Analysis of the Key Factors of Pumping Well System Efficiency for Oil Field Based on Multiple Regression

Bin Yao*
Sinopec Northwest Oil Field Company, Wulumuqi 830011, China

*Corresponding author email: xinjiangyaobin@163.com

Abstract. Many factors are affecting the efficiency of the pumping well system and the lack of pertinence in oil well governance. It is proposed to use the block as a unit to use the multiple regression method to study the key influencing factors of the system efficiency, determine the degree of influence of various factors on the system efficiency through standardized regression coefficients, and grasp the main contradictions that restrict the efficiency of the oil well system. Guide the adjustment and optimization of oil wells, provide theoretical support for the targeted management of pumping well system efficiency, and improve the comprehensive management level of the oilfield.

Keywords: System efficiency; Multiple regression; Key factors; Standardization

1. Introduction
The improvement of the efficiency of the pumping well system has always been an important part of the oilfield to achieve energy-saving and consumption reduction, increase production, and efficiency[1-3]. With the continuous improvement in the coverage of various energy-saving products and the level of oilfield management in recent years, although the energy-saving effect of pumping wells has shown an upward trend, the increase has decreased significantly, and the energy-saving space for system efficiency has been continuously reduced[4,5]. Since many specific factors affect system efficiency, the key factors that affect the system efficiency of different oil wells are not the same, and it is impossible to determine the degree of influence of each factor and it is impossible to carry out targeted governance work. In the same block, the reservoir types, development stages, supporting capabilities, and other indicators between different wells are similar. The main factor causing the difference in system efficiency is concentrated on the oil production parameters.

Zhang[6] found that when the polished rod load of a pumping unit is subject to random uncertain disturbance, the reduced-order observer is used to reconstruct the load disturbance, and the influence of the load disturbance on the system output can be eliminated through the feedforward control. Song[7] found that the appropriate variable speed operation scheme can improve the system efficiency and power-saving rate, reduce the motor power peak and torque peak value, and reduce the gearbox torque peak value and suspension point load peak value. Variable speed optimization technology can improve the comprehensive performance of pumping well systems, increase production, reduce energy consumption, and improve safety.

Li[8] uses the method based on curve moment and PSO-SVM to diagnose the downhole conditions of the rod pumping system. Li[9] proposed a multi-fault diagnosis method based on Freeman chain code and designated meta-analysis. Ren[10] proposed a fault diagnosis method for pumping units based on a biogeographic optimization algorithm. Liang[11] proposed a hierarchical diagnosis method for rod pumping system faults based on indicator diagrams.

Content from this work may be used under the terms of the CreativeCommons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd
Therefore, taking the block as a unit to research the key influencing factors of the system efficiency can be the key to further improving the efficiency of the pumping well system. This article introduces a method of using multiple regression method to analyze the key factors affecting the efficiency of the pumping well system. Emphasizes the method of establishing regression model, precautions, selection of influencing factors, and how to use the data obtained from regression to find out the key factors affecting the efficiency of the pumping well system, and formulate targeted improvement plans accordingly to improve the comprehensive management of the oilfield level.

2. Theoretical Basis of Multiple Regression

Regression analysis is a mathematical statistics method that can extract important information hidden in large-scale original data groups, simplify the data while constructing simple functional relationships between variables, and use standardized regression coefficients to reflect the degree of influence of each independent variable on the dependent variable. This article uses a multiple regression method to statistically analyze the many factors that affect the efficiency of the pumping well system, to find out the main factors restricting the improvement of efficiency, and implement targeted transformations.

Applying the idea of regression analysis to the system efficiency analysis of pumping wells, oil extraction parameters such as pump diameter, submersion, and motor installed power can be regarded as "independent variables" in regression analysis, while system efficiency can be regarded as "dependent variable". As a result, the corresponding relationship between multiple "independent variables" and a single "dependent variable" is formed. According to this corresponding relationship, the multiple regression method in regression analysis can be used to model it. By substituting various factors affecting the efficiency of the pumping well system into the mathematical model of multiple regression, the equation with the system efficiency can be established:

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \epsilon \]  

To study the influence of various factors on system efficiency, it is necessary to perform regression operations on multiple sets of data, and the equations they constitute can be written in the form of a matrix.

\[ Y = X \beta + \epsilon \]  

among them: \( Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}; \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}; \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \)

Where, \( x_{ij} \) — efficiency influencing factors such as stroke, stroke order, etc. \((j=1,2,\ldots,k)\); \( y_i \) — single well efficiency; \( \beta_0, \beta_1, \ldots, \beta_p \) — regression coefficient corresponding to each factor; \( \epsilon \) — random error. Researchers have done a lot of research on the key factors affecting the efficiency of the pumping well system through theoretical research and field experimental data and concluded that the main factors affecting efficiency include: stroke frequency, stroke frequency, pump efficiency, load utilization, motor power utilization, etc. Over twenty items. These factors affect each other. Researchers can make judgments based on their own experiences. To prevent strong multicollinearity between the respective variables, some factors with high correlation should be eliminated, such as the suspension of the same well. There is a strong correlation between the point maximum load and the maximum torque of the output shaft of the gearbox, and only one of them should be kept. In the regression method, the stepwise regression method is used to ensure the significance level of the regression equation.

After the regression model is established, it needs to be tested for regression significance. The content of the test mainly includes two aspects: one is to test the entire equation (F test) to ensure that there is a certain linear relationship between the respective variables and the dependent variable, and the other is the significance analysis of each factor (T-test). To ensure that the regression coefficient corresponding to each factor is not zero.
After the regression model is established, it needs to be tested for regression significance. The content of the test mainly includes two aspects: one is to test the entire equation (F test) to ensure that there is a certain linear relationship between the respective variables and the dependent variable, and the other is the significance analysis of each factor (T-test). To ensure that the regression coefficient corresponding to each factor is not zero.

After the equation is tested, the standardized regression coefficients in the regression results can be used to quantitatively describe the degree of correlation between various factors and system efficiency, find the main factors affecting system efficiency, and formulate targeted improvement plans. The whole analysis process is shown in Figure 1.

### Figure 1. Multiple regression analysis processes.

#### 3. Example Analysis of Factors Affecting System Efficiency of Pumping Wells

Based on the above analysis, a total of 10 factors affecting the efficiency of the pumping well system are selected, namely: pump diameter \((D)\), pump efficiency \((\eta_p)\), submergence \((H)\), stroke \((S)\), stroke frequency \((n)\), motor active power Power \((P_t)\), load utilization \((\eta_L)\), motor power utilization \((\eta_e)\), torque utilization \((\eta_m)\), balance \((\phi)\).

A block is randomly selected as the research object. There are 198 wells in this block. The above 10 influencing factors are used as independent variables, and the system efficiency is used as the dependent variable for multiple regression calculation. The calculation process is carried out in SPSS software. Table 1 is the generated efficiency regression model. The ‘R’, ‘R-square’, and ‘Adjusted R-square’ in the table are the multiple correlation coefficient, the coefficient of determination, and the corrected coefficient of determination, respectively. Under normal circumstances, the calibration coefficient of determination is used to measure the quality of the model built. Its range is between 0 and 1. The closer its value is to 1, the better the model fits. The ‘F test’ has a true rejection probability value Sig. F of 0.000, so the null hypothesis that there is no linear relationship between the independent variable and the dependent variable can be significantly rejected.

### Table 1. Regression model information.

| Model | R    | R²   | Adjustment R² | Standard estimation error | F test  | Sig. F |
|-------|------|------|---------------|---------------------------|---------|--------|
| 1     | 0.887| 0.789| 0.782         | 5.73                      | 119.695 | 0.000  |

The regression coefficient and variance test table of the model are shown in Table 2. The stepwise regression analysis method has automatically eliminated some factors that have a low degree of influence on the dependent variable. Table 2 lists the standardized regression coefficients of various factors on system efficiency. The standardized regression coefficients can be used to construct the standardized multiple Regression equation. ‘T-test’ shows that the probability of rejection of truth for all factors except the constant term is less than 0.05, so all meet the regression requirements. The VIF
value of each parameter in the collinearity diagnosis item is less than 5, and all of them pass the collinearity diagnosis.

### Table 2. Regression coefficient and variance test table.

| Model                      | Standardized regression coefficient | T-Test | Sig. T | Collinearity tolerance | VIF  |
|-----------------------------|-------------------------------------|--------|--------|------------------------|------|
| Constant                   | ---                                 | -0.071 | 0.843  | ---                    | ---  |
| Pump diameter (D)          | 0.820                               | 20.400 | 0.022  | 0.680                  | 1.470|
| Motor active power (Py)    | -0.611                              | -14.553| 0.000  | 0.624                  | 1.602|
| Submergence (H)            | -0.603                              | -16.003| 0.002  | 0.774                  | 1.292|
| Stroke times (n)           | 0.520                               | 13.249 | 0.000  | 0.715                  | 1.399|
| Pump efficiency (\(\eta_d\)) | 0.441                               | 12.683 | 0.000  | 0.907                  | 1.102|
| Stroke (S)                 | 0.268                               | 7.662  | 0.015  | 0.900                  | 1.111|

The parameters in the standardized regression equation are the data after standardized processing.

Standardized regression equation:

$$y_s = 0.82 \times D - 0.611 \times P_y - 0.603 \times H + 0.520 \times n + 0.441 \times \eta_d + 0.268 \times S$$  \(3\)

Where, \(y_s\) - Standardized regression equation; \(D\) - pump diameter, mm; \(P_y\) - active power of the motor, kW; \(H\) - Submergence depth, m; \(n\)-frequency of the pump stroke, 1/min; \(\eta_d\) - pump efficiency, %; \(S\)-Pump stroke, m.

The regression equation is used to calculate the system efficiency regression value of 198 wells. Table 3 lists the comparison results between the regression value and the measured value. From the error analysis, it can be seen that the absolute error between the two is about 2.9%, The relative error is close to 4.8%.

### Table 3. Comparison of regression values and measured values.

| System efficiency | Average value | Standard deviation | Absolute error | Relative error |
|-------------------|---------------|--------------------|----------------|---------------|
| Measured value    | 0.2369        | 0.1227             | 0.029          | 0.048         |
| Regression value  | 0.2373        | 0.1091             |                |               |

According to the above test results, it can be proved that the goodness of fit of the regression model is good, the error is within a reasonable interval, and the regression equation is statistically significant. Therefore, the absolute value of the standardized regression coefficient of the model can reflect the degree of influence of various factors on system efficiency.

### 4. Evaluation of Regression Results

It can be seen from the regression results that the factors that have a greater impact on the system efficiency of the block are mainly located in the downhole part. The largest impact is the pump diameter, followed by the submergence and pump efficiency; the stroke and the stroke rate of the swabbing parameters have an impact on the system efficiency. The effect of the motor is also more obvious; the active power of the motor is the sum of the useful work and the lost work of the entire system, and its impact on the system efficiency is second only to the pump diameter.

Therefore, the focus of improvement for this block should be focused on the improvement of downhole efficiency, giving priority to the adjustment of wells with unreasonable submergence settings and low pumping efficiency in the block, checking the pumping parameters of each well, and preventing oil wells from overwhelming due to pumping parameters. Problems such as insufficient liquid supply, empty pumping, and high pump inspection rate have occurred to ensure the balance of oil supply and production. The second is to improve the technical management level of the block, reduce the loss of work in the entire pumping system, and increase the energy conversion rate of each node.
According to the analysis results, within half a year, 103 wells in the block were inspected and improved in terms of submergence, swabbing parameters, and technical management level and the average system efficiency of the pumping wells in the block was optimized. This is an increase of about 3.7% compared to before optimization. It can be seen from the experimental results that the multiple regression method has a good predictive ability for the system efficiency of oilfield blocks, and the algorithm has considerable stability. Compared with the traditional test method, the multiple regression method is more accurate. The main body of the model establishment relies on the historical data of the tested block. The test data of the current month is only a supplement, which can effectively reduce the monthly test workload of each block and reduce the production of oil fields. It is burdensome to run, and the algorithm of the multiple regression method is relatively fast and efficient, easy to realize computer programming, and easy to integrate into the digital management of the oil field. It has good application prospects and is worthy of further research and promotion.

5. Conclusion

(1) The regression model established on the theoretical basis of multiple regression can quantitatively describe the degree of influence of various factors on system efficiency, and provide theoretical support for the analysis and optimization of pumping well efficiency.

(2) Guide the adjustment and optimization of pumping wells according to the degree of influence of various factors, arrange the implementation priority of various improvement measures, provide a scientific basis for the targeted management of oil wells, and improve the comprehensive management level of the oil field.

(3) For different oilfield blocks, the key factors affecting the efficiency of pumping wells are often different. The original concept should be changed, the relative importance of various factors for efficiency improvement should be studied, and the main contradictions that restrict the efficiency of oil wells should be grasped, and measures should be improved. Transformation effect benefit.

References

[1] Long Y., Zhang W., Wu W.Q. Application method of node management in improving pumping unit system efficiency. China Equipment Engineering, 2020, 4:14-15.

[2] Zhang C.Y., Gao X.H., Wang L.H., et al. Energy saving research of hydraulic regulating pumping units. Manufacturing automation, 2020, 5:106-110.

[3] Wang Y., Ma S.J., He X.L., et al. Study on the selection of energy saving motor for pumping units in Tahe Oilfield. Scientific and technological horizon, 2020, 13: 62-64.

[4] Feng W. Application status of energy-saving motors for pumping units. Chemical engineering and equipment, 2020, 4: 163-164.

[5] Dong S.T., Liu X.Z, Yang Y.G. Energy saving control of beam pumping unit. Petrochemical technology, 2020, 3: 333.

[6] Zhang W.P., Peng L.L. Research on dynamic power matching of hydraulic pumping unit. Machine tool and hydraulic, 2020, 16: 84-89.

[7] Song W., Feng Z.M., Zhang D.S., Yu N. Optimization modeling and energy saving mechanism of variable speed drive of beam pumping unit. Chemical machinery. 2020, 03: 348-354.

[8] Li K., Gao X. W., Tian Z. D., et al. Using the curve moment and the PSO-SVM method to diagnose downhole conditions of a sucker rod pumping unit. Petroleum Science, 2013, 10(1): 73-80.

[9] LI K., GAO X. W., YANG W., et al. Multiple fault diagnosis of down-hole conditions of sucker-rod pumping wells based on Freeman chain code and DCA. Petroleum Science, 2013, 10(3): 347-360.

[10] Ren W.J., Zhao Y.J., Wang T.R., et al. Research on pump-jack fault diagnosis method based on biogeography-based optimization algorithm. Journal of System Simulation, 2014, 26(6): 1243-1250, 1273.

[11] Liang H. Hierarchical fault diagnosis of rod pumping system based on fault distinguishing. Journal of Southwest Petroleum University, 2015, 37(1): 165-171.