0.0105)
Average fracturing interval (m)	0.1991 (0.0034)
Fracturing fluid intensity (m^3/m)	−0.0914 (0.1827)
Proppant intensity (t/m)	−0.1021 (0.1366)
Average hydraulic fracturing fluid displacement (m^3/min)	−0.1020 (0.1370)
Drilling length in high-quality reservoir (m)	−0.2640 (0.0009)
EUR (10^8m^3)	−0.0918 (0.1811)

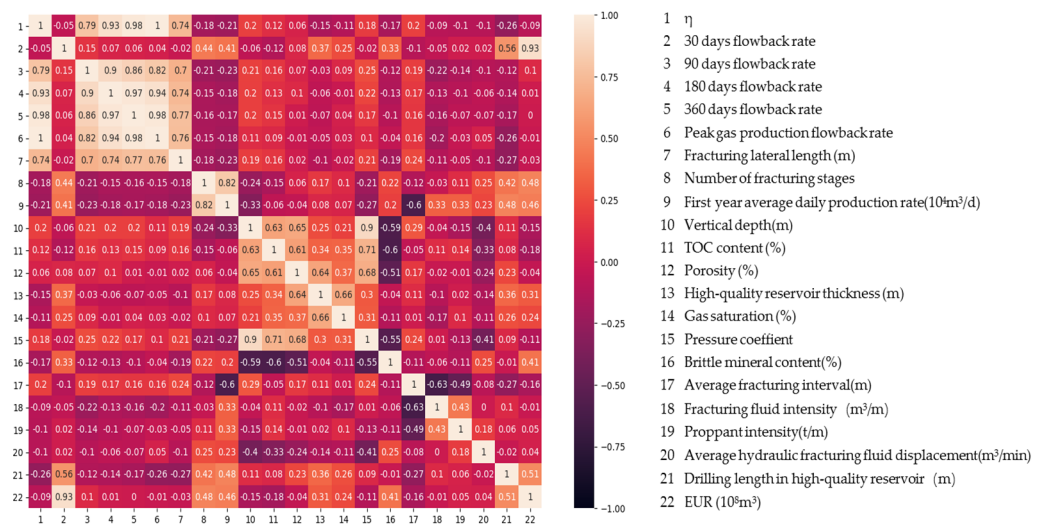


Figure 7. Hot map of correlations between η and static parameters.

From the correlation between various static parameters and the first-year average daily production rate and η , it is reasonable to simplify and characterize the flowback characteristics of shale gas wells by the flowback characteristic coefficient η . In our previous study [17], based on the actual data in the Weiyuan shale gas block, combined with a deep learning algorithm, two different depth feedforward neural networks were designed, which are the single output neural network and the multi-output neural network, to predict the shale gas well flowback rate. The 30 day, 90, 180, 360 day, and peak gas production flowback rates of shale gas wells were predicted. This method can predict the flowback rate of a specific day, but the main factors affecting the flowback rate at any time cannot be obtained. Combining the deep learning algorithm and the backflow characteristic coefficients proposed in this article, the flowback rate at any time in the future can be predicted. The method proposed in this paper also has some limitations. For gas wells with intermittent shut-in or long-term shut-in or for long-term flooded gas wells, there will be large errors in the prediction of the flowback rate. In addition, the research does not consider the impact of gas–liquid two-phase flow on flowback.

4. Summary and Conclusions

Based on the seepage theory of fracturing fluid in multi-stage fractured horizontal wells, the study uses convolution and Laplace transform methods to abstract the flowback curve into two characteristic parameters, the daily flowback rate coefficient γ and the flowback decline coefficient β . Taking the Weiyuan shale gas block as a study case, the flowback data of 214 wells were fitted, and the distribution characteristics of the fitted characteristic parameters were studied. The flowback characteristic coefficients to characterize the flowback potential of shale gas wells were established. The Spearman correlation coefficient method was used to study the correlation between the geological and engineering static parameters of 214 wells, the characteristic flowback coefficients, and the first-year average daily production rate. There are several conclusions obtained from this research:

(1) The fitting results of the flowback curve for 214 production wells show that the average daily flowback rate coefficient and the flowback decline coefficient of all wells are 0.0121 and 0.0293, and the average value of R^2 of all fitting results is 0.9212, the mean value of the mean absolute error is 0.0157, and the fitting effect is better. Both the daily flowback rate coefficient γ and the flowback decline coefficient β are right-skewed distributions.

(2) The comparative study on the correlation between the flowback characteristic coefficient and the first-year average daily production rate and static parameters shows that the thickness $\times$ the drilling length of a high-quality reservoir is best correlated with the characteristic flowback coefficient and the first-year average daily production rate, and their correlation symbol is opposite, which means that the larger the thickness $\times$ the drilling

length of the high-quality reservoir, the higher the average first-year daily production rate and the lower the flowback rate of the shale gas well. The factors affecting the flowback rate mainly include geological factors and engineering factors. The order of correlation coefficients is as follows: thickness $\times$ drilling length of high-quality reservoir, number of fracturing stages, pressure coefficient $\times$ vertical depth, fracturing stage interval, fracturing lateral length, brittle mineral content, and gas saturation.

(3) Through the method established in this paper, the shale gas flowback curve containing complex flow mechanism can be abstracted into simple characteristic parameters and characteristic coefficients. The method proposed in this paper can provide a novel way for machine learning and other big data analysis methods to study the flowback characteristics of shale gas horizontal wells. In future studies, the time-dependent gas output will be related to the flowback rate. Combining with machine learning, the flowback rate and gas output at any time can be predicted in the future.

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

Funding: This research was financially supported by the R&D Department of Petrochina (No. 2021DJ2005).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

Nomenclature

K	permeability, m^2
Φ	porosity, dimensionless
t	time, day
μ	viscosity, $\text{Pa}\cdot\text{s}$
C_t	comprehensive compressibility coefficient, dimensionless
x_{fi}	half-length of the i th fracture, m
P	pressure, Pa
P_D	dimensionless pressure, dimensionless
t_D	dimensionless time, dimensionless
x_D	dimensionless distance in x direction, dimensionless
y_D	dimensionless distance in y direction, dimensionless
C	wellbore storage constant, Pa/m^3
C_D	dimensionless wellbore storage constant, dimensionless
S	total skin factor of the fracture system, dimensionless
u	Laplace transform variable
α	daily flowback volume coefficient, m^3/d
β	decreasing flowback coefficient, $1/\text{d}$
q	daily flowback volume, m^3/d
Q_{total}	total quantity of injected fracturing fluids, m^3
γ	daily flowback rate coefficient, $1/\text{d}$
ρ	Spearman correlation coefficient, dimensionless
R_l	daily flowback rate, dimensionless
η	flowback characteristic coefficient, dimensionless

References

- Williams-Kovacs, J.D.; Clarkson, C.R. A modified approach for modeling two-phase flowback from multi-fractured horizontal shale gas wells. *J. Nat. Gas Sci. Eng.* **2016**, *30*, 127–147. [\[CrossRef\]](#)
- Liu, J.; Li, P.; Sun, Z.; Lu, Z.; Du, Z.; Liang, H.; Lu, D. A new method for analysis of dual pore size distributions in shale using nitrogen adsorption measurements. *Fuel* **2017**, *210*, 446–454. [\[CrossRef\]](#)
- Rongze, Y.; Xiaowei, Z.; Yanan, B.; Yang, L.; Mingxiang, H. Flow mechanism of shale gas reservoirs and influential factors of their productivity. *Nat. Gas. Ind.* **2012**, *32*, 10–15.
- Zoback, M.D.; Kohli, A.; Das, I.; McClure, M. The importance of slow slip on faults during hydraulic fracturing stimulation of shale gas reservoirs. In Proceedings of the SPE Americas Unconventional Resources Conference, Pittsburgh, PA, USA, 5–7 June 2012.
- Du, W.Y.; Fu, J.; Sun, Z.; Qiao, N. Shale gas seepage mechanism and transient pressure analysis. *Chin. J. Comput. Phys.* **2015**, *32*, 51–57.
- Jia, B.; Chen, Z.; Xian, C. Investigations of CO₂ storage capacity and flow behavior in shale formation. *J. Pet. Sci. Eng.* **2022**, *208*, 109659. [\[CrossRef\]](#)
- Dai, J.; Zou, C.; Liao, S.; Dong, D.; Ni, Y.; Huang, J.; Wu, W.; Gong, D.; Huang, S.; Hu, G. Geochemistry of the extremely high thermal maturity Longmaxi shale gas, southern Sichuan Basin. *Org. Geochem.* **2014**, *74*, 3–12. [\[CrossRef\]](#)
- Kohlia, A.; Arora, P. Application of artificial neural net-works for well logs. In Proceedings of the International Petroleum Technology Conference, Doha, Qatar, 19 January 2014.
- Hu, J.; Lihui, D. Oil reservoir water flooding flowing area identification based on the method of streamline clustering artificial intelligence. *Pet. Explor. Dev.* **2018**, *45*, 312–319.
- Adibifard, M.; Tabatabaei-Nejad, S.; Khodapanah, E. Artificial Neural Network (ANN) to estimate reservoir parameters in Naturally Fractured Reservoirs using well test data. *J. Pet. Sci. Eng.* **2014**, *122*, 585–594. [\[CrossRef\]](#)
- Ghaffarian, N.; Eslamloueyan, R.; Vaferi, B. Model identification for gas condensate reservoirs by using ANN method based on well test data. *J. Pet. Sci. Eng.* **2014**, *123*, 20–29. [\[CrossRef\]](#)
- Tian, C.; Horne, R.N. Recurrent neural networks for permanent downhole gauge data analysis. In Proceedings of the SPE Annual Technical Conference and Exhibition, San Antonio, TX, USA, 9–11 October 2017.
- Thanh, H.V.; Lee, K.K. Application of machine learning to predict CO₂ trapping performance in deep saline aquifers. *Energy* **2021**, *239*, 122457. [\[CrossRef\]](#)
- Bai, B.; Goodwin, S.; Carlson, K. Modeling of frac flowback and produced water volume from Wattenberg oil and gas field. *J. Pet. Sci. Eng.* **2013**, *108*, 383–392. [\[CrossRef\]](#)
- Zhou, Q.; Dilmore, R.; Kleit, A.; Wang, J.Y. Evaluating fracture-fluid flowback in Marcellus using data-mining technologies. *SPE Prod. Oper.* **2016**, *31*, 133–146. [\[CrossRef\]](#)
- Lin, B.; Guo, J.; Liu, X.; Xiang, J.; Zhong, H. Prediction of flowback ratio and production in Sichuan shale gas reservoirs and their relationships with stimulated reservoir volume. *J. Pet. Sci. Eng.* **2020**, *184*, 106529. [\[CrossRef\]](#)
- Yuyang, L.; Xinhua, M.; Xiaowei, Z.; Wei, G.; Lixia, K.; Rongze, Y.; Yuping, S. Shale gas well flowback rate prediction for Weiyuan field based on a deep learning algorithm. *J. Pet. Sci. Eng.* **2021**, *203*, 108637. [\[CrossRef\]](#)
- Li, Q.; Li, P.; Pang, W.; Li, D.; Liang, H.; Lu, D. A new method for production data analysis in shale gas reservoirs. *J. Nat. Gas Sci. Eng.* **2018**, *56*, 368–383. [\[CrossRef\]](#)
- Lu, D.T. *Modern Well Testing Theory and Application*; Petroleum Industry Press: Beijing, China, 2009.
- Zwillinger, D.; Kokoska, S. *Standard Probability and Statistics Tables and Formulae*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2000; pp. 346, 366–367.
- Xinhua, M.A.; Xizhe, L.I.; Liang, F.; Yujin, W.A.N.; Qiang, S.H.I.; Yonghui, W.A.N.G.; Zhang, X.; Mingguang, C.H.E.; Wei, G.U.O. Dominating factors on well productivity and development strategies optimization in Weiyuan shale gas play, Sichuan Basin, SW China. *Pet. Explor. Dev.* **2020**, *47*, 594–602.