Energy Engineering

Tech Science Press

Doi:10.32604/ee.2025.074170

ARTICLE

Statistical Modeling and Prediction of Hydraulic Fracture Propagation in Carbonate Reservoirs

V. V. Poplygin[1,*], A. Dieng[2], Min Wang[3] and Xian Shi[3]

[1]Kogalym Educational Center, Perm National Research Polytechnic University, Kogalym, 628482, Russia
[2]Department of Oil and Gas Technologies, Perm National Research Polytechnic University, Perm, 614990, Russia
[3]China University of Petroleum (East China), Qingdao, 266580, China

*Corresponding Author: V. V. Poplygin. Email: poplygin@bk.ru

Received: 04 October 2025; Accepted: 11 December 2025; Published: 18 June 2026

ABSTRACT: Hydraulic fracturing in carbonate reservoirs presents unique challenges due to their complex pore structures and heterogeneous mechanical properties. This paper explores the application of statistical methods to improve fracture prediction and optimization in carbonate formations. Hydraulic fracturing is actively carried out on these formations. In order to properly plan hydraulic fracturing, it is necessary to identify the main factors affecting oil production after hydraulic fracturing. This study introduces an integrated framework combining information amount theory (IAT) and Gray relational analysis (GRA) to identify and rank the dominant parameters controlling hydraulic fracturing performance in heterogeneous carbonate reservoirs. Based on a dataset of twenty-one fractured wells in the Perme region, twelve geological and operational parameters were evaluated to determine their impact on post-fracturing oil production rate. Results consistently indicate that fracturing fluid volume and fracture width exert the greatest influence, while fracture length ranks lower due to the complex fracture networks typical of carbonates. The proposed IAT-GRA method offers a computationally efficient, interpretable tool for data-limited reservoirs, and the findings provide clear engineering guidelines for optimizing hydraulic fracturing design and execution. The study used the theory of the amount of information and the gray method to identify the main factors influencing the results of hydraulic fracturing. Having data on core parameters before hydraulic fracturing, it is possible to predict the results of hydraulic fracturing with a high degree of reliability. The regression model is based on the method of multiple linear regression. Oil production after hydraulic fracturing increases with an increase in the flow rate of oil after hydraulic fracturing and the width of the crack. Hydraulic fracturing creates multiple cracks and microcracks, forming a complex network of cracks in the formation. Therefore, the use of statistical methods helps to make an operational assessment of the result, but does not negate the use of more accurate and complex models.

KEYWORDS: Oil reservoir; hydraulic fracturing; information amount theory; water cut

1 Introduction

1.1 The Critical Role of Hydraulic Fracturing in Modern Energy Security

Hydraulic fracturing (HF), commonly known as fracking, stands as one of the most transformative well stimulation technologies developed in the oil and gas industry over the past half-century. It is a cornerstone technique for the economic development of unconventional resources and the enhanced recovery from mature conventional fields. The fundamental principle of HF involves the high-pressure injection of a specialized fluid, laden with propping agents (proppant), into a target reservoir formation. This process induces tensile failure in the rock, creating a network of conductive fractures that serve as high-permeability pathways, thereby bypassing near-wellbore damage and connecting a larger volume of the reservoir to the wellbore. This dramatic increase in effective permeability is what transforms otherwise non-productive or marginally economic formations into prolific sources of hydrocarbons.

The global significance of HF cannot be overstated. As conventional hydrocarbon reserves continue to deplete and global energy demand persists, the ability to unlock vast resources trapped in low-permeability "tight" rocks—such as shales, tight sandstones, and complex carbonates—has redefined global energy markets. Hydraulic fracturing has single-handedly enabled the "shale revolution," particularly in North America, leading to a surge in oil and gas production and contributing significantly to national energy security. Beyond unconventional plays, HF is equally critical in revitalizing mature conventional fields, where it is used to enhance recovery factors, manage water production, and improve sweep efficiency. Despite its well-documented benefits, the application of hydraulic fracturing is a complex, high-stakes undertaking, accompanied by significant challenges including environmental concerns (e.g., water usage, potential aquifer contamination, and induced seismicity), substantial capital and operational expenditures, and a pressing need for highly precise forecasting to optimize treatment designs and mitigate economic and operational risks [1].

1.2 The Unique Challenge of Carbonate Reservoir Stimulation

The challenges associated with HF are magnified when applied to carbonate reservoirs, which are estimated to hold over 60% of the world's conventional oil reserves and a significant portion of its gas reserves. Carbonates, primarily composed of limestone ($CaCO_3$) and dolomite ($CaMg(CO_3)_2$), are fundamentally different from the more homogeneous sandstone reservoirs. They are characterized by extreme geological heterogeneity across multiple scales, resulting from complex post-depositional diagenetic processes such as dissolution, dolomitization, and cementation. This heterogeneity manifests in complex pore systems that are a triple-porosity mix of interparticle matrix porosity, vugs (large, irregular pores), and natural fractures.

This inherent complexity dictates highly irregular and often unpredictable fracture propagation during stimulation treatments. An induced hydraulic fracture may be arrested by a hard, dense stylolite layer; it may be diverted, branched, or swallowed by a pre-existing network of natural fractures and faults; or it may exhibit excessive and uncontrolled height growth in a mechanically weak zone. These phenomena often lead to the creation of inefficient, complex fracture networks rather than the ideal, planar, bi-wing fracture geometry assumed in simple models. The consequences include poor proppant placement, inadequate fracture conductivity, early screen-outs, and ultimately, suboptimal production results that fail to justify the significant investment. The oilfields in the Perm region of Russia, which form the empirical basis for this research, are a textbook example of these challenges. They are mature assets characterized by high levels of reserve depletion and complex mining and geological conditions, involving the development of heterogeneous carbonate reservoirs with low capacitance-filtration properties and reservoirs containing high-viscosity oil. In this challenging context, traditional development methods like natural depletion or conventional water flooding have proven grossly inefficient, yielding disappointingly low recovery factors typically ranging from a mere 2.5% to 30% [2]. Consequently, for these fields, hydraulic fracturing is not merely an optimization tool but an essential, albeit high-risk, technological prerequisite for any economically viable development.

1.3 Limitations of Traditional Forecasting and Design Methods

Analytical models are built upon a set of simplifying assumptions—homogeneous and isotropic rock properties, linear elastic fracture mechanics, and simple, planar fracture geometry—that are consistently violated in heterogeneous carbonates. They fail to adequately account for the complex interactions between the hydraulic fracture and pre-existing natural fractures, the impact of rock plasticity, and the significant fluid loss into vuggy and fractured zones. Empirical methods, while useful within the specific geological context from which they were derived, lack generalizability and can produce highly misleading results when applied to a new field or a different reservoir unit with subtly different characteristics. This reliance on oversimplified models often leads to inaccurate predictions of key fracture dimensions (length, height, width) and consequent production rates. The operational ramifications are profound: suboptimal design decisions, such as inappropriate proppant and fluid selection, ill-advised pumping schedules, or misplaced perforation clusters, which result in poor well performance, heightened operational risks (e.g., screen-outs, casing failures), and a failure to maximize the economic return on a multi-million-dollar investment [3,4].

The success of an HF operation hinges on the accuracy of its design, which in turn relies on the ability to predict fracture propagation and resulting production. Traditional forecasting methods have historically depended on a combination of historical production data from analogous wells, analytical models (e.g., PKN, KGD), and empirical correlations [5,6]. While these methods provide a valuable starting point for initial screening and design, they possess inherent limitations that render them inadequate for the complexities of carbonate systems [7–9].

1.4 The Rise of Advanced Modeling and Data-Driven Approaches

In response to the shortcomings of traditional methods, the industry has witnessed the rapid development of two parallel advanced approaches: high-fidelity numerical modeling and data-driven machine learning (ML) techniques.

1.4.1 High-Fidelity Numerical Models

Sophisticated numerical simulators represent the state-of-the-art in physically modeling the HF process. These tools, such as those based on the Finite Element Method (FEM), Discrete Ele-ment Method (DEM), and more recently, the Numerical Manifold Method (NMM) and Ele-ment-Free Galerkin (EFG) methods, can incorporate a vast array of physical phenomena. They model fully coupled hydro-mechanical processes, simulate complex fracture networks interact-ing with natural discontinuities, and account for rock anisotropy and inhomogeneity. The primary advantage of these models is their high potential predictive accuracy. Their primary disadvantage, however, is their immense computational cost and their insatiable appetite for detailed, high-quality input parameters (e.g., precise *in-situ* stress profiles, full mechanical rock properties, detailed characterization of natural fractures), which are often prohibitively ex-pensive or technically impossible to acquire at the required resolution for every well.

1.4.2 Data-Driven and Statistical Machine Learning Methods

Concurrently, the advent of big data analytics and machine learning has provided a powerful complementary approach. ML algorithms can learn complex, non-linear relationships between input parameters (e.g., treatment data, reservoir properties) and output outcomes (e.g., production) directly from field data, without requiring an explicit physical model. Li et al. [10,11] showcased this powerfully by integrating numerical simulation with ML proxy models (including neural networks and random forests) to predict Net Present Value (NPV), achieving a 16% improvement in prediction accuracy over standalone neural networks and demonstrating superior resilience to noisy data. Guo et al. [12] effectively used Grey Relational Analysis (GRA) and Principal Component Analysis (PCA) to identify key influencing factors on tight oil production, followed by a genetic algorithm to optimize fracturing parameters. These data-driven methods offer the advantage of computational speed and the ability to handle “messy” real-world data, making them highly suitable for rapid screening and operational decision-support.

1.5 Research Gap and Novelty

While both advanced numerical models and ML offer significant advances, a gap remains for pragmatic, robust methodologies that balance predictive power with practical applicability, especially in mature assets where data may be abundant but of variable quality, and where computational or budgetary resources for high-end simulation are limited. There is a particular need for techniques that can provide engineers with clear, interpretable insights into which parameters most significantly control HF success.

This study seeks to address this gap by applying and integrating two powerful yet underutilized statistical methodologies—**Information Amount Theory (IAT)** and **Grey Relational Analysis (GRA)**. IAT, rooted in information theory, provides a quantitative measure of the "information" each input variable contributes to reducing the uncertainty in the output (e.g., production rate). It is exceptionally effective for ranking the influence of various parameters in systems with limited data. GRA, a pillar of Grey System Theory, is designed to analyze systems with partial information and small samples. It measures the strength of relationships between variables based on the similarity of their data sequences, making it ideal for identifying which operational and geological factors most closely "follow" the trend of successful production outcomes.

The novelty of this work lies in the synergistic application of IAT and GRA to the problem of HF optimization in complex carbonate reservoirs. While some studies have used similar methods in silos, their integrated use provides a more robust validation: IAT identifies the most influential parameters, and GRA independently verifies the strength of their relationship with the production target. This dual-method approach mitigates the limitations of using a single statistical technique and provides a more reliable, data-driven foundation for decision-making.

1.6 Objectives and Structure of the Study

The primary objectives of this research are:

1. To compile and analyze a comprehensive dataset of hydraulic fracturing treatments and their outcomes from carbonate reservoirs in the Perm region.
2. To apply **Information Amount Theory** to quantitatively rank a suite of twelve key parameters (including fracture geometry, fluid and proppant volumes, and pre-frac production indices) based on their influence on post-fracturing oil production rate (Qo_post-frac).
3. To employ **Grey Relational Analysis** to independently assess and validate the relational strength between these parameters and the production outcome.
4. To synthesize the findings from both methods to provide a robust, prioritized list of the most critical success factors for HF in the studied carbonates, thereby offering a practical tool for optimizing future fracturing designs.

This paper is structured as follows: Following this introduction, Section 2 describes the geological setting of the study area and the dataset used for analysis. Section 3 details the theoretical framework and computational steps of Information Amount Theory, presenting the ranking results. Section 4 similarly elaborates on Grey Relational Analysis and its findings. Finally, Section 5 provides a comprehensive discussion, synthesizes the conclusions, and outlines practical recommendations and avenues for future research. By pursuing this roadmap, this study aims to contribute a practical, data-driven framework for enhancing the efficiency and economic viability of hydraulic fracturing in the world's challenging, yet critically important, carbonate reservoirs.

2 Study Area

According to micro–descriptions of the first study reservoir, the oil saturated part of the reservoir is composed predominantly of limestones (more than 60) and dolomites and rare siltstones. Limestones are overwhelmingly biomorphic, dominated by alga-foraminiferous limestone, with clotty-algal and detrital-foraminiferous varieties. Dolomites are micro-fines grained, less often micro-granular and fine micro-granular, often with quartz admixture of silt rock size (up to 15%–20%), with sulfate nests, coal admixture, rarely calcareous. Fine grained sand siltstones with dolomite and clay cement, porous. In the reservoir, the porosity and the permeability have a values of 16 and 19 × 10^{-3} μm^2 and the pay net value represents 3 m.

The oil saturated part of the second study reservoir consists of dolomites biomorphic, clotty limestone and much less frequently coagulated and organo-detritic limestone, according to micro-descriptions. Dolomites have fine grains and micro-fine grains, more rarely micro-fine grains. In the reservoir, the porosity and the permeability have a values of 19 and 7.3 × 10^{-3} μm^2 and the pay net value represents 4 m.

The productive formations of the reservoirs studies horizons are characterized by low natural oil saturation and productivity are represented in Table 1.

Table 1: Geological and physical characteristics

	1st Reservoir	2nd Reservoir
Average depth, m	1101	1026
Deposit type	Layer uplifted deposit	Layer uplifted deposit
Reservoir type	Gross and carbonate reservoir rock	Gross and carbonate reservoir rock
Absolute elevation of OWC	−890	−812
Average oil net pay, m	3	4
Initial formation temperature	26.5	25
Initial formation pressure, MPa	11.7	11.2
Average oil saturation	0.671	0.681
Porosity, %	16.6	20.4
Permeability, μm^2	24.2 × 10^{-3}	18.7 × 10^{-3}
Oil viscosity, MPa·s	42.8	18.2
Oil density in reservoir conditions, kg/m^3	892	880
Oil volume factor,	1.027	1.026
Bubble pressure with gas, Mpa	5.13	7.52
Gas oil ratio, m^3/t	8.1	12.6

Application of HF technology in the field has been actively in 2015 at the development target to enhance oil recovery.

Results show fracture length, fracture width, Propant Height, bubble pressure, pre-frac productivity index, Proppant Total, Specific polymer consumption for placement of 1 t of proppant, Main Frac fluid volume, pre-frac fluid production rate, pre-frac water cut and pre-frac oil production rate as secondary parameters (Table 2).

Table 2: Parameters for oil wells fractured of the productive formations in the reservoirs carbonate deposits of one field in the Perm region

No Well	Qo_post-frac, t/d	Lf, m	Wf, mm	Hf, m	h, mm	Pb, MPa	Ip, m^3/d*MPa	mp, t	qp, Kg/t	Vf	Qf, t/d	Wc, %	Qo_pre-frac, t/d
256	0.3	117.5	3	10.3	8.4	3.9	0.9	20	10.2	93	1.3	32.2	0.8
9044	1.6	251.6	2.8	5	4	7.5	0.7	23	10.5	116	2.8	2.3	2.5
16	2.7	153.5	1.8	5.4	3.4	5.3	0.5	26	10.6	111	0.3	22.9	0.2
112	3.7	287.2	2.8	4.7	3.2	7.5	4.0	25	11.3	123	16.0	90.2	1.4
510	4.1	163.5	3	4.9	5	5.3	0.5	26	10.6	119	2.9	11.2	2.3
29	4.9	219.5	2.5	5	4.2	4.2	0.6	29	10.6	132	1.9	41.9	1.0
452	5.5	248.7	2.4	5.5	3.8	5.3	0.2	24	9.7	104	1.0	33.0	0.5
483	5.6	152.9	3	4.4	3.2	5.3	0.4	27	8.7	111	0.5	20.4	0.3
471	5.6	175.4	2.5	4.9	3.8	5.0	1.1	29	10.5	111	1.8	20.5	1.3
515	5.7	292.9	3.9	4.9	4	5.3	0.6	30	9.4	117	0.9	11.3	0.7
484	5.8	132.2	2	4.5	3.8	5.3	1.4	29	11.7	152	1.3	10.5	1.2
73	6.1	149.2	2.6	4	3	5.3	5.0	29	9.5	118	3.0	36.0	1.7
165	6.2	246.8	2.6	4.9	3.8	5.3	1.6	25	9.1	101	4.0	48.0	1.7
503	6.6	273.0	3.5	4.5	4.4	5.3	0.5	32	9.7	133	2.4	25.7	1.6
73	6.7	213.0	2.7	10.5	2.6	7.52	4.4	26	9.8	118	3.7	36.0	2.0
29	7.0	213.4	4.2	5.9	2.6	7.5	0.6	20	11.0	129	1.7	32.6	1.0
522	8.2	160.9	4.4	10.3	3.8	7.52	0.1	26	9.8	123	0.6	19.0	0.4
484	8.3	111.1	5.3	0.5	4.4	5.3	1.4	34	10.8	139	1.9	10.5	1.7
451	8.6	170.7	4.2	5	4	4.5	1.9	31	8.5	120	5.2	22.7	3.6
50	8.8	112.8	3.8	4.5	4.2	6.3	0.9	25	9.8	109	4.1	22.2	2.9
16	11.3	193.4	2.5	6.6	3.4	7.52	0.5	25	12.0	135	0.8	23.8	0.6

To normalize the Qo_post-frac in the Table 1, used typically of the following methods: Max–Min normalization: scales the values to a fixed range, usually.

3 Information Amount Theory

Information amount (IAT) is a quantitative approach used to measure the amount of information that a variable contributes to a system, focusing on the amount of information on each variables change interval and the total amount information. The total amount of information on each variables change interval, a concept from information theory, quantities the level of uncertainty and thus, more potential total amount information when the variables are considered. Correlation coefficients and mutual information, which are effective for evaluating dependent relationships, are used to assess the correctness of the model [13].

Liang et al. calculated and verified the effects of productivity parameters of fractured horizontal wells in Bakken tight oil reservoirs using information amount theory. The results show that the method is effective in verifying the impact a variety parameters on the wells productivities [14]. The authors have proved that information amount theory, orthogonal experiment design (OED) and Grey Relational Analysis have similar effects and suggest they are practical and consistent.

In order to create a multi-variable model to estimate the oil production post-fracking increase using Z–score normalization, the wells sample is divided into two groups, comprising wells have negative values indicating that Qo_post-frac value is below the mean 5.87 t/day group A and wells have positive values indicating that Qo_post-frac above the mean t/day group B.

The following is a description of the fundamental information amount theory process: divide the wells into two groups or intervals, A and B, based on a criterion; count the frequency of the factors in each group; compute the frequencies again to verify the degree of difference between group A and B. The greater degree of the difference. The more different, the bigger information amount is, and the bigger influence degree is. Calculate and analyze the amount of information for each factor using this method. Procedures are as follows: Each factor will be counted separately in different ranges and the frequencies of group A and group B will be calculated. Further calculation can obtain the distribution of the difference between A and B. The difference is greater so the amount of information is greater. The IAT framework has been successfully used in similar contexts [5,6].

The calculation steps are as follows:

– Count the separate frequency of parameters in groups A and B.
– Convert the frequency into probability (%) $y_{\mathrm{A}\mathfrak{z}}$ and $y_{\mathrm{B}\mathfrak{z}}$ where $\mathfrak{z}$ is the interval serial number.
– Calculate average probability ratio $\overline{y}_{\mathfrak{z}}$ in each interval. The formula is

$$\overline{y}_{A\mathfrak{z}} = 0.1(y_{\mathfrak{z}-2} + 2y_{\mathfrak{z}-1} + 4y_{\mathfrak{z}} + 2y_{\mathfrak{z}+1} + y_{\mathfrak{z}+2} \tag{1}$$

Calculate the average frequency ratio

$$\overline{y}_{\mathfrak{z}} = \overline{y}_{B\mathfrak{z}} \tag{2}$$

Calculate the diagnosis coefficient $Z_{\mathfrak{z}}$:

$$Z_{\mathfrak{z}} = 10\log(\overline{y}_{A\mathfrak{z}}/\overline{y}_{B\mathfrak{z}}) \tag{3}$$

Calculate the amount of information on each parameter change interval $I_{\mathfrak{z}}$:

$$I_{\mathfrak{z}} = \frac{1}{2}Z_{\mathfrak{z}}\left(\overline{y}_{A\mathfrak{z}} - \overline{y}_{B\mathfrak{z}}\right) \tag{4}$$

Calculate the total amount of information

$$I = \sum I_{\mathfrak{z}} \tag{5}$$

Table 2 provides a comprehensive dataset of 21 hydraulically fractured wells in carbonate reservoirs, listing 12 key parameters measured before and after fracturing treatments. The data is used to analyze the impact of these parameters on post-fracturing oil production (*Qo_post-frac*). Table 2 serves as the primary dataset for statistical analysis, linking fracturing design (e.g., *Vf*, *mp*) and reservoir properties (e.g., *Pb*, *Ip*) to post-frac performance. The variability in outcomes underscores the need for data-driven optimization in carbonate reservoirs.

Table 3 demonstrates the step-by-step application of Information Amount Theory (IAT) to evaluate the influence of fracturing fluid volume (*Vf*) on post-fracturing oil production. This quantitative method measures how much information each parameter interval contributes to predicting production outcomes. The table complements Fig. 1, which graphically represents the IA values across all parameters, showing *Vf*'s dominant position in the ranking.

Table 4 presents the ranking of 12 hydraulic fracturing parameters based on their **Information Amount (IA)** values, quantifying their relative influence on post-fracturing oil production (*Qo_post-frac*) in carbonate reservoirs. Range from **60.95** (highest, *Vf*) to **3.4** (lowest, *h*).

Table 3: Determining the information content of the "Main Frac fluid volume" attribute

No	Vf	Shooting frequency		Probability %		Average probability %		Average probability ratio	Diagnosis co-efficient	Information amount
		A	*B*	y_A	y_B	$\bar{y}_A$	$\bar{y}_B$	$(\bar{y}_A)/(\bar{y}_A)$	*Dc*	*IA*
1	[0–0.2]	0	1	0.0	8.3	5	13.3	0.38	−4.26	17.75
2	[0.2–0.4]	1	4	10.0	33.3	14	22.5	0.62	−2.06	8.76
3	[0.4–0.6]	3	4	30.0	33.3	24	23.3	1.03	0.12	4.05
4	[0.6–0.8]	4	1	40.0	8.3	27	15.0	1.80	2.55	15.32
5	[0.8–1]	2	1	20.0	8.3	19	8.3	2.28	3.58	19.09
Sum	–	10	11	100	100	89	82.5	6.11	−0.07	60.95

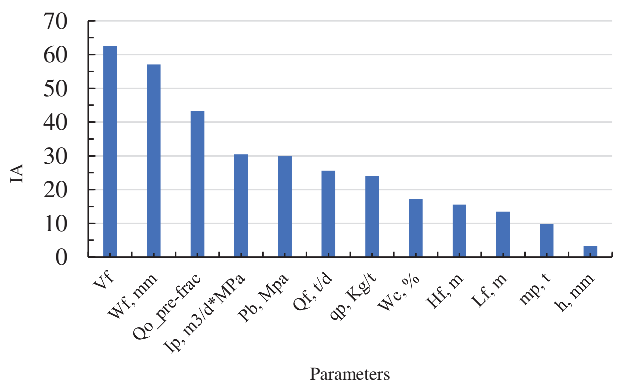

Figure 1: Information amount comparison of different parameters

Table 4: Parameters ranking of information amount theory

Parameters	Vf	Wf, mm	Qo_pre-frac,	Ip, m^3/d*MPa	Pb, Mpa	Qf, t/d	qp, Kg/t	Wc, %	Hf, m	Lf, m	mp, t	h, mm
IA	60.95	57.0	43.3	30.4	29.8	25.6	23.9	17.3	15.6	13.5	9.7	3.4
Rank	1	2	3	4	5	6	7	8	9	10	11	12

Fig. 2 the information content (and thus predictive influence on post-frac production) is not uniform across different ranges of fluid volume. The chart highlights that very low, high, and very high fluid volumes carry the most significant information for forecasting hydraulic fracturing outcomes in your carbonate reservoir study.

4 Gray Relational Analysis

Gray relational analysis (GRA) is a technique used to analyze and measure the strength of relationships between variables in systems with uncertainty. It has been applied in reservoir management [15], petroleum safety analysis [16], power equipment forecasting [17], and system modeling [18–21]. In this study, GRA helps identify which parameters have the strongest influence on post-fracturing production performance, GRA helps in identifying which variables have a significant impact on the outcome and understanding their interrelationships.

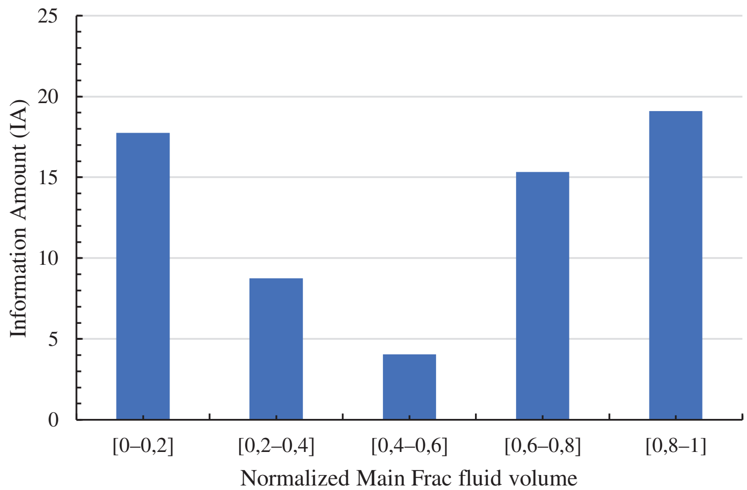

Figure 2: *C*hart effectively communicates the key insight from Table 3

Gray relational analysis (GRA) is a technique used to analyze and measure the strength of relationships between variables in systems with uncertainty. GRA helps in identifying which variables have a significant impact on the outcome and understanding their interrelationships.

The gray correlation analysis is described by Wu, Kuo et al. as a multivariate statistical analysis method that resolves the essential relationships between different system variables, identifies the key factors most influencing and gets the main statistical signification. It is a quantitative analysis of system evolution and comparison methods. The relative change of variables in the system development process is represented by the change trend, size and speed of the degree of gray correlation. If the relative changes of two variables or systems show a higher degree of gray correlation; conversely, they show a lower degree of gray correlation [22,23]. Wang used the normalization method for data analysis to pre-process index data in various dimensionless situations. The mean value method, initial value method, interval value method and maximum range method are the most frequently used in the study [24].

The production rate after hydraulic fracturing is chosen as the reference column X0, while the other parameters are used as comparison columns X1. The values of each parameter are represented n, which corresponds to the number of the wells, and m, corresponds to the number of the comparison columns (the parameter numbers). Typically, the data scaled to a range 0. . .1.

$$X_0 = \{X_0(k)|k = 1, 2, \ldots, n)\} \tag{6}$$

$$X_i = \{Xi_i(k)|k = 1, 2, \ldots, n)\}(i = 1, 2, \ldots, m) \tag{7}$$

The correlative coefficient between $X_i(k)$ and $X_0(k)$ is calculated as follows:

$$\xi_i(k) = \frac{\text{Min Min}\Delta_i\ (\text{k}) + \rho\text{Max Max}\Delta_i\ (\text{k})}{\Delta_i\ (\text{k}) + \rho\text{Max Max}\Delta_i\ (\text{k})} \tag{8}$$

The correlative coefficient: measures the degree of similarity the reference sequence and the comparison sequence. Higher coefficient indicate stronger relationships.

Where ρ is distinguishing coefficient! The smaller ρ is, the stronger the distinguishing ability is, Generally, the range of ρ is (0, 1), Mostly, $\rho = 0.5$. Different situation have different ρ value, at the moment k, absolute difference between X_i and X_0 is calculated as follows:

$$\Delta_i(k) = |X_0(k) - X_1(k)| \tag{9}$$

where $\Delta_i(k)$ is the absolute difference the reference and comparison sequences, $\Delta_{\min}(k)$ and $\Delta_{\max}(k)$ are the minimum and maximum differences.

$$\xi_i = \{\xi_i(k)|k = 1, 2, \ldots, n)\} \tag{10}$$

Gray relational grade (GRG): aggregates the GRGs for multiple factors to provide an overall assessment of their relational strength. The steps of the grey relational analysis are as follows:

- Determine the analytical indicator system according to the purpose of analysis and collect the analytical data.
- Generate the referential series.
- Normalize the data.

Gray relational coefficient comparison of different parameters is shown in Fig. 3. Figs. 1 and 2 in the paper highlight correlations (e.g., *Vf* vs. *Qo_post-frac*). Table 4 complements Figs. 1 and 2, where *Vf* and *Wf* also exhibit high IA and gray relational coefficients, reinforcing their significance. *Vf* and *Wf* (fluid volume and fracture width) dominate, suggesting they are primary drivers of production enhancement. Comparable analytical applications of GRA across engineering systems have been documented.

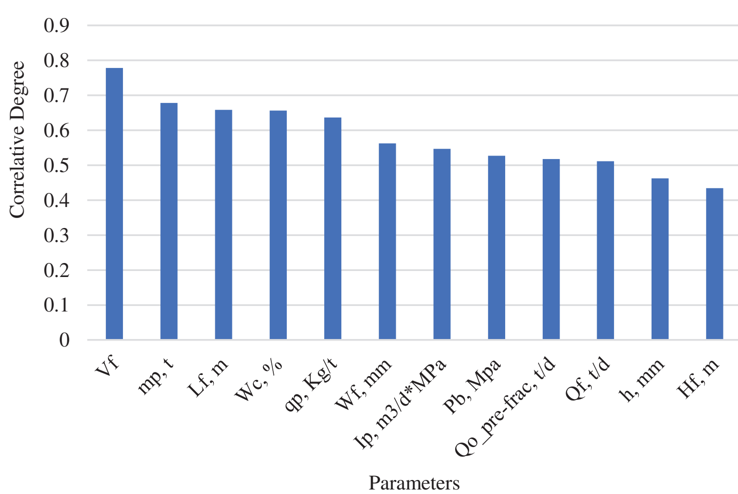

Figure 3: Gray relational coefficient comparison

The dominance of Vf and Wf explained through rock-mechanics principles: Vf directly affects fracture propagation pressure and simulated reservoir volume, while Wf governs post-closure conductivity and hydrocarbon mobility. In carbonates reservoirs, where the fracture systems interact with natural discontinuities, fracture complexity and aperture are more critical than planar length, explaining the low sensitivity of Lf observed in this study.

5 Operations Implications

Statistical methods offer powerful tools for understanding and predicting hydraulic fracturing behavior in complex carbonate reservoirs, though integration with geological and engineering knowledge remains essential for practical application. This study explores advantages methodologies by integrating Information amount theory and Gray relational analysis. The approach demonstrated that the IAT effectively identifies Vf and Wf as key factors affecting the production rate after hydraulic fracturing operation and GRA reveals important relationships between Qo_pre frac and Vf and mp. The results suggest that these methodologies, when combined, provide Vf for improving the accuracy of the production rate after hydraulic fracturing forecasts. These methods involve Vf in the rock create fractures, allowing hydrocarbons to flow more freely.

During hydraulic fracturing, multiple cracks and microcracks are created, forming a complex network of fractures in the formation. Therefore, the use of statistical methods aids in making a quick assessment of the results, but does not replace the need for more accurate and complex [25].

6 Conclusion

This study demonstrates that integrating Information amount theory and Gray relational analysis provides a robust, data-driven approach to evaluate and rank the factors influencing hydraulic fracturing outcomes in carbonates reservoirs. Vf and Wf were identified as the key parameters governing production enhancement, while Lf ranked lower due to the importance of fracture network complexity over linear geometry. Mechanistically, Vf controls the stimulated reservoir volume, and Wf ensures conductive flow channels. The proposed workflows offer interpretable, computationally lightweight guidance for field engineers, complementing traditional numerical simulations. Limitations include the relatively small dataset size and the assumption of linearity. Future work should integrate nonlinear models and field validation to enhance predictive capability.

Acknowledgement: The authors would like to thank Perm National Research Polytechnic the Government of Perm Krai for their continued support. The authors also acknowledge valuable comments from the reviews that improved the manuscript.

Funding Statement: The reported study was partially supported by the Government of Perm Krai, research project No. Сэд-26-08-08-32 from 25 January 2024.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, V. V. Poplygin and A. Dieng; methodology, V. V. Poplygin; software, A. Dieng; validation, Min Wang, Xian Shi and A. Dieng; formal analysis, A. Dieng; investigation, Min Wang; resources, A. Dieng; data curation, A. Dieng; writing—original draft preparation, V. V. Poplygin; writing—review and editing, Min Wang; visualization, A. Dieng; supervision, Xian Shi; project administration, V. V. Poplygin; funding acquisition, V. V. Poplygin. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the Corresponding Author, V. V. Poplygin, upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

Nomenclature

mp	Proppant Total
qp	Specific polymer consumption for placement of 1 t of proppant

Vf	Main Frac fluid volume
Qf	Pre-frac fluid production rate
Wc	Pre-frac water cut
Qo_pre-frac	Pre-frac oil production rate
Lf	Fracture length
Wf	Fracture width
Hf	Propant Height
h	Thickness
Pb	Bubble pressure
Ip	Pre-frac productivity index

References

1. Poplygin V, Qi C, Guzev M, Kozhevnikov E, Kunitskikh A, Riabokon E, et al. Assessment of the elastic-wave well treatment in oil-bearing clastic and carbonate reservoirs. Fluid Dyn Mater Process. 2023;19(6):1495–505. doi:10.32604/fdmp.2023.022335.
2. Poplygin VV, Wiercigroch M. Research of efficiency of complex non-stationary impact on layer with high-quality oil. Bull Tomsk Polytech Univ Geo Assets Eng. 2020;331(1):7–12. doi:10.18799/24131830/2020/1/2442.
3. Poplygin VV, Dieng A, Shi X. Forecasting hydraulic fracturing results using information amount theory. Perm J Petrol Min Eng. 2024;24(2):93–100.
4. Dieng A, Khiznyk GP, Poplygin VV. Prediction of the efficiency of hydraulic fracturing based on reservoir parameters. Int J Eng. 2023;36(12):2169–74. doi:10.5829/ije.2023.36.12c.05.
5. Xu C, Huang K, Ye J, Hu K. Directional microwave heating based on time reversal quantified by information theory. Int J Heat Mass Transf. 2023;216:124621. doi:10.1016/j.ijheatmasstransfer.2023.124621.
6. Ping X, Yang F, Zhang H, Zhang J, Xing C, Yan Y, et al. Information theory-based dynamic feature capture and global multi-objective optimization approach for organic Rankine cycle (ORC) considering road environment. Appl Energy. 2023;348:121569. doi:10.1016/j.apenergy.2023.121569.
7. Rapeti P, Pasam VK, Rao Gurram KM, Revuru RS. Performance evaluation of vegetable oil based nano cutting fluids in machining using grey relational analysis–a step towards sustainable manufacturing. J Clean Prod. 2018;172(1):2862–75. doi:10.1016/j.jclepro.2017.11.127.
8. Wang X, Wang Y, Tseng YY, Gao Y, Li K, Wang MH, et al. Integration of the grey relational analysis with machine learning for sucrose anaerobic hydrogen production prediction. Int J Hydrog Energy. 2024;68(45):388–97. doi:10.1016/j.ijhydene.2024.04.242.
9. Dieng A, Poplygin VV. Study on application of arps decline curves for gas production forecasting in Senegal. Int J Eng. 2023;36(12):2207–13. doi:10.5829/ije.2023.36.12c.10.
10. Li L, Zhou F, Zhou Y, Cai Z, Wang B, Zhao Y, et al. The prediction and optimization of Hydraulic fracturing by integrating the numerical simulation and the machine learning methods. Energy Rep. 2022;8:15338–49. doi:10.1016/j.egyr.2022.11.108.
11. Belytschko T, Lu YY, Gu L. Crack propagation by element-free Galerkin methods. Eng Fract Mech. 1995;51(2):295–315. doi:10.1016/0013-7944(94)00153-9.
12. Guo D, Kang Y, Wang Z, Zhao Y, Li S. Optimization of fracturing parameters for tight oil production based on genetic algorithm. Petroleum. 2022;8(2):252–63. doi:10.1016/j.petlm.2021.11.006.
13. Fu X, Sun H, Guo Q, Pan Z, Xiong W, Wang L. Uncertainty analysis of an integrated energy system based on information theory. Energy. 2017;122(1):649–62. doi:10.1016/j.energy.2017.01.111.
14. Liang T, Chang Y, Guo X, Liu B, Wu J. Influence factors of single well's productivity in the Bakken tight oil reservoir, Williston Basin. Petrol Explor Dev. 2013;40(3):383–8. doi:10.1016/s1876-3804(13)60047-6.
15. Zhang J, Wang M, Gong P, Yang Z, Liu X. Study on water injection formula by grey correlation method for offshore water flooding reservoir. J Geosci Environ Prot. 2018;6(8):1–11. doi:10.4236/gep.2018.68001.
16. Chu Z, You Z, Wang F, Chen X, Zhang B. A targeted risk prediction method based on statistical analysis of accidents in petroleum geophysical exploration. J Petrol Sci Eng. 2020;192(362):107314. doi:10.1016/j.petrol.2020.107314.

17. Qin J, Zhou C, Lin Y, Bai D, Zheng W. Based on the combination prediction method for the characteristic parameters prediction of power transmission and transformation equipment. Energy Rep. 2022;8(5):589–95. doi:10.1016/j.egyr.2021.11.125.
18. Deng JL. Gray control system. J Huazhong Inst Technol. 1982;3:9–17.
19. Deng JL. Introduction to grey system theory. Inn Mong Electr Power Technol. 1993;3:51–2.
20. Liu H, Sun GB. Application of the grey relation method in groundwater quality evaluation in Huai'an City. Jiangsu Environ Sci Technol. 2007;20:51–3.
21. Bezuglov A, Comert G. Short-term freeway traffic parameter prediction: application of grey system theory models. Expert Syst Appl. 2016;62(5):284–92. doi:10.1016/j.eswa.2016.06.032.
22. Wu HH. A comparative study of using grey relational analysis in multiple attribute decision making problems. Qual Eng. 2002;15(2):209–17. doi:10.1081/qen-120015853.
23. Kuo Y, Yang T, Huang GW. The use of grey relational analysis in solving multiple attribute decision-making problems. Comput Ind Eng. 2008;55(1):80–93. doi:10.1016/j.cie.2007.12.002.
24. Wang X. Application of grey relation analysis theory to choose high reliability of the network node. J Phys Conf Ser. 2019;1237(3):032056. doi:10.1088/1742-6596/1237/3/032056.
25. Yi D, Yang Z, Yi L, Liu J, Yang C, Gou L, et al. Phase-field model of hydraulic fracturing in thermoelastic-plastic media. Int J Mech Sci. 2024;283(1):109750. doi:10.1016/j.ijmecsci.2024.109750.