Efficient evaluation model of beam pumping unit based on principal component regression analysis

Zhewei Ye\textsuperscript{1,2}, Zhiyang Liu\textsuperscript{1}, Cheng Cheng\textsuperscript{1}, Lang Tan\textsuperscript{1} and Kang Feng\textsuperscript{1}

\textsuperscript{1}School of Mechatronic Engineering, Southwest Petroleum University, Chengdu, P.R. China
\textsuperscript{2}School of Aeronautics & Astronautics, Sichuan University, Chengdu, P.R. China

Abstract
The beam pumping unit is widely used in oil fields as an oil extraction machine. However, due to its structural characteristics and the changing conditions of underground reservoirs, there are many factors affecting the efficiency of the beam pumping unit. How to quickly and accurately reflect the inefficient factors, and whether the efficiency of the pumping unit is reasonable, it is an urgent problem to be solved. In this article, we propose a pumping unit efficiency evaluation model consisting of the scoring model and the predictive model. The scoring model reflects the inefficient factors, and the predictive model judges whether the efficiency of the pumping unit is reasonable. Finally, the actual production data of the oilfield were used to verify the effectiveness of the method.

Keywords
Beam pumping unit efficiency, principal component analysis, multiple linear regression method, efficient evaluation model, scoring model, predictive model

Introduction
Oil and gas are one of the main energy resources that are difficultly replaced in the modern society. They are called black gold to economic and are the main materials for the petrochemical industry. With the development of economy, the target of energy-saving, emission-reduction, green and low-carbon production are our goal.\textsuperscript{1}
Oilfields are great customers for power consumption, and had the responsibility to design the beam pumping units in order to improve their performance. The total consumed power of the Daqing Oilfield is 9900 million kW h, where the artificial lift occupied one-third. The “big motor drive small rod pumping” always existed in the past, which was constantly denounced.\(^2\)

Electric motors are operated with cyclical loading on Nodding Donkey oil pump. This varying load impacts motor efficiency, energy consumption which is the largest part cost for the operation of oil-well pumping units. And high loss of energy is a common and urgent problem in all oil fields.\(^3\)

Studying efficiency of beam pumping unit, the people completed the following work.

In order to keep initiating and operating the prime motor smoothly and safely, its actual rated power is high much more than the needed. With the intention of improving the efficiency of the prime motor and decreasing the cost of rod pumping, a set of rated power calculated methods and a set of matching templates was built up under the regarding conditions such as the changing rule of the system load, the heating of the prime motor, and the overload torque in the initiate processing.\(^4\)

Calculating the efficiency of the beam pumping unit by analyzing the pattern of motion of the sucker rod and using the dynamometer card.\(^5\)

In petroleum exploitation, the beam pumping unit is used widely. But the phenomenon “the motor power is greater than the power required by the pumping unit” exists, which is caused by the reasons to satisfy the starting requirements of pumping unit and ensure sufficient overload capacity. Therefore, eventually it enhances the efficiency of electric motor of the unit, and causes the pumping unit to often work in light load. People have developed many kinds of energy-saving beam pumping unit at home and abroad. What’s more, the technology of power factor correction, soft start, adjustable speed motor, thyristor voltage regulator has appeared in the sphere of saving energy control beam pumping unit studies.\(^6\)

Analyzing the impact of balance on beam pumping unit of efficiency. At the same time, a self-balancing beam pumping unit was designed.\(^7\)

The predecessors studied the effects of motors, balance and sucker rod power on the efficiency of the pumping unit. Single factor analysis of pumping unit efficiency has certain limitations, ignoring the direct or indirect interaction between various factors. What’s more, the factors that affect the degree of the pumping unit efficiency cannot be directly reflected. Therefore, the principal component method in the statistical category is used, the principal component analysis is carried out on many factors, the primary and secondary relations of each factor are scientifically determined, and the deviation caused by subjective randomness is avoided.

In this article, the scoring model and the prediction model of pumping unit efficiency are established by principal component analysis of measured parameters. Model coefficients are trained and revised to ensure the accuracy of the model. The scoring model is used to judge inefficient factors, and the predictive model is used to predict system efficiency. The scoring model and the predictive model constitute the efficient evaluation model.
Principal component analysis is a dimensionality reducing multivariate statistical method. By means of principal component analysis, the original multiple related variables can be transformed into several principal components, usually expressed as a linear combination of the original related variables. The information principal components contain does not overlap and is unrelated to each other, and most of the information of the original variable is reflected. In general, when the question involves many variables, and the information contained in the variables overlaps, that is, there is a clear correlation between the variables, then principal component analysis can be considered, which can more easily capture the main contradiction of things and simplify the problem.\textsuperscript{8,9}

Parameter selection and data acquisition that affect the efficiency of the beam pumping unit

The choice of parameters and the acquisition of data are related to the accuracy of the evaluation model.

Parameter selection

The calculation formula of the beam pumping unit efficiency can be calculated by

\[
\eta = \frac{P_2}{P_1} \times 100\%
\]

\[
P_2 = \frac{QH_e \rho_l g}{86,400 \times 1000}
\]

\[
H_e = H_a + \frac{(F_a - F_b) \times 10^6}{\rho_l g}
\]

\[
\rho_l = f_w \rho_w + (1 - f_w) \rho_o
\]

where \(P_1\) is the motor input power, \(P_2\) is the output power, \(Q\) is the daily production of oil well, \(H_e\) is the effective lift, \(\rho_l\) is the mixed density, \(g\) is the gravitational acceleration, \(H_a\) is the liquid depth, \(F_a\) is the line pressure, \(f_w\) is the moisture content, \(\rho_o\) is the oil density, and \(\rho_w\) is the water density.

The above formula can directly derive the efficiency of the pumping unit. However, there is no explanation for the indirect factors affecting the efficiency of the beam pumping unit. In the previous studies, the pumping unit was divided into a ground part and a downhole part, and the ground part was composed of a motor part and a transmission part; the downhole part was mainly a reservoir characteristic. Zhang et al.\textsuperscript{10} analyzes the influencing factors of pump efficiency, and introduces factors that are closely related to reservoir characteristics, such as pumping depth, pump diameter, sinking degree, daily liquid volume, crude oil density, and so on. Feng et al.\textsuperscript{4} proposes the matching of the polished rod power of the pumping unit with the rated power of the motor, which is related to active power,
reactive power, apparent power, power factor, and current. The predecessors mainly studied the efficiency of the pumping unit from the ground part or the downhole part, but the factors of the ground part and the downhole part are mutually influential. Therefore, it is necessary to comprehensively consider the efficiency of the pumping unit, summarized as follows.\textsuperscript{11}

The factors affecting the efficiency of the beam pumping unit are mainly divided into three categories, the first type is the power source part, which is mainly the motor. The second type is the transmission part, which is mainly the pumping unit and the sucker rod. The third type is working part, which is mainly reservoir characteristics, including parameters shown in Tables 1–3.

Through the above summary, the factors affecting the efficiency of the beam pumping unit are 22 indicators.

**Selection of data**

These data were selected from the production data of 45 oil wells by the 22 kW motor unit of the 10 type pumping unit in the Xinjiang Oilfield of PetroChina. See Table 4 for details.

### Table 1. Index of power source part.

| Index                        | Unit       |
|------------------------------|------------|
| Upstroke maximum current     | $I_{u}, \text{A}$ |
| Downstroke maximum current   | $I_{d}, \text{A}$ |
| Motor input average current  | $I_{m}, \text{A}$ |
| Motor input reactive power   | $P_{c}, \text{kW}$ |
| Motor input apparent power   | $P_{a}, \text{kW}$ |
| Motor input active power     | $P_{b}, \text{kW}$ |
| Consumed power               | $P_{e}, \text{kW}$ |
| Power factor                 | $\varphi$  |
| Generated output             | $P_{d}, \text{kW}$ |

### Table 2. Index of transmission part.

| Index                        | Unit       |
|------------------------------|------------|
| Sucker rod power             | $P_{r}, \text{kW}$ |
| Effective head               | $C, \text{m}$ |
| Balance degree               | $T$        |
| Sucker rod maximum load      | $W_{u}, \text{kN}$ |
| Sucker rod minimum load      | $W_{d}, \text{kN}$ |
| Stroke                       | $S, \text{m}$ |
| Frequency                    | $n, \text{min}^{-1}$ |
It can be seen from Table 4 that the original data has large differences and many influencing factors, and it is difficult to conduct comprehensive and systematic analysis by conventional methods. Therefore, the principal component analysis method is used to establish the evaluation model to achieve more objective and accurate analysis.

### Logical block diagram of the evaluation model

In Figure 1, starting from the sample parameters, the principal component analysis is performed, and the comprehensive evaluation function equation and the regression equation can be obtained, respectively. Through the data training of the coefficients of the comprehensive evaluation function equation and the regression equation.

#### Table 3. Index of working part.

| Index                        | Unit   |
|------------------------------|--------|
| Liquid depth                 | $H_a$, m |
| Pump setting depth           | $H_b$, m |
| Dynamometer card of area     | $\beta$, m$^2$/m$^2$ |
| Submergence                  | $L$, m  |
| Line pressure                | $F_b$, MPa |
| Wellhead pressure            | $F_a$, MPa |

![Logical block diagram of the evaluation model](image)
Table 4. Statistical value of 22 indicators in 45 oil wells.

|       | Amount | Mean | Maximum | Minimum | Variance |
|-------|--------|------|---------|---------|----------|
| $I_u$, A | 45     | 25.23| 9.40    | 48.10   | 99.47    |
| $I_d$, A | 45     | 24.04| 8.6     | 44.1    | 93.57    |
| $I_m$, A | 45     | 13.60| 6.62    | 26.28   | 37.87    |
| $P_b$, kW | 45     | 5.26 | 2.20    | 9.42    | 2.51     |
| $P_c$, kW | 45     | 6.20 | 0.80    | 15.89   | 20.63    |
| $P_o$, kW | 45     | 8.59 | 3.98    | 17.11   | 15.27    |
| $\varphi$ | 45     | 0.68 | 0.28    | 0.99    | 0.05     |
| $P_{ch}$, kW | 45 | 0.46 | 0.00    | 1.84    | 0.25     |

|       | Amount | Mean | Maximum | Minimum | Variance |
|-------|--------|------|---------|---------|----------|
| $P_e$, kW | 45     | 4.80 | 2.05    | 8.52    | 2.69     |
| $F_a$, MPa | 45     | 0.86 | 0.60    | 1.50    | 0.03     |
| $F_b$, MPa | 45     | 0.91 | 0.00    | 1.50    | 0.12     |
| $W_u$, kN | 45     | 56.55| 43.69   | 69.28   | 40.35    |
| $W_d$, kN | 45     | 36.44| 27.31   | 45.23   | 18.13    |
| $S$, m | 45     | 4.45 | 2.88    | 4.93    | 0.27     |
| $n$, min$^{-1}$ | 45 | 4.18 | 2.20    | 4.60    | 0.22     |
| $H_a$, m | 45     | 1231.24 | 1721.40 | 250,338.98 |

|       | Amount | Mean | Maximum | Minimum | Variance |
|-------|--------|------|---------|---------|----------|
| $C$, m | 45     | 1225.78 | 26.60  | 1742.03 | 249,079.00 |
| $P_f$, kW | 45     | 1.62 | 0.05    | 3.77    | 0.92     |
| $T$ | 45     | 108.30| 21.00   | 252.52  | 2850.51  |
| $H_b$, m | 45 | 1735.75 | 1581.95 | 1883.60 | 6413.43 |
| $L$, m | 45     | 504.51 | 5.22    | 1684.45 | 240,455.44 |
| $\beta$, (m$^2$/m$^2$) | 45 | 0.28 | 0.04    | 0.76    | 0.03     |
equation, the scoring model and the predictive model which can objectively and accurately reflect the efficiency of the target pumping unit pumping unit are obtained. The two models together constitute the efficiency evaluation model of the beam pumping unit.

Substituting the target parameters into the prediction model, the efficiency of the predicted beam pumping unit is obtained, and the efficiency of the actual beam pumping unit and the efficiency of the predicted beam pumping unit are judged. If the actual beam pumping unit efficiency is greater than or equal to predicting the efficiency of the beam pumping unit by 90%, the output is that the actual beam pumping unit efficiency is in line with expectations. If the actual beam pumping unit efficiency is less than the forecasting efficiency of the beam pumping unit by 90%, the target parameters are substituted into the scoring model to find the inefficient factors, and then the adjusted target parameters are substituted into the forecasting model until the actual beam pumping unit efficiency is greater than or equal to the forecasting efficiency, output result.

**Establishing the scoring model based on principal component analysis**

**Solving the principal component from the correlation matrix**

In Table 4, \( n = 45 \) oil wells (samples), \( p = 22 \) parameters (variables), so that a total of \( n \times p = 45 \times 22 \) data, the original data matrix are

\[
x = \begin{pmatrix}
x_{11} & x_{12} & \ldots & x_{1p} \\
x_{21} & x_{22} & \ldots & x_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
x_{n1} & x_{n2} & \ldots & x_{np}
\end{pmatrix}
\]

(5)

\[
S = \frac{1}{n-1} \sum_{k=1}^{n} (x_{ki} - \bar{x}_i)(x_{ki} - \bar{x}_i)'
\]

(6)

\[
\bar{x}_i = \frac{1}{n} \sum_{k=1}^{n} x_{ki}, i = 1, 2, \ldots, p
\]

(7)

\[
R = (r_{ij})_{p \times p}, r_{ij} = \frac{S_{ij}}{\sqrt{S_{ii} \times S_{jj}}}
\]

(8)

where \( S \) is the sample covariance matrix, as an unbiased estimate of the overall covariance matrix, and \( R \) is the sample correlation matrix, which is the estimate of the overall correlation matrix. According to the definition of the overall principal component, the covariance of the principal component \( Y \) is

\[
\text{cov}(Y) = \Lambda
\]

(9)
where $\Lambda$ is the diagonal array

$$
\Lambda = \begin{pmatrix}
\lambda_1 & 0 & \ldots & 0 \\
0 & \lambda_2 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \lambda_p
\end{pmatrix}
$$

(10)

For $Y_1$, there is a maximum variance, $Y_2$ has a large variance, ..., and the covariance is

$$
cov(Y_i, Y_j) = \text{cov}(\gamma'_i X, \gamma'_j X) = \gamma'_i R \gamma_j $$

$$
= \gamma'_i \left( \sum_{a=1}^{p} \lambda_a \gamma_a \gamma'_a \right) \gamma_j $$

$$
= \sum_{a=1}^{p} \lambda_a (\gamma'_i \gamma_a) (\gamma'_a \gamma_j) = 0, \quad i \neq j
$$

It can be seen that the new comprehensive variables (principal components) $Y_1, Y_2, \ldots, Y_p$ are not related to each other. And the variance of $Y_i$ is $\lambda_i$, then $Y_1 = \gamma'_1 X$, $Y_2 = \gamma'_2 X$, ..., $Y_p = \gamma'_p X$ are called first, second, ..., $p$th principal components, respectively. Proportion of the variance of the $i$ principal component $y_i$ in the total variance $\lambda_i / \sum_{j=1}^{p} \lambda_j (i = 1, 2, \ldots, p)$ contribution rate is called principal component $y_i$. The contribution rate of the principal component reflects the ability of the principal component to synthesize the original variable information, and can also be understood as the ability to interpret the original variable.\(^{12,13}\)

The sum $\lambda_i / \sum_{j=1}^{p} \lambda_j$ of the contribution rates of the first $m$ ($m \leq p$) principal components is called the cumulative contribution rate of the first $m$ principal components, which reflects the ability of the first $m$ principal components to synthesize the original variable information (or explain the original variables).\(^{14,15}\)

The calculation process of the principal component is implemented in SPSS software, and results are shown in Table 5.

The correspondence between the correlation coefficient and the correlation degree is as follows.

Corresponding Table 5 correlation coefficient matrix and Table 6 correlation coefficient can be obtained according to the correlation degree.

There is a strong correlation between the upstroke maximum current ($I_u$), downstroke maximum current ($I_d$), motor input active power ($P_a$), motor input reactive power ($P_r$), and motor input apparent power ($P_d$), and they reflect the efficiency of the motor output.

There is a strong correlation between the sucker rod maximum load ($W_U$), sucker rod minimum load ($W_d$), liquid depth ($H_a$), effective head ($C$), and pump setting depth ($H_b$), and they reflect the output power of sucker rod.

There is a strong correlation between the liquid depth ($H_a$) and effective head ($C$), and they reflect the rise in crude oil.
Table 5. Index correlation coefficient matrix.

|   | $I_u$ | $I_d$ | $I_m$ | $P_b$ | $P_c$ | $P_a$ | $\varphi$ | $P_d$ | $P_e$ | $F_a$ | $F_b$ | $W_U$ | $W_d$ | $S$ | $n$ | $H_a$ | $C$ | $P_f$ | $T$ | $H_b$ | $L$ | $\beta$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $I_u$ | 1.000 | 0.250 | 0.651 | 0.644 | 0.474 | 0.650 | -0.186 | 0.342 | 0.518 | 0.232 | 0.188 | 0.437 | -0.184 | 0.415 | 0.116 | 0.172 | 0.167 | 0.045 | -0.632 | 0.209 | -0.141 | -0.317 |
| $I_d$ | 1.000 | 0.421 | 0.427 | 0.348 | 0.435 | -0.102 | -0.204 | 0.476 | 0.055 | 0.169 | 0.087 | -0.345 | 0.403 | 0.215 | -0.039 | -0.049 | 0.466 | 0.488 | 0.118 | 0.059 | 0.123 |
| $I_m$ | 1.000 | 0.475 | 0.944 | 0.999 | -0.714 | 0.318 | 0.362 | 0.146 | 0.098 | -0.050 | -0.252 | 0.280 | -0.059 | -0.084 | -0.086 | -0.11 | -0.247 | 0.240 | 0.125 | -0.098 |
| $P_b$ | 1.000 | 0.945 | -0.887 | 0.336 | 0.073 | 0.050 | -0.009 | -0.237 | -0.293 | 0.112 | -0.162 | -0.167 | -0.165 | -0.189 | -0.167 | 0.253 | 0.212 | -0.148 |
| $P_c$ | 1.000 | -0.712 | 0.323 | 0.360 | 0.138 | 0.090 | -0.055 | -0.278 | 0.277 | -0.050 | -0.078 | -0.080 | -0.013 | -0.238 | 0.259 | 0.122 | -0.102 |
| $P_a$ | 1.000 | -0.320 | 0.324 | 0.034 | 0.181 | 0.480 | 0.267 | 0.124 | 0.343 | 0.277 | 0.266 | 0.415 | 0.145 | -0.116 | -0.301 | 0.218 |
| $\varphi$ | 1.000 | -0.258 | 0.048 | 0.007 | 0.064 | 0.126 | 0.072 | -0.124 | 0.187 | 0.188 | -0.419 | -0.326 | 0.278 | -0.145 | 0.324 |
| $P_d$ | 1.000 | 0.201 | 0.283 | 0.469 | -0.031 | 0.532 | 0.296 | 0.043 | 0.030 | 0.583 | -0.089 | 0.008 | -0.043 | 0.186 |
| $P_e$ | 1.000 | 0.593 | 0.095 | -0.125 | 0.087 | -0.205 | 0.043 | 0.037 | 0.157 | -0.106 | -0.104 | -0.061 | 0.080 |
| $F_a$ | 1.000 | 0.246 | 0.111 | 0.145 | 0.034 | 0.080 | 0.028 | 0.202 | 0.060 | 0.149 | -0.057 | 0.158 |
| $W_U$ | 1.000 | 0.329 | 0.269 | 0.117 | 0.430 | 0.417 | 0.198 | -0.312 | 0.246 | -0.399 | -0.213 |
| $W_d$ | 1.000 | 0.061 | -0.034 | -0.032 | -0.045 | -0.094 | -0.026 | 0.170 | 0.061 | 0.079 |
| $S$ | 1.000 | 0.380 | 0.008 | 0.001 | 0.238 | -0.006 | -0.041 | -0.015 | -0.201 |
| $N$ | 1.000 | 0.058 | 0.048 | 0.200 | 0.170 | 0.217 | -0.024 | -0.117 |
| $H_a$ | 1.000 | 0.998 | 0.174 | -0.100 | 0.203 | 0.987 | -0.003 |
| $C$ | 1.000 | 0.165 | -0.109 | 0.189 | 0.988 | -0.012 |
| $P_f$ | 1.000 | 0.274 | -0.160 | -0.203 | 0.742 |
| $T$ | 1.000 | -0.032 | 0.096 | 0.345 |
| $H_b$ | 1.000 | -0.044 | -0.139 |
| $L$ | 1.000 | -0.019 |
| $\beta$ | 1.000 |
As shown in Table 7, the top seven principal components whose eigenvalues are greater than 1, the cumulative contribution rate reaches 86.013%. The first seven principal components represent most of the information for the 22 indicators. Therefore, 22 indicators affecting the efficiency of the pumping unit can be integrated into seven principal components.

Based on the selected seven comprehensive indicators, the initial factor load can be calculated. The principal component load matrix needs to divide the data in the initial factor load matrix by the eigenvalue corresponding to the principal component and then find the square root, and obtain the principal component load value corresponding to each index of the seven principal components. The principal component loads are shown in Table 8.9,16

Table 6. Corresponds to the degree of correlation.

| Correlation coefficient ($r$) | Relevance     |
|-------------------------------|---------------|
| $0.8 \leq r$                 | Highly correlated |
| $0.5 \leq r < 0.8$           | Moderate correlation |
| $0.3 \leq r < 0.5$           | Low correlation  |
| $r < 0.3$                     | Irrelevant     |

Table 7. Eigenvalue, variance contribution rat, and cumulative.

| Component | Initial eigenvalue | Extracting sum of squared loads |
|-----------|-------------------|---------------------------------|
| Component | Total          | % of variance | Cumulative | Total          | % of variance | Cumulative |
| 1         | 5.115          | 23.249        | 23.249     | 5.115          | 23.249        | 23.249     |
| 2         | 4.345          | 19.748        | 42.997     | 4.345          | 19.748        | 42.997     |
| 3         | 3.236          | 14.707        | 57.704     | 3.236          | 14.707        | 57.704     |
| 4         | 2.075          | 9.431         | 67.135     | 2.075          | 9.431         | 67.135     |
| 5         | 1.687          | 7.667         | 74.802     | 1.687          | 7.667         | 74.802     |
| 6         | 1.336          | 6.074         | 80.877     | 1.336          | 6.074         | 80.877     |
| 7         | 1.130          | 5.136         | 86.013     | 1.130          | 5.136         | 86.013     |
| 8         | 0.813          | 3.694         | 89.707     |                |                |            |

As shown in Table 7, the top seven principal components whose eigenvalues are greater than 1, the cumulative contribution rate reaches 86.013%. The first seven principal components represent most of the information for the 22 indicators. Therefore, 22 indicators affecting the efficiency of the pumping unit can be integrated into seven principal components.

Based on the selected seven comprehensive indicators, the initial factor load can be calculated. The principal component load matrix needs to divide the data in the initial factor load matrix by the eigenvalue corresponding to the principal component and then find the square root, and obtain the principal component load value corresponding to each index of the seven principal components. The principal component loads are shown in Table 8.9,16

Establishment of a scoring model

According to the main component load of Table 8, the analysis of the seven principal components is as follows (Table 9 to Table 15).

According to the principal component load value and the normalized data of each original index, the corresponding principal component expression and comprehensive evaluation function can be obtained. Due to space limitation, just the calculation expression of the first principal component can be listed
Table 8. Principal component load.

| Index | Component | 1   | 2   | 3   | 4   | 5   | 6   | 7   |
|-------|-----------|-----|-----|-----|-----|-----|-----|-----|
| $I_u$ | 0.839     | 0.146 | -0.245 | -0.221 | 0.062 | -0.209 | -0.012 |
| $I_d$ | 0.519     | 0.153 | 0.508 | 0.334 | -0.303 | 0.134 | -0.204 |
| $I_m$ | 0.912     | -0.327 | -0.006 | 0.161 | 0.014 | 0.034 | 0.120 |
| $P_b$ | 0.723     | 0.503 | 0.244 | -0.226 | 0.030 | -0.076 | 0.153 |
| $P_c$ | 0.759     | -0.565 | -0.063 | 0.250 | -0.029 | 0.086 | 0.082 |
| $P_a$ | 0.914     | -0.327 | -0.007 | 0.173 | -0.006 | 0.038 | 0.108 |
| $\varphi$ | -0.422 | 0.817 | 0.153 | -0.279 | -0.009 | -0.035 | -0.039 |
| $P_d$ | 0.305     | -0.216 | -0.612 | -0.120 | 0.044 | 0.232 | 0.002 |
| $P_e$ | 0.606     | 0.553 | 0.424 | -0.182 | 0.016 | -0.145 | 0.148 |
| $F_a$ | 0.253     | 0.165 | 0.067 | 0.038 | 0.743 | 0.070 | -0.454 |
| $F_b$ | 0.246     | 0.318 | 0.149 | -0.069 | 0.533 | 0.481 | -0.373 |
| $W_U$ | 0.249     | 0.677 | -0.270 | -0.380 | 0.016 | 0.075 | 0.099 |
| $W_d$ | -0.242    | 0.166 | -0.099 | -0.509 | 0.091 | 0.488 | 0.432 |
| $S$   | 0.502     | 0.311 | 0.216 | -0.305 | -0.266 | -0.090 | -0.218 |
| $N$   | 0.102     | 0.341 | 0.203 | -0.214 | -0.620 | 0.089 | -0.249 |
| $H_a$ | 0.050     | 0.639 | -0.625 | 0.426 | -0.074 | 0.021 | -0.014 |
| $C$   | 0.042     | 0.624 | -0.635 | 0.433 | -0.085 | -0.012 | -0.004 |
| $P_f$ | 0.138     | 0.619 | 0.555 | 0.340 | 0.090 | -0.064 | 0.251 |
| $T$   | -0.317    | 0.050 | 0.535 | 0.423 | -0.297 | 0.407 | -0.223 |
| $H_b$ | 0.281     | 0.020 | -0.292 | -0.052 | -0.284 | 0.695 | 0.091 |
| $L$   | -0.005    | -0.649 | 0.390 | -0.443 | 0.029 | 0.092 | 0.030 |
| $\beta$ | -0.166 | 0.246 | 0.506 | 0.457 | 0.311 | 0.158 | 0.485 |

Table 9. First principal component analysis.

| First principal component factor | Component load | Contribution rate | Analysis |
|---------------------------------|----------------|------------------|---------|
| Upstroke maximum current ($I_u$) | 0.839          | 23.249%          | The absolute value of the input motor input average current and the motor input apparent power correlation coefficient exceeds 0.9, which is highly correlated, mainly reflecting the actual output efficiency of the motor. |
| Downstroke maximum current ($I_d$) | 0.519          |                  |         |
| Motor input average current ($I_m$) | 0.912          |                  |         |
| Motor input apparent power ($P_a$) | 0.723          |                  |         |
| Motor input active power ($P_b$) | 0.759          |                  |         |
| Motor input reactive power ($P_c$) | 0.914          |                  |         |
| Consumed power ($P_e$) | 0.606          |                  |         |

\[ Y_1 = 0.839x_{I_u} + 0.519x_{I_d} + 0.912x_{I_m} + 0.723x_{P_b} + 0.759x_{P_c} + 0.914x_{P_a} - 0.422x_\varphi + 0.305x_{P_d} + 0.606x_{P_e} + 0.253x_{F_a} + 0.246x_{F_b} + 0.249x_{W_U} - 0.242x_{W_d} + 0.502x_S + 0.102x_N + 0.050x_{H_a} + 0.042x_C + 0.138x_{P_f} - 0.317x_T + 0.281x_{H_b} - 0.005x_L - 0.166x_\beta \]
### Table 10. Second principal component analysis.

| Second principal component factor | Component load | Contribution rate | Analysis |
|-----------------------------------|----------------|-------------------|----------|
| Sucker rod maximum load ($W_u$)   | 0.660          | 19.748%           | It mainly reflects the work done by the power of the polished rod to overcome the lift of the crude oil. |
| Power factor ($\phi$)             | 0.817          |                   |          |
| Liquid depth ($H_a$)              | 0.639          |                   |          |
| Effective head ($C$)              | 0.624          |                   |          |
| Pump setting depth ($H_b$)        | -0.649         |                   |          |
| Sucker rod power ($P_f$)          | 0.619          |                   |          |

### Table 11. Third principal component analysis.

| Third principal component factor | Component load | Contribution rate | Analysis |
|----------------------------------|----------------|-------------------|----------|
| Liquid depth ($H_a$)             | -0.625         | 14.707%           | Mainly reflects the impact of motor power generation. |
| Power factor ($\phi$)            | -0.612         |                   |          |
| Effective head ($C$)             | -0.635         |                   |          |
| Downstroke maximum current ($l_d$) | 0.508       |                   |          |

### Table 12. Fourth principal component analysis.

| Fourth principal component factor | Component load | Contribution rate | Analysis |
|-----------------------------------|----------------|-------------------|----------|
| Sucker rod minimum load ($W_d$)   | -0.509         | 9.431%            | It mainly reflects the influence of the sucker rod minimum load. |
| Liquid depth ($H_a$)              | 0.426          |                   |          |
| Pump setting depth ($H_b$)        | 0.433          |                   |          |
| Submergence ($L$)                 | -0.443         |                   |          |
| Dynamometer card of area ($\beta$) | 0.457       |                   |          |
| Balance degree ($T$)              | 0.423          |                   |          |

### Table 13. Fifth principal component analysis.

| Fifth principal component factor | Component load | Contribution rate | Analysis |
|----------------------------------|----------------|-------------------|----------|
| Wellhead pressure ($F_a$)        | 0.743          | 7.667%            | Mainly reflects the impact of wellhead pressure. |
| Line pressure ($F_b$)            | 0.533          |                   |          |
| Frequency ($n$, min$^{-1}$)      | -0.620         |                   |          |
where $Y_1$ represents the first principal component, and the index ($x_{Iu}, x_{Id}, \ldots, x_{\beta}$) on the right side of the equation is obtained by normalizing the original index ($I_u, I_d, \ldots, \beta$); the coefficient terms of each standardized data are as shown in Table 8. Lists the load values of the principal components. The rest of the principal component expressions are analogous\(^{17}\)

$$Y = \frac{\sum_{i=1}^{7} \lambda_i Y_i}{(\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 + \lambda_6 + \lambda_7)}$$

$$= \frac{5.115}{18.923} Y_1 + \frac{4.345}{18.923} Y_2 + \frac{3.236}{18.923} Y_3 + \frac{2.075}{18.923} Y_4 + \frac{1.687}{18.923} Y_5$$

$$+ \frac{1.336}{18.923} Y_6 + \frac{1.130}{18.923} Y_7$$

$$= 0.270Y_1 + 0.230Y_2 + 0.171Y_3 + 0.110Y_4 + 0.089Y_5$$

$$+ 0.071Y_6 + 0.060Y_7$$

### Comprehensive value analysis

Substitute the data of 40 oil wells into the comprehensive evaluation function equation (12), and the comprehensive value ranking was obtained (as shown in Table 16).

Figure 2 is the corresponding relationship between the comprehensive value $Y$ of the No. 40 oil well and the efficiency of the pumping unit of the pumping unit. As shown in Figure 2, the comprehensive score interval is divided into A (−1.0 to −0.5), B (−0.5 to 0), C (0 to 0.5), and D (0.5 to 1.0). The efficiency range of the

| Table 14. Sixth principal component analysis. |
|---------------------------------------------|
| Sixth principal component factor | Component load | Contribution rate | Analysis                                      |
|---------------------------------|----------------|--------------------|-----------------------------------------------|
| Motor input active power ($F_b$) | 0.481          | 6.074%             | Mainly reflects                              |
| Sucker rod minimum load ($W_d$)  | 0.488          |                    |                                               |
| Pump setting depth ($H_b$)       | 0.695          |                    |                                               |

| Table 15. Seventh principal component analysis. |
|-----------------------------------------------|
| Seventh principal component factor | Component load | Contribution rate | Analysis                                      |
|-----------------------------------------|----------------|--------------------|-----------------------------------------------|
| Wellhead pressure ($F_a$)               | −0.454         | 5.136%             | Reflects the impact of the dynamometer card of area. |
| Sucker rod minimum load ($W_d$)         | 0.432          |                    |                                               |
| Dynamometer card of area ($\beta$)      | 0.485          |                    |                                               |
beam pumping unit is divided into I (0%–20%) low efficiency zone, II (20%–40%) medium efficiency zone, III (40%–60%) high efficiency zone, and IV (60%–100%) high efficiency zone. 

**Figure 2.** Beam pumping unit efficiency and comprehensive score.
extremely efficiency zone, and the distribution points of each interval are shown in Table 17.

As shown in Table 17, 66.7% of the distribution points in B zone are concentrated in the I (0%–20%) low zone; 84.6% of the C zone distribution points are concentrated in the II (20%–40%), III (40%–60%), and IV (60%–100%); 100% of D zone distribution point concentrated in II (20%–40%) and III (40%–60%).

As illustrated in Figure 3, When the comprehensive score is located in Zone C or Zone D (the comprehensive score is greater than 0), the probability of pumping unit efficiency reaching 20% (or above 20%) is significantly improved. Revealing that in the current working conditions, the higher comprehensive score corresponds to the higher probability of the beam pumping unit efficiency, especially when the composite score is greater than 0.

The purpose of the scoring model is to find the inefficient factors that affect the efficiency of the beam pumping unit, and the prediction model is used as a means of judging whether the target well is meeting the efficiency of the expected beam pumping unit.

**Table 17.** Comprehensive score and efficiency distribution of beam pumping unit.

|       | A (−1.0 to −0.5) | B (−0.5 to 0) | C (0 to 0.5) | D (0.5 to 1.0) | Total |
|-------|------------------|--------------|-------------|----------------|-------|
| I (0%–20%) | 1 | 6 | 4 | 0 | 11 |
| II (20%–40%) | 0 | 0 | 11 | 2 | 13 |
| III (40%–60%) | 0 | 2 | 9 | 2 | 13 |
| IV (60%–100%) | 0 | 1 | 2 | 0 | 3 |

**Figure 3.** The probability that the data points of the four category of scoring areas fall in the efficiency II zone and above zone.
Table 18. Top seven main components of 45 wells.

| No. | First principal component | Second principal component | Third principal component | Fourth principal component | Fifth principal component | Sixth principal component | Seventh principal component |
|-----|---------------------------|----------------------------|---------------------------|----------------------------|---------------------------|---------------------------|----------------------------|
| 1   | -0.07081                  | 0.04537                    | 0.6656                    | 1.27099                    | -0.67484                  | 1.35706                   | -1.43023                   |
| 2   | -1.07796                  | 0.65268                    | 0.64078                   | 0.65722                    | -0.21234                  | 0.73437                   | -0.18845                   |
| 3   | -1.01938                  | 1.56914                    | 0.62406                   | 0.96411                    | -0.0104                   | 0.0247                    | 0.49155                    |
| 4   | -0.63951                  | 1.52511                    | 1.13558                   | 0.36953                    | 0.61334                   | 0.16023                   | -0.87263                   |
| 5   | -0.11725                  | 2.05968                    | 0.66045                   | -0.06291                   | -0.03724                  | -0.06872                  | 0.50852                    |
| 6   | -0.35921                  | 1.64612                    | 0.35109                   | 0.21794                    | 0.21507                   | 0.46674                   | 0.62474                    |
| 7   | -0.24316                  | 1.15282                    | 1.07974                   | -0.25711                   | -0.23799                  | -0.36396                  | 0.80206                    |
| 8   | -0.33924                  | 1.68316                    | 0.38057                   | 0.41869                    | 0.20527                   | -0.15243                  | 0.24433                    |
| 9   | 0.15038                   | 0.21433                    | 0.42034                   | 2.09949                    | -0.55269                  | 0.93468                   | -0.69299                   |
| 10  | 0.10173                  | -0.45529                   | 0.61957                   | 0.24598                    | 0.73637                   | 0.33569                   | -0.41954                   |
| 11  | -1.52934                  | -1.23517                   | 1.64559                   | 0.10657                    | -0.36772                  | -1.5593                   | 0.33713                    |
| 12  | 0.8221                   | 0.1066                     | 1.6677                    | 0.5392                     | -1.10809                  | -0.96914                  | 0.99388                    |
| 13  | -1.08224                  | -0.48265                   | 2.4032                    | 0.20634                    | -0.4732                   | -1.0684                   | 0.48388                    |
| 14  | -1.31991                  | 0.8048                     | 0.46983                   | 0.36819                    | 0.58541                   | 0.8813                    | 0.52783                    |
| 15  | -1.13305                  | 0.63889                    | 1.21624                   | 0.19666                    | 0.0655                    | 0.87727                   | 0.70762                    |
| 16  | 0.63721                   | -0.05403                   | 1.29974                   | -0.22259                   | 0.33976                   | -0.33819                  | 0.12287                    |
| 17  | -0.1944                  | 1.78055                    | 0.84286                   | -0.29221                   | 0.6499                    | 0.48465                   | 0.52979                    |
| 18  | -0.58918                  | 0.79636                    | 1.45459                   | -0.23181                   | 1.31469                   | 2.17721                   | 0.93294                    |
| 19  | -0.05366                  | 1.17182                    | 0.76903                   | -0.41755                   | 1.5885                    | 1.03641                   | 0.76609                    |
| 20  | 0.36092                   | 1.67326                    | 0.56826                   | -0.31616                   | 0.69327                   | 0.36382                   | 0.85603                    |
| 21  | -0.03847                  | 0.32927                    | 1.52295                   | -1.17604                   | 0.9899                    | -1.05957                  | 0.67258                    |
| 22  | -0.14278                  | 0.84431                    | 2.33547                   | 0.44291                    | 0.9547                    | 1.64394                   | 1.63271                    |
| 23  | 0.10398                   | -0.60516                   | 2.27916                   | -1.29852                   | 0.8336                    | -0.93252                  | 1.29948                    |
| 24  | -0.60892                  | 0.15739                    | 0.91382                   | -0.86511                   | 1.73742                   | -0.02516                  | -0.10143                   |
| 25  | 0.78106                  | -0.53065                   | 2.13997                   | 0.49988                    | 0.0579                    | -0.13425                  | 0.13514                    |
| 26  | -0.60803                  | 0.79209                    | 0.37826                   | -1.49833                   | 2.16608                   | 0.16506                   | -0.18092                   |
| 27  | -0.81149                  | 0.95339                    | 0.83719                   | -0.21618                   | 2.37102                   | 0.85229                   | 0.49346                    |
Table 18. (Continued)

| No. | First principal component | Second principal component | Third principal component | Fourth principal component | Fifth principal component | Sixth principal component | Seventh principal component |
|-----|---------------------------|----------------------------|---------------------------|---------------------------|--------------------------|---------------------------|-----------------------------|
| 28  | -0.85824                  | 0.46394                    | 0.67045                   | 0.0657                    | -0.84704                 | -2.56433                  | 0.05895                     |
| 29  | 0.0678                    | 0.12016                    | 1.52656                   | 1.77804                   | -2.02369                 | -1.42059                  | -0.07124                    |
| 30  | -0.54264                  | 1.0765                     | 0.51208                   | 1.48216                   | -0.99451                 | 0.33097                   | -0.06786                    |
| 31  | -0.86555                  | 0.84045                    | 1.13614                   | 0.55067                   | -1.71166                 | 1.683                     | 0.02809                     |
| 32  | -0.25412                  | -0.32881                   | 1.07399                   | 0.95503                   | -2.22444                 | 0.05964                   | -0.26024                    |
| 33  | 0.34386                   | 0.36889                    | 0.53306                   | 2.22385                   | -2.70923                 | 1.75291                   | 0.60017                     |
| 34  | 0.14396                   | 1.91592                    | 0.47181                   | -0.1392                   | -0.55031                 | -0.81151                  | 0.2075                      |
| 35  | -0.03799                  | 1.80419                    | 0.79015                   | -0.5598                   | -0.4641                  | -0.01905                  | -0.91123                    |
| 36  | -0.3799                   | 2.2514                     | 0.27356                   | 0.23486                   | -1.55107                 | -0.83694                  | -0.70881                    |
| 37  | -0.96806                  | -0.29647                   | 1.0606                    | -0.56068                  | 0.68975                  | -0.64758                  | -1.25853                    |
| 38  | 0.42608                   | 0.56759                    | 0.44098                   | 0.5326                    | -0.36242                 | 0.29178                   | -1.31283                    |
| 39  | -0.77436                  | 0.59107                    | 0.11115                   | 1.24399                   | -0.40417                 | 0.45468                   | -0.35982                    |
| 40  | 0.4212                    | 1.0087                     | 0.55641                   | 1.64997                   | -0.39521                 | -0.58101                  | -0.37509                    |
| 41  | 0.2912                    | 1.17858                    | 0.30893                   | 2.13772                   | -1.00643                 | 0.17996                   | -0.56308                    |
| 42  | -0.05513                  | 0.30409                    | -0.95279                  | -1.71845                  | 0.88299                  | 0.3431                    | 0.54475                     |
| 43  | -0.42355                  | 1.6786                     | 0.49097                   | -0.24202                  | 1.21417                  | -0.70683                  | -0.37162                    |
| 44  | -0.67817                  | 0.23843                    | 1.87342                   | -0.79021                  | -0.2243                  | -2.72484                  | 0.41878                     |
| 45  | -1.16773                  | 1.56484                    | 0.02676                   | 0.4688                    | -0.72801                 | -0.10021                  | -0.95109                    |
Establishing the predictive model based on multiple linear regression equation

**Establishment of regression equation based on principal component**

The raw data of the No. 40 oil well were used to establish a predictive model that is the efficiency of the beam pumping unit by suing multiple linear regression analysis. And the Nos 41–45 oil well was used as the inspection data.

The multiple linear regression analysis was performed using the first seven principal component load scores of Nos 1–40 oil wells in Table 18, and the regression equation coefficients were obtained, Table 19.

As illustrated in Table 19, listing the significance test results of independent variables (using Student’s *t*-test). Regression factor 1, regression factor 3, regression factor 4, and regression factor 5 pass the *t*-test with a significance level of 0.05, indicating that the above regression factors have significant effects on the dependent variables.8

The regression equation can be obtained

\[ Z = 48.126 + 11.261x_1 - 8.465x_3 - 4.872x_4 + 10.537x_5 \]  \hspace{1cm} (14)

The regression equation needs to be verified by data before it can be used as a predictive model.

**Inspection and application of predictive models**

The error of the prediction model of the beam pumping unit efficiency is not clearly defined in the seniors’ research. Therefore, we delimit the relative error of the predicted pumping unit efficiency is less than 10%, which is considered to be in line with expectations. The error of the prediction model can be calculated by
The method is used to predict the efficiency of the beam pumping unit of five wells, and the average relative error is 7.4%. It shows that the regression equation can accurately predict the efficiency of the beam pumping unit of the 10 type 22 kW motor unit. As illustrated in Figure 4, no. 43 is located above 10% of the relative error line and is considered to be much lower than the predicted beam pumping efficiency. This point can be placed in the scoring model to find out the inefficient factors of the well, and then improve beam pumping unit efficiency.

**Conclusion**

For the oil well parameters adopted in this article, the following conclusions are obtained:

1. A total of 22 original indicators affected the efficiency of the pumping unit. Through principal component analysis, this article selects the top seven for regression analysis (the total contribution rate of the top seven principal components has reached 86.013%). This not only reduces the number of indicators that need to be considered and the dimension of the problem but also reflects the information of the original parameters as much as possible, and the components are independent of each other, which lays the foundation for the next step to establish an evaluation model.

2. Based on the principal component eigenvalues, the established scoring model, the integrated value obtained by the data, reflects the
correspondence between the comprehensive value and the efficiency of the beam pumping unit. When the comprehensive score is greater than or equal to 1.0, the probability of the corresponding medium-high beam pumping unit efficiency is significantly improved; for oil wells with a comprehensive score less than 1.0, it can be considered to find the inefficient factor from the seven principal component components.

3. Based on the principal component components, the multi-step linear regression method was used to predict the efficiency of the pumping unit. The results show that the prediction model has higher accuracy and provides a reference standard for judging the efficiency of the beam pumping unit.

4. The efficiency evaluation model of the beam pumping unit is composed of a scoring model and a predictive model, and can judge and improve the pumping unit efficiency of the target oil well.

5. According to the pumping unit model, motor power, reservoir characteristics, and other parameters to refine the classification, and then establish an evaluation model, can more accurately and reasonably evaluate the efficiency of the target beam pumping unit.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work is supported by the key Laboratory of Oil and Gas Equipment Foundation of Ministry of Education.
References

1. Pang Y-H, Liu F, Yang X, et al. An analysis of power economizing for motor used for beam pumping unit. *Oil Field Equip* 2004; 6: 79–81.
2. Feng J. Improvement on energy saving effect of electromotor for beam pumping unit. *J Xinjiang Oil Gas* 2006; 2(3): 91–92.
3. Zhou L, Li H, Rang Y, et al. Study on energy saving for nodding donkey oil pump. In: *Proceedings of the 2011 2nd international conference on artificial intelligence, management science and electronic commerce (AIMSEC)*, Dengleng, China, 8–10 August 2011, pp. 4362–4364. New York: IEEE.
4. Feng Z-M, Tan J-J, Liu X-L, et al. Selection method modelling and matching rule for rated power of prime motor used by beam pumping units. *J Petrol Sci Eng* 2017; 153: 197–202.
5. Grana D and Daly C. Petroleum geostatistics. *Math Geosci* 2017; 49(4): 439–440.
6. Liu X and Xu J. The method of energy saving in beam pumping unit based on genetic algorithm. *AASRI Proc* 2012; 1: 441–447.
7. Wu J, Wang Q and Han Y. Finite element analysis of walking beam of a new compound adjustment balance pumping unit. In: *Proceedings of the 1st international conference on frontiers of materials synthesis and processing (FMSP 2017)*, Changsha, China, 28–29 October 2017.
8. Huang Y, Shen L and Liu H. Grey relational analysis, principal component analysis and forecasting of carbon emissions based on long short-term memory in China. *J Clean Prod* 2019; 209: 415–423.
9. Fan S, Zhang Y, Zhang Y, et al. Motion process monitoring using optical flow-based principal component analysis-independent component analysis method. *Adv Mech Eng* 2017; 9(11): 733231.
10. Zhang X-D, Xie X-H, Li Z-Y, et al. Application of principal component analysis in influence factor evaluation of oil well pump efficiency. *J Southwest Petrol* 2011; 33(5): 176–180.
11. Wang X, Yu W and Li X-L. Calculation analysis and utilization of rod power. *Agr Equip Vehicle Eng* 2009; 5: 42–44.
12. Liu RX, Kuang J, Gong Q, et al. Principal component regression analysis with SPSS. *Comput Programs Biomed* 2003; 71(2): 141–147.
13. Zhang J, Xie J, Zhang H. Production capacity and mining plan optimization of fault/fracture-controlled EGS model in Gonghe Basin. *Energy Sci Eng* 2019; 00: 1–18.
14. Chen H, Jiang B, Lu N, et al. Multi-mode kernel principal component analysis-based incipient fault detection for pulse width modulated inverter of China Railway High-speed 5. *Adv Mech Eng* 2017; 9(10): 727383.
15. Wang C, Fu Z and Cui G. A neural-network-based approach for diagnosing hardware faults in cloud systems. *Adv Mech Eng* 2019; 11(2): 819236.
16. He Y, Pang Y, Zhang Q, et al. Comprehensive evaluation of regional clean energy development levels based on principal component analysis and rough set theory. *Renew Energ* 2018; 122: 643–653.
17. Zhang J, Wang C, Liu L, et al. Investigation of carbon dioxide emission in China by primary component analysis. *Sci Total Environ* 2013; 472: 239–247.

**Author biographies**

Zhewei Ye is a teacher at Southwest Petroleum University and a PhD from Sichuan University. He is mainly engaged in drilling tool research.

Zhiyang Liu, Master of Southwest Petroleum, he is mainly engaged in oil field big data research.

Cheng Cheng, Master of Southwest Petroleum, he is mainly engaged in screw drill research.

Lang Tan, Master of Southwest Petroleum, he is mainly engaged in screw pump research.

Kang Feng, Master of Southwest Petroleum, he is mainly engaged in screw motor research.