

基于主成分分析法的地质工程参数定量表征研究

唐毅

成都理工大学能源学院, 四川 成都

收稿日期: 2025年4月29日; 录用日期: 2025年6月11日; 发布日期: 2025年6月23日

摘要

本文针对地质工程中压裂井的复杂性, 提出了一种基于主成分分析法(PCA)的地质工程参数定量表征方法。通过对地质工程参数的相关性分析, 筛选出与无阻流量具有较强相关性的参数, 并利用主成分分析法对其进行定量分析, 实现了对压裂井压前和压后产能的详细评估。该方法能够有效降低数据维度, 保留关键信息, 并为地质工程一体化压后评估体系提供科学依据。

关键词

主成分分析法, 地质工程参数, 定量表征, 压裂井评估, 无阻流量

Research on the Quantitative Characterization of Geological Engineering Parameters Based on the Principal Component Analysis Method

Yi Tang

College of Energy, Chengdu University of Technology, Chengdu Sichuan

Received: Apr. 29th, 2025; accepted: Jun. 11th, 2025; published: Jun. 23rd, 2025

Abstract

Aiming at the complexity of fractured wells in geological engineering, this paper proposes a quantitative characterization method of geological engineering parameters based on the Principal Component Analysis (PCA). Through the correlation analysis of geological engineering parameters, the

parameters with a strong correlation with the open flow capacity are screened out, and the PCA is used for their quantitative analysis, realizing a detailed evaluation of the pre-fracturing and post-fracturing productivity of the fractured wells. This method can effectively reduce the data dimension, retain key information, and provide a scientific basis for the integrated post-fracturing evaluation system of geological engineering.

Keywords

Principal Component Analysis, Geological Engineering Parameters, Quantitative Characterization, Evaluation of Fractured Wells, Open Flow Capacity

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在地质工程中,压裂技术是一种重要的增产措施,广泛应用于油气田开发中。压裂过程通过在地层中形成裂缝,改善流体的流动通道,从而提高油气井的产能。然而,影响压后产能的因素众多且相互关系复杂,包括地质参数(如裂缝密度、孔隙度、渗透率等)和工程参数(如米液量、米砂量、分层工艺等)。传统的定性分析方法难以全面、准确地评估这些因素对压后产能的影响,因此需要一种定量的分析方法来提高评估的科学性和准确性。

刘子龙[1]针对博孜1区块开展了全面且深入的压后评估工作。先是引入滤失系数 ζ ,借助修正G函数和双对数曲线,对区内7口井进行分析,初步确定人工裂缝形态和闭合压力。陈满[2]等针对四川长宁地区页岩气水平井压后评估开展工作。选取H36平台9口井,用改进的返排优化模型评估其中3口控压返排井,通过等效压降模型确定闷井时间、砂拱模型确定油嘴制度,对比原方案和现方案明确返排潜力空间。廖凯等[3]针对页岩油井压后评估开展了多项工作。先是构建基于返排动态分析的评估方法,将有效裂缝分类并建立数学模型与迭代求解法;随后借助商业油藏模拟软件进行数值模拟,验证方法的适用性;最后以吉木萨尔页岩油藏四口井为例,运用该方法剖析返排数据,反演关键参数以评估压后效果。

本文旨在通过主成分分析法(PCA)对地质工程参数进行定量表征[4],建立一套地质工程一体化压后评估体系。通过该体系,可以对压裂井的压前和压后产能进行详细评估,为压裂井的优化设计和产能预测提供科学依据。

主成分分析法(PCA)是一种多元统计分析方法,通过线性组合将多个变量转换为少数几个不相关的主成分,从而降低数据维度并保留关键信息。本文将通过以下步骤实现研究目标:

- 1) 对地质工程参数进行相关性分析,筛选出与无阻流量具有较强相关性的参数。
- 2) 应用主成分分析法对筛选后的参数进行定量分析,计算各主成分的权重。
- 3) 基于主成分得分,建立地质工程一体化压后评估体系,并对压前和压后产能进行评估。

2. 地质工程参数定性论证结果

通过对地质工程参数的相关性分析,筛选出14项与无阻流量具有较强相关性的参数,如表1所示。这些参数包括裂缝密度(缝周)、孔隙度(缝周)、渗透率(缝周)、饱和度(缝周)、米液量、米砂量、分层工艺、G函数特征值、K值(缝周)、多裂缝特征和人工裂缝参数等。这些参数涵盖了地质条件和施工工艺两个方面,能够全面反映压裂井的产能变化。

表 1 单因素定性分析结果详细列出了各参数与无阻流量的相关性。其中，裂缝密度、孔隙度、渗透率等参数与无阻流量具有较强的相关性，而地层压力、闭合压力等参数与无阻流量具有一般相关性。

Table 1. Results of univariate qualitative analysis

表 1. 单因素定性分析结果

因素	相关性	因素	相关性
G 函数特征值	较强	地层压力(缝周)	一般
裂缝密度(缝周)	较强	闭合压力	一般
K 值	较强	压降速率	一般
分层工艺	较强	饱和度(缝周)	一般
米液量	较强	孔隙度(缝周)	一般
米砂量	较强	渗透率(缝周)	一般
多裂缝特征	较强		
人工裂缝参数	较强		

3. 地质工程评价模型建立

3.1. 数据标准化

在主成分分析法的实际应用中，用协方差矩阵求解主成分是关键步骤，但由于协方差矩阵易受指标数量级和量纲的影响，所以对原始数据进行无量纲化处理至关重要。以往在无量纲化处理时，常采用数据标准化方法。然而，无量纲化中的除法本质上是集合上的相似变换，无论采用何种具体的无量纲化方式，都不可避免地会造成信息丢失。这是因为相似变换会改变数据结构，进而改变变量间的相关信息，导致变换前后的方差之和出现差异。原始数据蕴含两部分关键信息：一是各指标变异程度的信息，通过方差大小来反映各指标的差异情况；二是各指标间相互影响程度的相关信息，由相关系数矩阵体现。

本研究选取了孔隙度(缝周)、渗透率(缝周)、饱和度(缝周)、米液量、米砂量、裂缝密度(缝周)、地层压力(缝周)、压降速率、闭合压力、G 函数特征值、多裂缝特征、人工裂缝参数、K 值、分层工艺等 14 个参数，构建分析矩阵 X (见表 2)。

Table 2. Basic data table for correlation analysis

表 2. 相关性分析基础数据表

井号	孔隙度 (井周)	渗透率 (井周)	饱和度 (井周)	地层压力 (井周)	裂缝密度 (井周)	米液量	米砂量
	x1	x2	x3	x4	x5	x6	x7
BZ1-1	7.90	0.35	70.597	131.23	0.84	20.69613	1.192265
BZ3-K2	7.80	0.39	73.128	99.06	0.81	1.878061	0
BZ3	7.70	0.31	67.767	113	0.65	36.01214	3.873571
BZ101-2	6.95	0.34	69.814	134.945	0.64	10.65088	0.810063
BZ105	7.22	0.33	67.426	109.472	0.58	1.690073	0
BZ102-1	6.93	0.33	66.621	131.226	0.46	9.257234	0.560106
BZ102-4	7.09	0.30	66.608	134.077	0.58	12.4489	0.691463
BZ3-2X	7.15	0.29	66.229	109.472	0.658	12.25345	0.803448

如表 3 所示，采用极差标准化处理矩阵 X，其中 X10 的闭合压力因素栏，由于闭合压力与无阻流量呈现负相关，因此采用补集计算。

Table 3. Results of range standardization**表 3.** 极差标准化结果

井号	x1	x2	x3	x4	x5	x6	x7
BZ1-1	0.839450	0.011057	0.814443	0.867680	1.000000	0.553756	0.307795
BZ3-K2	0.816514	0.013514	1.000000	0.352960	0.952381	0.005477	0.000000
BZ3	0.793578	0.008600	0.606965	0.576000	0.698413	1.000000	1.000000
BZ101-2	0.621560	0.010442	0.757038	0.927120	0.682540	0.261080	0.209126
BZ105	0.683486	0.009828	0.581965	0.519552	0.587302	0.000000	0.000000
BZ102-1	0.616972	0.009828	0.522947	0.867616	0.396825	0.220475	0.144597
BZ102-4	0.653670	0.007985	0.521994	0.913232	0.587302	0.313467	0.178508
BZ3-2X	0.667431	0.007371	0.494208	0.519552	0.711111	0.307772	0.207418

3.2. KMO 和 Bartlett 检验

为确保获得可信度较高的数据集，需对样本进行分析可行性检验。KMO 检验和巴特利特(Bartlett)球形检验是统计学中常用的两种多元数据分析方法，其中 KMO 检验用于评估数据的样本合适性，Bartlett 球形检验则用于确认数据的统计独立性。

(1) KMO 检验

KMO 检验(Kaiser-Meyer-Olkin Test)是一种用于衡量变量之间相关性的方法。KMO 检验通常用于检测数据样本的合适性，以确定是否适合进行因子分析或其他多元数据分析。

KMO 计算公式为：

$$KMO = \frac{\sum \sum_{i \neq j} r_{ij}^2}{\sum \sum_{i \neq j} r_{ij}^2 + \sum \sum_{i \neq j} r_{ij,1,2,\dots,k}^2} \quad (1)$$

Kaiser 给出了常用的 KMO 度量标准：0.9 以上表示非常适合；0.8 表示适合；0.7 表示一般；0.6 表示不太适合；0.5 以下表示极不适合。KMO 统计量是取值在 0 和 1 之间。当所有变量间的简单相关系数平方和远远大于偏相关系数平方和时，KMO 值接近 1。KMO 值越接近于 1，意味着变量间的相关性越强，原有变量越适合作因子分析；当所有变量间的简单相关系数平方和接近 0 时，KMO 值接近 0。KMO 值越接近于 0，意味着变量间的相关性越弱，原有变量越不合作因子分析。

(2) Bartlett 球形检验

Bartlett 球形检验(Bartlett's test of sphericity)用于检验数据之间的独立性，一般在进行因子分析前都要进行此项检验，以判断变量是否适合用于做因子分析。该检验以变量的相关系数矩阵为出发点，其零假设是相关系数矩阵为单位阵，即相关系数矩阵对角线上的所有元素都是 1，所有非对角线上的元素都为零。Bartlett 球形检验的统计量根据相关系数矩阵的行列式得到，如果该值较大，且其对应的相伴概率值小于用户设定的显著性水平，那么应拒绝零假设，认为相关系数不可能是单位阵，即原始变量之间存在相关性，适合进行因子分析；反之则不合作因子分析。

Bartlett 球形检验步骤如下：

a. 建立原假设

H0：这 n 个样本所代表的各总体方差相同；

H1：这 n 个样本所代表的各总体方差不同。

b. 计算统计量

$$X^2 = \frac{\sum_{i=1}^m (n_i - 1) \ln s^2 - \sum_{i=1}^m (n_i - 1) \ln s_i^2}{C} \quad (2)$$

式中： s_i^2 为每组总体样本方差， $s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$ ， $(i=1, 2, \dots, m)$ ；

$$s^2 = \sum_{i=1}^m (n_i - 1) s_i^2 / \sum_{i=1}^m (n_i - 1) \quad (3)$$

$$C = 1 + \frac{1}{3(m-1)} \left[\sum_{i=1}^m \frac{1}{n_i - 1} - \frac{1}{\sum_{i=1}^m (n_i - 1)} \right] \quad (4)$$

在原假设成立的条件下，计算检验 s^2 分布服从卡方分布。

Table 4. Measurement criteria for KMO and Bartlett's sphericity test

表 4. KMO 与 Bartlett 球形检验度量标准

检测类别	值的范围	因子分析适合情况
KMO 值	大于 0.9	非常适合
	0.8~0.9	很适合
	0.7~0.8	适合
	0.6~0.7	勉强适合
	0.5~0.6	不太适合
	小于 0.5	不适合
Bartlett P 值	小于或者等于 0.01	适合

KMO 与 Bartlett 球形检验度量标准如表 4 所示。将极差标准化结果代入 KMO 与 Bartlett 球形检验中得到检验值，从图 1 可以看出，KMO 值为 $0.855 > 0.8$ ，Bartlett 球形检验显著性 < 0.001 ，根据表 4 度量标准，认为样本很适合做主成分分析。

KMO和巴特利特检验

KMO 取样适切性量数	0.855	
巴特利特球形度检测	近似卡方	185.474
	自由度	91
	显著性	<0.001

Figure 1. Results of KMO and Bartlett's sphericity test

图 1. KMO 与 Bartlett 球形检验结果

3.3. 特征向量计算

在计算主成分的时候，根据式 $u_{k1}^2 + u_{k2}^2 + \dots + u_{kp}^2 = 1$ ， $(k=1, 2, \dots, p)$ ，那么特征向量： $|u_k| = 1$ ，则特征值 λ_k 对应的特征向量为 $\pm u_k$ 。

在选择特征向量的正负的条件时，假使有较为理想的最优样本 a 与最差样本 b 为：

$$a = \left\{ \max_i Z_{ij} \mid i=1, \dots, n; j=1, \dots, p \right\} = \{Z_1^+, \dots, Z_p^+\} \quad (5)$$

$$b = \left\{ \min_i Z_{ij} \mid i=1, \dots, n; j=1, \dots, p \right\} = \{Z_1^-, \dots, Z_p^-\} \quad (6)$$

从上式中可以看出主成分 Y 对样本和的得分必须得满足 $Y_{ai} > Y_{bi}$ ，与

$$\sum_{i=1}^p u_{ij} Z_i^+ > \sum_{i=1}^p u_{ij} Z_i^-, i=1, 2, \dots, p \quad (7)$$

是一种等价的满足条件。

所以单位特征向量 u_k 满足如下的条件:

$$(1) u_{k1}^2 + u_{k2}^2 + \dots + u_{kp}^2 = 1$$

$$(2) \sum_{k=1}^p u_{k,i} u_{k,j} = 0$$

$$(3) \sum_{i=1}^p u_{i,j} Z_i^+ > \sum_{i=1}^p u_{i,j} Z_i^-$$

上面三个条件中的第三个条件是可以作为选择特征向量 u_k 的条件。如果采用极差标准化进行无量纲化时, 上面的最优和最劣的样本为: $a = \{1, 1, \dots, 1\}$, $b = \{0, 0, \dots, 0\}$ 。那么上面的第三个条件就会变为 $\sum_{k=1}^p u_{k,j} > 0$ 的形式, 只要是非负即可。也就是单位特征向量满足非负就可以。当 u_k 不满足大于零的条件时, 可以用负值来代替主成分的计算。

在实际应用中, 特征向量 u_k 的正负需要满足至多以下三个条件:

(1) 独立性。主成分所表示的意义各自不相关, 独立;

(2) 代表性。主成分各自所对应的系数值较大的那些样本所表示的意义也就越大, 即主成分的意义明确且有代表性;

(3) 综合性。所选主成分的意义应包括指标体系的所有指标且包含每一指标的次数尽可能都一样。

数据 x_1, x_2, \dots, x_p 是向量 μ 协方差矩阵为 Σ 的某一个 p 维的总体样本中的 n 个抽样, \bar{x} 为上述数据得到的样本均值向量, S 为样本的协方差矩阵, R 为样本相关矩阵。

假设特征值 - 特征向量对 $(\hat{\lambda}_1, \hat{e}_1), (\hat{\lambda}_2, \hat{e}_2), \dots, (\hat{\lambda}_n, \hat{e}_n)$ 的 $p \times p$ 的样本协方差阵为 S , 第 i 个主成分由下式得出:

$$\hat{y}_i = \hat{e}_i' x = \hat{e}_{i1}' x_1 + \hat{e}_{i2}' x_2 + \dots + \hat{e}_{ip}' x_p, i=1, 2, \dots, p \quad (8)$$

上式中 x 是 x_1, x_2, \dots, x_p 的任意一个观察值, 并且样品的方差:

$$\text{var}(\hat{y}_k) = \hat{\lambda}_k, k=1, 2, \dots, p \quad (9)$$

样品的协方差:

$$\text{cov}(\hat{y}_i, \hat{y}_k) = 0, i \neq k \quad (10)$$

样品的总方差:

$$\sum_{i=1}^p \text{var}(\hat{y}_i) = \hat{\lambda}_1 + \hat{\lambda}_2 + \dots + \hat{\lambda}_p \quad (11)$$

变量 x_i 在主成分 y_i 中的负荷量为:

$$r(\hat{y}_k, x_i) = \frac{\hat{e}_{ki} \sqrt{\hat{\lambda}_k}}{\sqrt{s_{ij}}}, i, k=1, 2, \dots, p \quad (12)$$

运用标准化的形式求得的主成分是:

$$z_i = D^{-\frac{1}{2}} (x_i - \bar{x}) = \left[\frac{x_{j1} - \bar{x}_1}{\sqrt{s_{11}}} \quad \frac{x_{j2} - \bar{x}_2}{\sqrt{s_{22}}} \quad \dots \quad \frac{x_{jp} - \bar{x}_p}{\sqrt{s_{pp}}} \right]^T \quad (13)$$

则把数据进行标准化之后的数据矩阵为:

$$Z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1p} \\ z_{21} & z_{22} & \cdots & z_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{np} \end{bmatrix} = \begin{bmatrix} \frac{x_{11} - \bar{x}_1}{\sqrt{s_{11}}} & \frac{x_{12} - \bar{x}_2}{\sqrt{s_{11}}} & \cdots & \frac{x_{1p} - \bar{x}_p}{\sqrt{s_{11}}} \\ \frac{x_{21} - \bar{x}_1}{\sqrt{s_{22}}} & \frac{x_{22} - \bar{x}_2}{\sqrt{s_{22}}} & \cdots & \frac{x_{2p} - \bar{x}_p}{\sqrt{s_{22}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{x_{p1} - \bar{x}_1}{\sqrt{s_{pn}}} & \frac{x_{p2} - \bar{x}_2}{\sqrt{s_{pn}}} & \cdots & \frac{x_{pn} - \bar{x}_p}{\sqrt{s_{pn}}} \end{bmatrix} \quad (14)$$

样本均值向量为:

$$\bar{z} = \frac{1}{n}(1'Z') = \frac{1}{n}Z'1' = \frac{1}{n} \left[\sum_{j=1}^n \frac{x_{j1} - \bar{x}_1}{\sqrt{s_{11}}} \quad \sum_{j=1}^n \frac{x_{j2} - \bar{x}_2}{\sqrt{s_{22}}} \quad \cdots \quad \sum_{j=1}^n \frac{x_{jp} - \bar{x}_p}{\sqrt{s_{pp}}} \right] = 0 \quad (15)$$

样品的协方差矩阵是:

$$S_z = \frac{1}{n-1}(Z - 1\bar{z}')(Z - 1\bar{z}') = \frac{1}{n-1}ZZ' = R \quad (16)$$

因此,第 i 个样本的主成分是:

$$\hat{y}_i = \hat{e}_1 z_1 + \hat{e}_2 z_2 + \cdots + \hat{e}'_p z_p \quad (17)$$

其中 $(\hat{\lambda}_i, \hat{e}_i)$ 是 R 的第 i 个特征值—特征向量对。

样品的方差:

$$\text{var}(\hat{y}_k) = \hat{\lambda}_k, k = 1, 2, \dots, p \quad (18)$$

样品的协方差:

$$\text{cov}(\hat{y}_i, \hat{y}_k) = 0, i \neq k \quad (19)$$

样品的总方差:

$$\sum_{i=1}^p \text{var}(\hat{y}_i) = \hat{\lambda}_1 + \hat{\lambda}_2 + \cdots + \hat{\lambda}_p \quad (20)$$

变量 x_i 在主成分 y_i 中的负荷量为:

$$r_i(\hat{v}_k \cdot x_i) = \hat{e}_{ki} \sqrt{\hat{\lambda}_{ki}}, i, k = 1, 2, \dots, p \quad (21)$$

第 K 个主成分的贡献率为:

$$\frac{\hat{\lambda}_i}{p}, i = 1, 2, \dots, p \quad (22)$$

也就是说样本进行标准化之后并没有改变协方差,那么也就说明主成分没有改变,主成分所包含的信息与标准化之前是相同的,因此可以通过标准化的过程进行主成分的计算及得分的计算。主成分可以通过累计贡献率来确定,阈值确定为 85%,主成分可以适当的做筛选工作,来获取最为适合的主成分。

采用显著性分析获得相关性矩阵,基于主成分分析法提取公因子方差,获得样本总方差解释表。从图 2 可以看出,当成分为 4 时,累计方差解释率大于 85%,表明样本合格,适合做主成分分析。通过公式 8 到公式 22 计算出样本特征值向量矩阵(图 3)。

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14
显著性(单尾)														
x1		.020	.003	.029	<.001	.085	.072	.007	.038	.005	.124	.065	.023	.011
x2			.028	.053	.142	.159	.230	.114	.232	.174	.378	.219	.233	.134
x3				.308	.005	.314	.184	.010	.417	.197	.111	.001	.001	.000
x4					.080	.308	.447	.044	.035	.000	.499	.157	.271	.235
x5						.144	.125	.013	.066	.061	.211	.026	.056	.006
x6							.000	.129	.410	.283	.240	.423	.440	.237
x7								.066	.393	.353	.132	.397	.181	.064
x8									.239	.007	.181	.145	.062	.031
x9										.262	.044	.416	.042	.440
x10											.043	.348	.058	.327
x11												.428	.001	.044
x12													.037	.000
x13														.002
x14														

Figure 2. Correlation calculation matrix
图 2. 相关性计算矩阵

	成分			
	1	2	3	4
x1	0.869	-0.308	0.085	0.081
x2	-0.507	0.253	-0.097	-0.123
x3	0.842	0.326	-0.139	0.203
x4	-0.282	0.883	-0.003	0.214
x5	0.790	-0.186	0.013	0.219
x6	0.282	-0.081	0.860	0.029
x7	0.360	0.097	0.864	0.051
x8	0.635	-0.257	-0.633	-0.127
x9	0.176	-0.641	0.002	0.592
x10	0.549	-0.538	-0.053	-0.532
x11	0.480	0.301	0.136	-0.690
x12	0.567	0.442	-0.225	0.477
x13	0.700	0.475	-0.079	-0.412
x14	0.807	0.432	-0.023	0.217

提取方法：主成分分析法。

a. 提取了4个成分。

Figure 3. Extraction of eigenvectors
图 3. 特征向量提取

3.4. 权重计算

根据获取的特征向量，为了使用方便，将特征向量进行旋转，采用迭代的方式计算在各主成分下的单井得分值。

设极差标准化矩阵单元值为 t_{ij} (i 为多因素 x_i , j 为井号)，主成分矩阵单元值为 H_{ij} (i 为多因素 x_i , j 为主成分)，单因素得分值为 T_{ij} (i 为多因素 x_i , j 为井号)，则：

$$T_{ij} = H_{ij} * t_{ij} \quad (23)$$

根据公式 23，计算各主成分迭代矩阵。

表 5 中展示的是旋转后的特征向量，其中列出了各个主成分对应的特征向量值。

如表 6 所示，第一次迭代矩阵揭示了在第一轮迭代之后，各个油井在不同主成分上的得分。

表 7 所示的是第二次迭代矩阵，其中列出了各井在完成第二次迭代后，针对不同主成分的得分情况。

Table 5. Rotated eigenvectors

表 5. 旋转后的特征向量

	x1	x2	x3	x4	x5	x6	x7
主成分 1	0.578	-0.335	0.900	0.218	0.641	0.034	0.172
主成分 2	0.679	-0.428	0.150	-0.907	0.506	0.135	0.031
主成分 3	0.006	0.095	0.202	0.180	-0.072	0.024	0.112
主成分 4	0.260	-0.204	0.046	-0.041	0.185	0.898	0.919

Table 6. The first iteration matrix

表 6. 第一次迭代矩阵

井号	x1	x2	x3	x4	x5	x6	x7
BZ1-1	0.485338	-0.003705	0.732976	0.189064	0.641274	0.019005	0.053047
BZ3-K2	0.472078	-0.004529	0.899972	0.076908	0.610737	0.000188	0.000000
BZ3	0.458817	-0.002882	0.546251	0.125508	0.447874	0.034320	0.172344
BZ101-2	0.359362	-0.003499	0.681313	0.202015	0.437695	0.008960	0.036042
BZ105	0.395166	-0.003294	0.523752	0.113208	0.376621	0.000000	0.000000
BZ102-1	0.356710	-0.003294	0.470638	0.189050	0.254474	0.007567	0.024920
BZ102-4	0.377927	-0.002676	0.469780	0.198989	0.376621	0.010758	0.030765
BZ3-2X	0.385884	-0.002470	0.444774	0.113208	0.456017	0.010563	0.035747
BZ301	0.430970	-0.002264	0.541765	0.112957	0.389854	0.012762	0.058543

Table 7. The second iteration matrix

表 7. 第二次迭代矩阵

井号	x1	x2	x3	x4	x5	x6	x7
BZ1-1	0.569957	-0.004727	0.122136	-0.787279	0.506218	0.074804	0.009491
BZ3-K2	0.554385	-0.005777	0.149962	-0.320254	0.482113	0.000740	0.000000
BZ3	0.538812	-0.003676	0.091022	-0.522627	0.353549	0.135086	0.030836
BZ101-2	0.422017	-0.004464	0.113527	-0.841211	0.345514	0.035268	0.006449
BZ105	0.464063	-0.004202	0.087273	-0.471409	0.297303	0.000000	0.000000
BZ102-1	0.418903	-0.004202	0.078422	-0.787221	0.200880	0.029783	0.004459
BZ102-4	0.443819	-0.003414	0.078279	-0.828610	0.297303	0.042345	0.005505
BZ3-2X	0.453163	-0.003151	0.074113	-0.471409	0.359978	0.041576	0.006396
BZ301	0.506109	-0.002889	0.090274	-0.470364	0.307749	0.050231	0.010475
BZ302	0.451605	-0.002363	0.066329	-0.402989	0.265162	0.111423	0.000000
BZ24	0.358170	-0.003151	0.049870	-0.722967	0.184810	0.042940	0.005550

设迭代次数为 n ($n = 1, 2, 3, 4$)，单井在第 n 次的得分为

$$S_{ni} = \sum_1^{14} T_{ij} \quad (24)$$

由公式 24 计算得到每次迭代的得分值，得到如表 8 所示的各主成分得分。

Table 8. Scores of each principal component
表 8. 各主成分得分

	主成分 1 得分	主成分 2 得分	主成分 3 得分	主成分 4 得分
BZ1-1	4.604452128	1.375723238	0.620189642	1.133646949
BZ3-K2	4.74312753	2.511824102	1.290208158	0.081216733
BZ3	3.517493712	1.861996965	1.941014334	2.390071578
BZ101-2	3.30877251	0.749528973	0.752685887	0.790931963
BZ105	3.361639485	2.127549606	2.095741634	0.163673874
BZ102-1	3.184711925	0.69363111	1.070362142	0.537547978
BZ102-4	3.407928133	0.605595328	1.087615659	0.715402967
BZ3-2X	3.412041012	1.853508132	1.996029885	0.838779145
BZ301	2.589879352	1.855266992	0.557327844	0.940514947
BZ302	2.174079957	2.052068532	1.165044677	0.893655463
BZ24	1.839236375	0.565025002	0.537850249	0.606733997

基于主成分得分矩阵和提取的总方差矩阵，得到如表 9 所示的基础解释力矩阵。

Table 9. Basic explanation table
表 9. 基础解释表

	x1	x2	x3	x4	x5	x6	x7
解释力	12.687	5.06	4.88	4.01	8.62	7.54	7.62
权重	0.12687	0.05713	0.05510	0.04527	0.09732	0.08513	0.08603
	x8	x9	x10	x11	x12	x13	x14
解释力	15.58	9.89	6.77	6.48	4.01	4	4.11
权重	0.17591	0.11166	0.07644	0.07316	0.04527	0.04516	0.04640

根据基础解释表计算综合值得分，与无阻流量对比，具有极强的相关性。

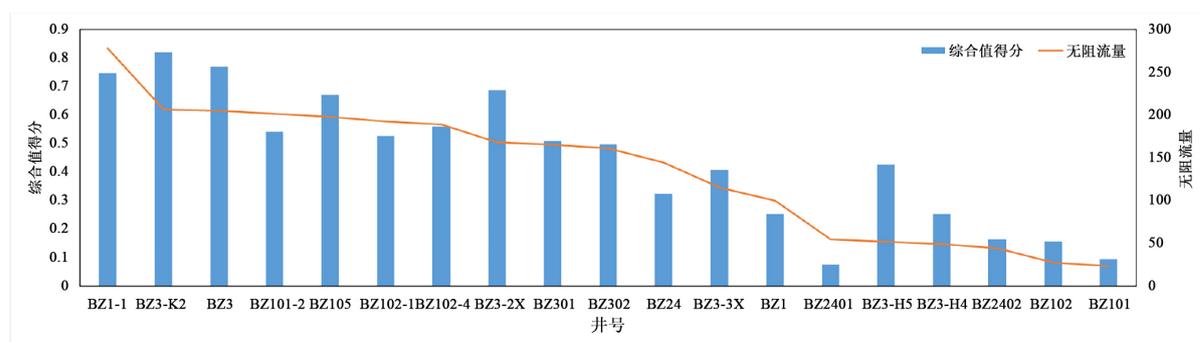


Figure 4. Relationship between comprehensive score of a single well and unrestricted flow rate
图 4. 单井综合得分与无阻流量关系

为了对压裂前的产能值进行一个大概的预估，将以上 14 种因素进行拆分，分为压前参数和压后参数，分别计算压前参数和压后参数各自的占比。

Table 10. Weight proportion of multiple factors in Bozi block
表 10. 博孜区块多因素权重占比

		评估参数	权重
压前参数	地质参数	裂缝密度(缝周)	25%
		孔隙度(缝周)	10%
		渗透率(缝周)	10%
	施工参数	饱和度(缝周)	8%
		分层工艺	17%
		米液量	15%
压后参数		米砂量	15%
		G函数特征值	30%
		K值(缝周)	20%
		多裂缝特征	13%
		人工缝网参数	13%
		地层压力(缝周)	8%
		压降速率	8%
		闭合压力	8%

4. 博孜区块单井得分评估

根据表 10, 以 100 分为满分, 对各类地质工程参数进行权重赋值。赋值结果见图 4 所示, 为了计算简便, 只保留了一位小数, 对各因素进行区间划分, 获得不同因素下不同赋分间隔。

基于单因素赋分细则, 对博孜区块各影响因素中的单井压前和压后参数进行打分, 打分结果见表 11 与表 12。

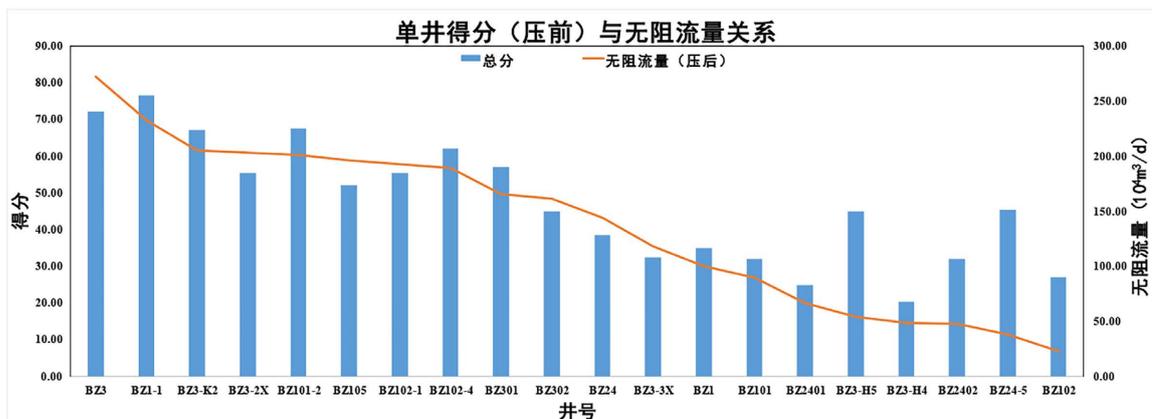
Table 11. Pre-fracturing score table of Bozi block
表 11. 博孜区块压前得分表

井号	裂缝密度(缝周) 得分	孔隙度(缝周) 得分	渗透率(缝周) 得分	饱和度(缝周) 得分	改造方式 得分	米液量 得分	米加砂量 得分	总分	无阻流量(压前)	无阻流量(压后)
BZ3	13.00	10.00	10.00	7.50	4	15.00	15.00	72.00	0.00	271.90
BZ1-1	17.00	10.00	10.00	7.50	17	10.00	5.00	76.50	0.00	232.02
BZ3-K2	17.00	10.00	10.00	10.00	15	2.50	2.50	67.00	15.74	204.98
BZ3-2X	13.00	7.50	7.50	7.50	10	5.00	5.00	55.50	0.00	202.95
BZ101-2	13.00	7.50	7.50	7.50	7	12.50	10.00	67.50	0.00	201.39
BZ105	13.00	10.00	10.00	7.50	4	5.00	2.50	52.00	51.02	196.30
BZ102-1	9.00	7.50	7.50	7.50	14	5.00	2.50	55.50	31.80	192.57
BZ102-4	13.00	7.50	7.50	7.50	14	7.50	5.00	62.00	0.00	189.14
BZ301	13.00	10.00	10.00	7.50	4	7.50	7.50	57.00	12.31	165.39
BZ302	9.00	7.50	7.50	5.00	1	12.50	2.50	45.00	31.80	161.25
BZ24	9.00	5.00	5.00	5.00	2	5.00	5.00	38.50	14.20	144.10
BZ3-3X	9.00	5.00	5.00	5.00	1	2.50	2.50	32.50	16.42	118.60
BZ1	9.00	7.50	7.50	5.00	1	2.50	2.50	35.00	31.04	100.07
BZ101	5.00	2.50	2.50	2.50	4	7.50	5.00	32.00	26.93	89.64
BZ2401	1.00	2.50	2.50	5.00	3	5.00	2.50	25.00	37.41	66.81
BZ3-H5	17.00	2.50	2.50	2.50	3	10.00	7.50	45.00	0.00	54.54
BZ3-H4	5.00	2.50	2.50	2.50	3	2.50	2.50	20.50	33.88	48.90
BZ2402	9.00	2.50	2.50	5.00	17	5.00	2.50	32.00	0.00	48.00
BZ24-5	1.00	7.50	7.50	2.50	3	7.50	5.00	45.50	19.94	38.69
BZ102	9.00	2.50	2.50	2.50	4	5.00	2.50	27.00	0.00	23.11

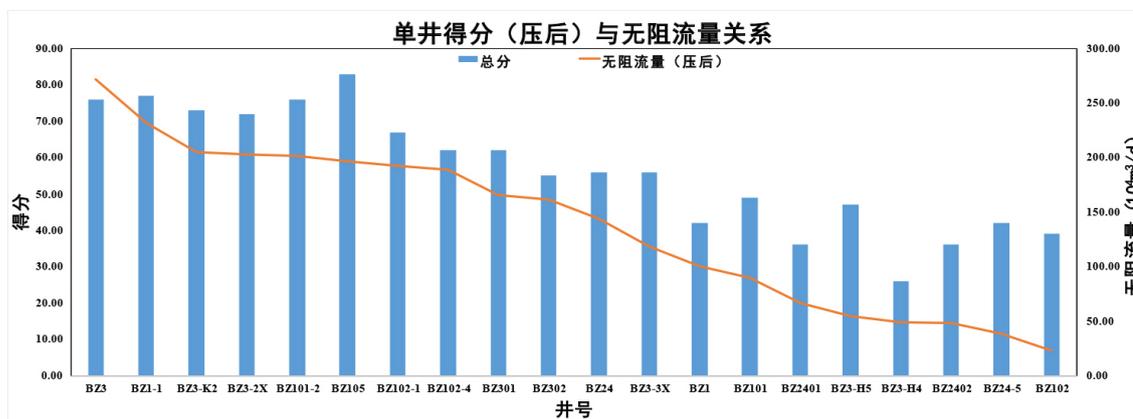
Table 12. Post-fracturing score table of Bozi block
表 12. 博孜区块压后得分表

井号	G 函数 得分	K 值 得分	地层压 力(井周) 得分	压降速 率 得分	闭合压 力(井周) 得分	多裂缝 特征 得分	人工缝 网参数 得分	总分	无阻流 量(压前)	无阻流 量(压后)
BZ3	20.00	16.00	6.00	4.00	4.00	13.00	13.00	76.00	0.00	271.90
BZ1-1	30.00	4.00	8.00	6.00	8.00	8.00	13.00	77.00	0.00	232.02
BZ3-K2	25.00	4.00	6.00	8.00	4.00	13.00	13.00	73.00	15.74	204.98
BZ3-2X	15.00	20.00	6.00	6.00	4.00	13.00	8.00	72.00	0.00	202.95
BZ101-2	30.00	8.00	8.00	4.00	8.00	8.00	10.00	76.00	0.00	201.39
BZ105	20.00	20.00	6.00	8.00	6.00	13.00	10.00	83.00	51.02	196.30
BZ102-1	20.00	8.00	8.00	4.00	6.00	13.00	8.00	67.00	31.80	192.57
BZ102-4	15.00	4.00	8.00	4.00	8.00	13.00	10.00	62.00	0.00	189.14
BZ301	30.00	4.00	6.00	4.00	4.00	8.00	6.00	62.00	12.31	165.39
BZ302	15.00	12.00	6.00	6.00	4.00	8.00	4.00	55.00	31.80	161.25
BZ24	20.00	4.00	8.00	4.00	8.00	8.00	4.00	56.00	14.20	144.10
BZ3-3X	15.00	8.00	6.00	6.00	4.00	13.00	4.00	56.00	16.42	118.60
BZ1	15.00	4.00	6.00	6.00	6.00	3.00	2.00	42.00	31.04	100.07
BZ101	20.00	4.00	8.00	4.00	6.00	3.00	4.00	49.00	26.93	89.64
BZ2401	5.00	4.00	8.00	4.00	8.00	3.00	4.00	36.00	37.41	66.81
BZ3-H5	20.00	4.00	4.00	6.00	6.00	3.00	4.00	47.00	0.00	54.54
BZ3-H4	5.00	4.00	4.00	4.00	4.00	3.00	2.00	26.00	33.88	48.90
BZ2402	5.00	4.00	8.00	4.00	8.00	3.00	4.00	36.00	0.00	48.00
BZ24-5	5.00	8.00	8.00	4.00	8.00	3.00	6.00	42.00	19.94	38.69
BZ102	5.00	4.00	8.00	4.00	8.00	8.00	2.00	39.00	0.00	23.11

图 5 所示根据单井得分表，与无阻流量对比，无论是压前还是压后参数均获得了不错的解释效果。压前参数评估压裂前预估产能，压后参数对压裂后产能进行分析，同时兼顾了预测与评估，形成了一套地质工程一体化压后评估体系。



(a) 压前得分与无阻流量相关性



(b) 压后得分与无阻流量相关性

Figure 5. Correlation between pre-fracturing and post-fracturing scores and unrestricted flow rate in Bozi block
图 5. 博孜区块压前、压后得分与无阻流量相关性

5. 结论

本文通过主成分分析法对地质工程参数进行了定量表征，建立了一套地质工程一体化压后评估体系。研究表明，主成分分析法能够有效降低数据维度，保留关键信息，并为压裂井的优化设计和产能预测提供科学依据。通过该体系，可以对压裂井的压前和压后产能进行详细评估，为地质工程的决策提供有力支持。

参考文献

- [1] 刘子龙. 博孜1区块地质工程一体化综合压后评估[D]: [硕士学位论文]. 成都: 成都理工大学, 2021.
- [2] 廖凯, 陈绩, 谢勃勃, 等. 基于返排动态分析的页岩油井压后评估方法[J]. 科学技术与工程, 2023, 23(24): 10273-10280.
- [3] 陈满, 常程, 岳文瀚, 等. 四川长宁地区页岩气水平井压后返排潜力评估[J]. 石油地质与工程, 2023, 37(6): 63-67.
- [4] Chen, L., Liu, X.Y., Zhou, H.L., *et al.* (2025) Carbonate Reservoirs Characterization Based on Frequency Bayesian Principal Component Analysis. *Geoenergy Science and Engineering*, **246**, Article ID: 213615. <https://doi.org/10.1016/j.geoen.2024.213615>