Fault diagnosis of electric submersible pump tubing string leakage

Jiali Yang^{1,†}, Wei Li^{2,3,†}, Jiarui Chen¹ and Sheng Li^{1,*}

¹ School of Mathematics and Computer, Guangdong Ocean University, Zhanjiang Guangdong, 524088, China

² Southern Marine Science and Engineering Guangdong Laboratory (Zhanjiang), Zhanjiang, Guangdong, 524088, China

³ CNOOC China Limited ZhanJiang Branch, Zhanjiang, Guangdong, 524057, China

[†] These authors contributed equally to this work.

Abstract. With the rapid development of the offshore oil industry, electric submersible pumps have become more and more important. They are the main pumping equipment in oil well production and have huge advantages in terms of displacement and production costs. Due to the complex structure of the electric submersible pump, the bad working environment will cause failures. The failure of tubing string leakage is a common failure in oilfields; tubing string leakage of the electric submersible pump will reduce oil production. In order to reduce the economic loss of oil well production. This paper uses PCA and Mahalanobis distance to make the tubing Fault diagnosis of leakage. The feasibility of the algorithm is verified through experiments. The result shows that it can diagnose the failure time of pipe string leakage in advance and hence help us to reduce the maintenance cost of offshore oilfields.

1 Introduction

Electric submersible centrifugal pump (electric submersible pump) is a kind of multi-stage centrifugal pump that works in the well. It uses tubing to lower the centrifugal pump and the submersible motor into the well. The motor drives the multi-stage centrifugal pump to rotate to generate centrifugal force, which reduces the crude oil in the well oil extraction equipment lifted to the ground. The electric submersible pump oil production process is widely used in high water cut wells, non-self-blown high-yield wells and offshore oil fields due to its simple equipment structure, large displacement, high efficiency and high degree of automation^[1]. Electric submersible pumps can lift crude oil to the surface under conditions of higher temperature and greater depth, but electric submersible pump wells are prone to failures during the production process, which will cause production interruptions and cause serious economic losses^[2].

Under normal circumstances, once the electric submersible pump fails, it needs to perform workover operations, which costs a high maintenance fee, which brings a heavy economic burden to the oil field enterprise, and also affects the normal production of the oil field. It mainly relies on manual calculations and expert experience to carry out post-mortem evaluation and forecasting. There are problems such as heavy workload, poor timeliness and low forecast accuracy, which cannot meet the needs of modern oilfield production. Therefore,

it is necessary to conduct in-depth research on the fault diagnosis technology of electric submersible pumps, apply new methods and new technologies to fault diagnosis of electric submersible pumps, ensure that the electric submersible pumps work better, and create better economic benefits for enterprises.

At present, Peng Long et al[3]proposes a fault diagnosis of the broken shaft of a submersible electric pump based on principal component analysis, and concludes that the principal component analysis algorithm has the potential to be used to detect dynamic changes in data. Matthews et al[4]. used random forest to classify and identify 7 types of oil well faults with an accuracy of 94%. Lu Li et al[5] used electric submersible pump wellhead electrical parameters and production parameters to extract the parameter characteristics of different faults through digital signal processing, and analyzed 9 types of faults. Data analysis is carried out for typical fault conditions. Zhang Panlong et al[6] proposed an algorithm based on feature extraction to diagnose electric submersible pump system faults, experiments show that the algorithm can achieve the expected goal well.

This paper uses principal component analysis and Mahalanobis distance algorithm to study the daily production data of electric submersible pumps. First, the data is preprocessed, then the principal component analysis algorithm is used to extract the principal components, and finally the electric submersible pump data is faulted through Mahalanobis distance. Diagnosis and analysis of experimental results show that this set of algorithms can diagnose pipe string leakage failures in

*Corresponding author's e-mail: lish_ls@gdou.edu.cn and lish_ls@sina.com

advance and achieve the benefit of reducing economic losses.

2 Principal component analysis

PCA(principal component analysis) is a multivariate statistical method. It is one of the most commonly used dimensionality reduction methods. A set of potentially correlated variable data is transformed into a set of linearly uncorrelated variables through orthogonal transformation. The transformed variable is called main ingredient^[7]. That is, given a set of variables $x_1, x_2, \dots, x_n$, through a linear transformation, transformed into a set of uncorrelated variables $Y_1, Y_2, \dots, Y_n$, in this transformation, the total variance of the variables (the sum of variances) is kept unchanged, and the maximum variance is called the first principal component. At the same time, Y_1 has the largest variance, called the first principal component; Y_2 has the second largest variance, called the second principal component. By analogy, once there were n variables, you can convert n principal components, and find out that most of the variance variables that can reflect the original n variables are $p(p \leq n)$.

First calculate the correlation between variables $\gamma_{ij} (i, j = 1, 2, \dots, n)$, where $\gamma_{ij} = \gamma_{ji}$.

$$\gamma_{ij} = \frac{\sum_{k=1}^m (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\left(\sum_{k=1}^m (x_{ki} - \bar{x}_i)^2 \sum_{k=1}^m (x_{kj} - \bar{x}_j)^2 \right)^{1/2}}. \quad (1)$$

Set a correlation matrix R as follows,

$$R = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m1} & \gamma_{m2} & \cdots & \gamma_{mn} \end{bmatrix}. \quad (2)$$

Calculate the eigenvalue λ of the matrix, and obtain the variance contribution rate and the cumulative variance contribution rate as the formula.

$$F_i = \frac{\lambda_i}{\sum_{k=1}^n \lambda_k}, (i = 1, 2, \dots, n), \quad (3)$$

$$S_i = \frac{\sum_{k=1}^i \lambda_k}{\sum_{k=1}^n \lambda_k}, (i = 1, 2, \dots, n). \quad (4)$$

Finally, the principal component is:

$$Y = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ t_{21} & t_{22} & \cdots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ t_{m1} & t_{m2} & \cdots & t_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}. \quad (5)$$

3 Mahalanobis distance

The Mahalanobis distance was developed by Indian statistician P. C. Mahalanobis to measure the covariance distance of data. It is an effective method to calculate the similarity of two unknown samples. Compared with Euclidean distance, Mahalanobis distance has more advantages and is not affected by dimension, that is, the Mahalanobis distance between two points has nothing to do with the original data, and the calculation speed is fast.

Moreover, the Mahalanobis distance can consider the relation between various characteristics and eliminate the correlation interference between variables. To sum up, Mahalanobis distance can easily measure the distance between the observed sample and the known sample set, so it is suitable for fault diagnosis^[8]. $Y = (Y_1, Y_2, \dots, Y_p)$ is the p -dimensional principal component obtained through data principal component analysis, where the mean vector is μ , the covariance matrix is Σ , and $y = (y_1, y_2, \dots, y_m)$ is a random sample obtained by principal component analysis of Y . Then the Mahalanobis distance D_M from Y to the mean vector μ is:

$$D_M(Y, \mu) = \left((Y - \mu)^T \Sigma^{-1} (Y - \mu) \right)^{1/2}, \quad (6)$$

where P is the dimension of the sample, M is the size of the sample.

The μ value is calculated as followed:

$$\mu = \frac{1}{p} \sum_{i=1}^m y_i, \quad (7)$$

where the covariance matrix is:

$$\Sigma = \frac{1}{p-1} \sum_{i=1}^m (y_i - \mu)(y_i - \mu)'. \quad (8)$$

4 Fault Diagnosis Process

The electric submersible pump data collected by the sensor has many missing and abnormal data, so the data should be cleaned first, and then the data should be standardized. For some attributes, only unique values need to be deleted. Then, the preprocessed data were analyzed by principal component analysis to calculate variance contribution rate and cumulative variance contribution rate, and each principal component. Finally, the data set is divided, the Markov distance and the maximum Markov distance of the training set are calculated, and then the new data is tested to diagnose the time of failure. The specific fault flow is as shown in Figure 1.

Figure 1. Fault diagnosis flow

5 Case study: diagnosis of the ESP Tubing String Leakage

5.1. Data source of the ESP Tubing String Leakage

The data in this article comes from the China National Offshore Oil Development and Production Database. The sensor records the daily data of electric submersible pumps (one record per day), among which 25 oil wells have recorded string leakage. Through data observation, only the historical data of 14 oil wells can be used. The well IDs are 831352627, 831353047, 831412597, 832012687, 833155477, 834652657, 834952567, 835852627, 861652417, 861652717, 951352447, 951352747, 951412537 and 951412567. There are four different types of data sets for oil well data. They includes:

- Daily status data of oil and gas wells: time, nozzle diameter, production time, bottom hole pressure, wellhead temperature, oil pressure, casing pressure, surface casing pressure, technical casing pressure, pump current, pump voltage, pump frequency, pump inlet temperature, moisture content.
- Daily measurement data of oil and gas wells: converted daily fluid production, converted daily oil production, converted daily water production, converted daily gas production, test fluid volume, test oil volume, test water volume, test gas volume, wellhead temperature, oil body density, gas-oil ratio, water content, separator pressure, separator temperature.
- Daily distribution data of oil and gas wells: daily fluid production, daily oil production, daily water production, daily gas production, gas-oil ratio, oil-gas ratio, water cut, water-gas ratio.
- Fault information data: the data information contains the specific time when the string leakage occurred in each oil well.

5.2. Data preprocessing

5.2.1. *Data cleaning.* Through data analysis, data cleaning is divided into the following steps:

- Daily measurement data of oil and gas wells are recorded multiple times a day at some time points, and one piece of data should be retained.
- The daily measurement data of oil and gas wells is not recorded every day but is recorded irregularly (sometimes once every two or three days, sometimes every five or six days), so linear interpolation of the measurement data is required.
- The daily status data of oil and gas wells, the daily metering data of oil and gas wells, and the daily distribution data of oil and gas wells have been merged at the same point in time.
- Delete the samples whose production time is 0, this kind of data will lead to fault misjudgment.
- Delete variables with a single value, such as nozzle diameter, surface casing pressure, technical casing pressure, etc. These variables have no effect on the data.

5.2.2. *Data transformation.* Data standardization is performed on the cleaned electric submersible pump data. Because the difference between different variables is large, the calculated relationship coefficients will differ greatly, and the optimal solution of the model cannot be quickly found in the optimization process. This paper uses the maximum-minimum normalization method to map each variable to the interval [0,1], as in equation (9).

$$x'_{ij} = \frac{x_{ij} - x_{\min}}{x_{\max} - x_{\min}}, \begin{cases} i = 1, 2, \dots, n \\ j = 1, 2, \dots, m \end{cases} \quad (9)$$

where $x_{\min}$ is the minimum value of x_{ij} , $x_{\max}$ is the maximum value of x_{ij} , m is the data sample size,

and n is the characteristic dimension of the data.

5.3. Principal component analysis

According to the data variables obtained by data preprocessing, input the principal component analysis model. According to the decreasing order of the variance of the principal components, the different principal components are sorted. Taking the well ID=831353047 as an example, through data preprocessing, an oil well has 18 parameters including wellhead temperature, pump current, pump voltage, oil pressure, daily gas production, daily liquid production, daily oil production, daily water production, gas-oil ratio, oil-gas ratio, water content, water-gas ratio, converted daily fluid production, converted daily oil production, converted daily water production, test fluid volume, test oil volume, test water volume. Through principal component analysis, the variance of the original input parameters with more than 99% of the first six principal components can be observed, as shown in Figure 2. Among them, the first principal component and the second principal component have the highest variance, including 81% of the variance of the original data.

Figure 2. Cumulative variance contribution rate of well 831353047 by principal components

Figure 3 shows the distribution of the two-dimensional scatter plot samples of principal component 1 and principal component 2. It can be seen that a large number

Figure 4. Mahalanobis distance

of data samples are distributed in the lower left corner (normal samples), and only a few data samples are distributed on the right and far from the left (faulty samples). Therefore, the two-dimensional curves of principal component 1 and principal component 2 are very important for real-time monitoring of the performance of the electric submersible pump. If the sample starts to deviate from the concentrated area, it can send an early warning of potential failure to the field engineer.

Figure 3. Well 831353047 scores plot of Principal Component 1 and Principal Component 2

5.4. Mahalanobis distance fault diagnosis

Extract more than 99% of the principal components of the original data. Well ID=831353047 as an example, extract the first 6 principal components to construct a data set, use the Mahalanobis distance model to train the data, calculate the Mahalanobis distance of each sample, and get Maximum Mahalanobis distance, and use this data value as the failure threshold. Through data visualization, the calculated threshold for marking anomalies is about 6.74, as shown in Figure 4.

Through the Mahalanobis distance of the training set and the test set, and the calculated threshold for marking anomalies. The next step is to compare the Mahalanobis distance of the data set with the threshold. When it is greater than the threshold, it is marked as abnormal. As shown in Figure 5, it can be seen that the oil well was abnormal after 2016/8/15, but the actual failure the date was 2016/8/22.

Figure 5. Predict date of failure

Similarly, taking the well ID=861652417 as an example, the threshold for marking anomalies is about 7.96. It was detected that the oil well was abnormal on

2016/3/10, and the actual failure date was 2016/3/16, as shown in the Figure 6 and Figure 7.

Figure 6. Mahalanobis distance

Table 1 shows the comparison between the calculation of the abnormal threshold value and the prediction of the string leakage of the electric submersible pump using the Mahalanobis distance model and the actual time. The analysis in Table 1 shows that the pipe string leakage time

Figure 7. Predict date of failure

predicted by the Mahalanobis distance diagnostic model is earlier than the actual pipe string leakage time. Therefore, the Mahalanobis distance diagnosis model has excellent accuracy in predicting the leakage time of the electric submersible pump string.

Table 1. Mahalanobis distance threshold and failure prediction date and true date for each well

Well ID.	MD Threshold	Predict date of failure	Actual date of failure
831352627	7.96	2017/3/14	2017/3/22
831353047	6.74	2016/8/15	2016/8/22
831412597	6.32	2017/7/6	2017/7/13
832012687	9.14	2019/11/18	2019/12/18
833155477	10.81	2019/7/13	2019/8/29
834652657	7.35	2019/4/26	2019/5/7
834952567	8.43	2019/4/17	2019/7/12
836452927	10.16	2019/5/9	2019/5/17
861652417	7.96	2016/3/10	2016/3/16
861652717	7.88	2018/9/12	2018/10/2
951352447	8.32	2018/12/9	2018/12/23
951352747	7.87	2015/7/28	2015/8/15
951412537	7.68	2016/5/9	2016/6/1
951412567	7.96	2018/3/13	2018/3/14

6 Conclusions

This paper proposes a fault diagnosis model based on principal component analysis and Mahalanobis distance to detect the leakage fault of the electric submersible pump string in advance. Firstly, the principal component 1 and principal component 2 can be used to judge the normal sample and the abnormal sample, and then the Mahalanobis distance can be used to diagnose the time when the fault occurs, which reduces the economic loss of oil well production.

Acknowledgements

This work was supported by the Fund of Southern Marine Science and Engineering Guangdong Laboratory (Zhanjiang) (ZJW-2019-04) and the Natural Science Foundation of Guangdong Province (2018A030307062).

References

1. Xie, W., Chen, L., Wei, H.G., et al. (2017) "Multi-Split VVVF" System for Electric Submersible Pump on Extraction Flat Roof on the Sea. In: Proceedings of the 2016 6th International Conference on

Advanced Design and Manufacturing Engineering (ICADME 2017). Zhuhai. pp. 506-511.

2. Reges, G., Fontana, M., Ribeiro, Marcos., et al. (2020) Electric submersible pump vibration analysis under several operational conditions for vibration fault differential diagnosis. *J. Sci. Ocean Engineering.*, pp. 108249-.
3. Peng, L., Han, G.Q., Pagou, A.I., Shu, J. (2020) Electric submersible pump broken shaft fault diagnosis based on principal component analysis. *J. Sci. Journal of Petroleum Science and Engineering.*, 191.
4. Marins, M.A., Barros, B.D., Santos, I.H., Barrionuevo, D.C., et al. (2020) Fault detection and classification in oil wells and production/service lines using random forest. *J. Sci. Journal of Petroleum Science and Engineering.*, 197.
5. Li, L., Hua, C.Q., Xu, X. (2018) Condition monitoring and fault diagnosis of electric submersible pump based on wellhead electrical parameters and production parameters. *J. Esci. Systems Science & Control Engineering.*, 6(3): 253-261.
6. Zhang, P.L., Chen, T.K., Wang, G.C., Peng, C.Z. (2017) Ocean Economy and Fault Diagnosis of Electric Submersible Pump applied in Floating platform. *J. Esci. International Journal of e-Navigation and Maritime Economy.*, 6: 37-43.
7. Konopleva, L., Il'yasov, K.A., Teo, S.J., et al. (2021) Robust intra-individual estimation of structural connectivity by Principal Component Analysis. *J. Sci. NeuroImage.*, 226.
8. Ji, H.Q. (2020) Statistics Mahalanobis distance for incipient sensor fault detection and diagnosis. *J. Sci. Chemical Engineering Science.*, pp. 116233-.