Predictive Approach to Perform Fault Detection in Electrical Submersible Pump Systems

Long Peng, Guoqing Han,* Xianfu Sui, Arnold Landjobo Pagou, Liying Zhu, and Jin Shu

ABSTRACT: It has been a great challenge for the oil and gas industry to timely identify any electrical submersible pump (ESP) abnormal performance to avoid ESP failure. Given the high cost of the ESP failure, more and more real-time surveillance systems are applied to monitor ESP performance to generate alarms in the case of failures. This paper presents a robust principal component analysis (PCA) model to perform fault detection for ESP systems continuously. A three-dimensional plot of scores of principal components was used to observe different patterns during the stable and failure periods. 47 cases of actual failure events and 40 cases of stable operating events were tested on the robust PCA model to generate prediction results. The testing results demonstrate that the robust PCA model has managed to identify 20 failure events before the actual failure time out of the 47 failure cases and has successfully distinguished all the 40 stable operating wells. This study has concluded that PCA has the potential to be used as a monitoring platform to recognize dynamic change and therefore to predict the developing failures in the ESP system.

1. INTRODUCTION

The electrical submersible pumps (ESPs) are one of the most widely employed artificial lift pumping techniques, accounting for over 60% of the total global oil production. More than 90% of the offshore oil wells require ESPs for pumping oil and other fluids to the surface. However, it is often observed that the ESP system reaches the point of production disruptions that will result in an economic loss of hundreds of millions of oil barrels.

ESP failures are common occurrences in the oil industry, and the ESP maintenance is an incredibly expensive issue for the oil operating company. A failure happens when key parameters start deviating from safe operating thresholds and an ESP stops working forever. Ammeter charts have been the earliest and simplest ESP diagnostic tool in the past few years and are still very common but require substantial human resources for frequent wellsite visits to make necessary adjustments. This monitoring method measures the motor current and records it as a function of time on a circular chart. Ammeter charts are limited to detect complete ESP failures because the change in power consumption is only linked to voltage fluctuations, fluid density, and flow rate. In recent years, nodal analysis has been applied to perform ESP system modeling and determine some common ESP failures such as broken shaft, blockage at pump intake, and so forth. However, nodal analysis is also limited in the ability to analyze performance and identify optimal operating conditions in real time. The oil and gas industry has been employing statistical modeling and machine learning techniques to detect patterns that allow for production optimization, failure prevention, and real-time event detection. Guo et al. used the support vector method to predict anomalous operation. Andrade Marin et al. analyzed random forest to obtain a high value of accuracy and recall of ESP failure prediction in 165 cases. Bhardwaj et al. applied the principal component analysis (PCA) and gradient boosting algorithm (XGBoost) for anomaly detection and failure prediction in six cases and two samples. Developments in predicting the potential ESP failures with machine learning technology, if successful and implemented, will improve the quality of forecasting and take timely measures.

The ESP was previously monitored by traditional means, such as collecting ESP-related parameters, visiting the physical well site or evaluating the manual amperage charts (Amps), hindering the ESP from reaching optimum performance. In an effort to better understand ESP behavior, numerous...
improvement has been made on ESP sensors, supervisory control, and data acquisition systems and surface remote terminal units over the past few decades.11 With the fast development of the digital oil field, more and more machine learning and data-driven models are applied to perform fault detection and predict impending failure for specific ESP operation systems.12 The full realization of a data-driven model lies beyond the availability of continuous data acquisition. Therefore, the key to perform fault detection on the ESP can be better defined as a problem to build an accurate data-driven model that describes the ESP system dynamics.

This presented work is concerned with the application of the PCA in three dimensions to provide an intuitive way of monitoring the ESP system. The least squares method is used to determine the safety threshold of the ESP stable operation system. With this method, a predictive model is constructed to perform fault detection in real time. The objective of this paper is to evaluate this predictive approach to detect developing ESP failure.

2. APPLICATION OF PCA

PCA is widely used as a preprocessing method for data visualization and dimensionality reduction.13 Jackson14 defined PCA as an unsupervised dimensionality reduction technique that converts a set of dependent variables into a set of linearly uncorrelated variables called “principal component”. As known to all, ESP data are highly correlated. For example, increasing the wellhead pressure will cause an increase in discharge pressure and intake pressure, eventually increasing the motor pressure. A PCA model takes advantage of this interdependence to reduce the numbers of ESP original data and to create a new principal component space (PCs). This reduction allows the evaluation of the original ESP system only by several principal components. Abdelaziz et al.11 built different PCA models for each installation to conclude that PCA had the potential to be used as a tool to identify dynamic changes in the ESP system. Gupta et al.15 proposed the PCA predictive model, diagnostic model, and prescriptive model to detect impending ESP failures.

The PCA basic model was represented as

\[
X = TP^T + E
\]

(1)

where \(X\) is the input matrix \((n \times p)\); \(T\) is the score matrix \((n \times k)\); \(P\) is the loading matrix \((p \times k)\); \(E\) is the residual matrix \((n \times p)\); \(n\) is the number of time steps;15 \(p\) is the number of original parameters;15 and \(k\) is the number of principal components.15

The first principal component contains the highest variance which means that the first principal component captures the most information. The second principal component captures the next highest variance. Each successive principal component is orthogonal to its preceding principal components. Several principal components are constructed to create new PCs to evaluate the original ESP system.

Gupta et al.17 proposed that the PCA model contained two steps: model construction and model prediction. The ESP data obtained from the stable operating zone \((X_{\text{training}})\) is used to build the PCA model. The loading matrix \(P_{\text{training}}\) for the stable operating zone will be stored for prediction at a later stage. Once a robust PCA model is constructed, any new input dataset can be selected as a testing dataset \((X_{\text{testing}})\). The testing dataset may be a stable operation zone or an unstable failure operation zone. The testing dataset is fed to the robust PCA model as a new score matrix \((T_{\text{testing}})\) is obtained. Patterns corresponding to stable periods or failure periods will be compared against patterns trained by the PCA model to draw conclusions for making predictions.18 By this means, any historical data in the ESP system can be detected as a stable operating zone or a failure zone. Figure 1 summarizes the comments above.

3. CONSTRUCTION OF A ROBUST PCA MODEL

3.1. Selection of ESP Variable Parameters. Much research has been done about the application of the PCA model in the XX oilfield XX block in China. The fast development of ESP sensors and data acquisition systems makes it possible for the ESP system to record dynamic data and historical data continuously. Dynamic production data were collected at a frequency of 20 min. Dynamic data recorded contain variables of casing pressure, intake pressure, discharge pressure, flow tubing pressure, intake temperature, motor current, motor leakage current, motor power, motor temperature, motor vibration, motor voltage, tubing choke,
and VFD frequency. Thirteen variable parameters were selected as training datasets or testing datasets. Historical data recorded include historical well events and well failure reports. The information when a trip or failure occurred in each ESP well was recorded in the historical data.

3.2. Comparison of the Original System and PCs. A data-driven model using the PCA method was constructed to reduce the dimensionality of the historical stable training dataset containing 13 key variable parameters. 174,538 historical stable operating data according to different production time of 47 wells were used to construct a robust PCA model. Three principal components captured more than 70% information of the original training dataset to create a new PCs. Taking well #X1 as an example, Figure 2 showed the original stable operating ESP data recorded containing 13 variable parameters shown. Three principal components were extracted to form the new PCs to evaluate the original system shown in Figure 3. It could be seen from the comparison of Figures 2 and 3 that the trends of the original 13 variable parameters were retained in the form of three principal components.
components according to the robust PCA model. Each principal component was determined by calculating the singular value decomposition of the selected datasets. Moreover, a special eigenvalue and eigenvector corresponding to a unique principal component would be determined. Each principal component explained the different percentage of the total variance, and the top-ranked principal components had larger eigenvalues. Only three principal components were used to evaluate the original ESP system, making the ESP system fault detection much easier. Three principal components on a three-dimensional plot provided an intuitive way of monitoring the ESP system.

3.3. Scores of Three Principal Components. A three-dimensional plot of scores of principal component 1, principal component 2, and principal component 3 was used to observe different patterns during stable and failure periods. The three principal components were extracted from the original training dataset. Figure 4 represented the scores plot of three principal components; it was viewed that the historical stable operating zone forms a large cluster. The X-axis represented the first principal component, the Y-axis represented the second principal component, and the Z-axis represented the third principal component. The new PCs also contained 174,537 data was determined based on the least square method, and the radius was the maximum distance from all stable operating data to the center of the sphere. There was only one sphere for all 47 wells. The radius of the sphere was the failure threshold of all the stable operating data of 47 wells. By this means, when the scores for the testing dataset corresponded to a stable operation period, the cluster formed inside the sphere, similar to the training dataset. However, when the scores for the testing dataset corresponded to a trip or failure, the cluster started deviating away from the training dataset and was finally located outside the sphere.

4. MODEL TESTING AND RESULTS

4.1. Testing of Potential Failure Wells. The historical stable operating data of 47 wells were used to construct the robust PCA model, and the unstable operating period of these wells served as a training dataset.

According to the ESP failure nomenclature standard and failure definitions by Alhanati et al., this paper represented the different types of ESP failures from the collected datasets shown in Figure 5. There were mainly 6 different types of failures, and the causes for each type of failure were also different.

![Figure 5. Different types of ESP failures in the XX oilfield.](image)

When the ESP wells were put into production, the ESP wells would start from the stable region to the failure region as the time step increased. The unstable operating region contained stable operating data before a trip/failure and unstable behavior ultimately leading to a trip/failure. 47 potential failure wells were tested by the robust PCA model. For each well of the testing dataset, the robust PCA model generated one prediction result. Taking four wells as examples, Figure 6 was an intuitive visual of ESP well monitoring for a historical failure event. It was observed that the stable operating zone of the testing dataset formed clusters inside the training sphere. In the meanwhile, a failure zone started deviating away from the sphere while a failure or trip took place.

4.2. Testing of Stable Operating Wells. The robust PCA model was tested against a total of 40 stable operating wells. These 40 wells were also from the XX oilfield XX block in China. However, the difference was that these 40 wells were the daily report data rather than real-time data. Similarly, these 40 wells were tested by the robust PCA model to generate the prediction results. Taking 4 wells as examples, Figure 7 represented three-dimensional plots of scores for ESP fault detection on the stable operating wells. The analysis of Figure 7 showed that the stable operating daily report data formed clusters inside the sphere, which meant that these four wells did not have any abnormal behavior during the operating periods.

4.3. Testing Results and Discussion. A confusion matrix was used to summarize the performance of the robust PCA model, showing what the prediction results were getting right and what types of errors the prediction results were making. There were four possible binary classification outcomes in a confusion matrix.
The robust PCA model was tested on 47 cases of actual failure events and 40 cases of stable operating events to generate prediction results. Figure 8 summarized the matrix results in a confusion matrix.

It could be drawn from Figure 8 that out of the 47 ESP failure events, the robust PCA model managed to detect 20 failure wells before the actual failure date but failed to identify the remaining 27 impending failure events. For the stable operating well testing results, the robust PCA model had successfully predicted all the 40 nonfailure cases. It was observed that the robust PCA model in three-dimension presented 42.5% accuracy for predicting developing ESP failure events before the actual failure date. The appearance of these prediction results came for some reasons. The main reason was that the cause for each type of failure was different. For instance, there were no obvious abnormal behaviors before the pump broken shaft, so the robust PCA model had failed to detect the failure events before the actual failure data but had identified when these failure events exactly happened. In the meanwhile, the robust PCA model had managed to detect the motor failures before the actual failure date because the motor current, motor voltage, or motor temperature performed abnormally before the actual failure time captured by the robust PCA model. Secondly, the sampling of 20 min was too long and much dynamic changes of different variable parameters leading to a failure had been missed. Lastly, the robust PCA model in three principal components only extracted 70% information of the original ESP operation.

Figure 6. ESP fault detection on potential failure wells.
system, and 30% of the variance was discharged, resulting in part of fault behavior missed. Future work must be focused on these ways to improve the prediction results in detecting developing failure events.

Overall, the PCA method serves as a real-time platform to automatically perform fault detection of any ESP system. This detection solution can be of significant importance in predicting the impending ESP failure in real-time. The three-dimensional plot of scores of three principal components creates a visual medium to continuously monitor the ESP operation for keeping users apprised of what is happening with the monitored system. Moreover, oilfield engineers will be reminded to take corrective action or timely solve the issues to minimize downtime if the clusters are far away from the sphere.

5. RESULTS AND DISCUSSION

This paper develops a robust PCA model to perform fault detection for the ESP systems. 47 cases of actual ESP failure events and 40 cases of stable operating events are tested by the PCA model. It can be concluded that the PCA technique has the potential to monitor dynamic change and predict the impending failure in the ESP system. The following conclusions has been developed from this work:

(1) A three-dimensional plot of scores of principal components is used to observe different patterns of stable and failure period in real time. The dataset
corresponding to a stable operation period forms the clusters inside the sphere and the dataset corresponding to a trip or failure starts deviating away from the sphere.

(2) Out of the 47 ESP failure events, the PCA model has managed to identify 20 impending failure events and mispredict 27 failure before the actual failure time.

(3) The PCA model has successfully detected all the 40 stable operating wells out of the 40 nonfailure cases.

(4) The PCA technique can serve as a real-time platform to perform fault detection for the ESP system and as an unsupervised machine learning technique to predict the impending ESP failures.

# AUTHOR INFORMATION

**Corresponding Author**

Guoqing Han — State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum (Beijing), Beijing 102249, China; Shaanxi Key Laboratory of Carbon Dioxide Sequestration and Enhanced Oil Recovery, Xi’an 710075, China; Email: CUP_706ac@163.com

**Authors**

Long Peng — State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum (Beijing), Beijing 102249, China; Shaanxi Key Laboratory of Carbon Dioxide Sequestration and Enhanced Oil Recovery, Xi’an 710075, China

Xianfu Sui — China National Offshore Oil Corporation (CNOOC) Research Institute, Beijing 100027, China

Arnold Landjobo Pagou — State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum (Beijing), Beijing 102249, China; Shaanxi Key Laboratory of Carbon Dioxide Sequestration and Enhanced Oil Recovery, Xi’an 710075, China

Liyang Zhu — State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum (Beijing), Beijing 102249, China; Shaanxi Key Laboratory of Carbon Dioxide Sequestration and Enhanced Oil Recovery, Xi’an 710075, China

Jin Shu — State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum (Beijing), Beijing 102249, China; Shaanxi Key Laboratory of Carbon Dioxide Sequestration and Enhanced Oil Recovery, Xi’an 710075, China

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.0c05808

---

**Notes**

The authors declare no competing financial interest.

# ACKNOWLEDGMENTS

The authors of this article would like to express their gratitude to the National Natural Science Foundation of China (grant no. 51574256) and Open Foundation of Shaanxi Key Laboratory of Carbon Dioxide Sequestration and Enhanced Oil Recovery (grant no.YJSYZX20SKF0011). The authors are grateful to the editor and the reviewers for their critical comments, which were helpful in the preparation of this article.

# REFERENCES

(1) Dunham, C. 27th ESP Workshop: Summary of Presentations; Oilfield Automation Consulting, 2013; p 1. Retrieved from https://www.specs.org/media/files/files/cebfcc3a/2013-ESP-Workshop-Summary-of-Presentations.pdf.

(2) Alamu, O. A.; Pandya, D. A.; Warner, O.; Debacker, I. ESP Data Analytics: Use of Deep Autoencoders for Intelligent Surveillance of Electric Submersible Pumps. Offshore Technology Conference, Houston, Texas, USA, May 2020.

(3) Bates, R.; Cosad, C.; Fielder, L.; Kosmala, A.; Hudson, S.; Romero, G.; Shannugam, V. Taking the pulse of production wells—ESP surveillance. Oilfield Rev. 2004, 16, 16–25.

(4) Gabor, T. Electrical Submersible Pumps Manual: Design, Operations, and Maintenance; Elsevier: Burlington-Massachusetts, 2009.

(5) Awaid, A.; Al-Muqbali, H.; Al-Bimani, A.; Al-Yazeedi, Z.; Al-Sukaity, H.; Al-Harthy, K.; Baillie, A. ESP Well Surveillance using Pattern Recognition Analysis, Oil Wells, Petroleum Development Oman. International Petroleum Technology Conference, Doha, Qatar, January 2014.

(6) Stone, P. Introducing predictive analytics: Opportunities. Digital Energy Conference and Exhibition, Houston, Texas, U.S.A., April 2007.

(7) Guo, D.; Raghavendra, C. S.; Yao, K.-T.; Harding, M.; Anvar, A.; Patel, A. Data driven approach to failure prediction for electrical submersible pump systems. SPE Western Regional Meeting, Garden Grove, California, USA, April 2015.

(8) Andrade Marin, A.; Busaidy, S.; Murad, M.; Balushi, I.; Ryyami, A.; Jahwari, S.; Ghadani, A.; Ferdiansyah, E.; Shukaili, G.; Amri, F. ESP Well and Component Failure Prediction in Advance using Engineered Analytics: A Breakthrough in Minimizing Unscheduled Subsurface Deferments. Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, November 2019.

(9) Bhardwaj, A. S.; Saraf, R.; Nair, G. G.; Vallabhaneni, S. Real-Time Monitoring and Predictive Failure Identification for Electrical Submersible Pumps. Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, November 2019.

(10) Shuwaikhat, H. A.; Ramos, M.; Aifan, A.-R.; Al-Sadah, A. A. Innovative Approach to Prolong ESP Run Life Using Algorithmic Models. Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, November 2017.

(11) Abdelaziz, M.; Lastra, R.; Xiao, J. ESP data analytics: Predicting failures for improved production performance. Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, November 2017.

(12) Jahnke, P. Machine Learning Approaches for Failure Type Detection and Predictive Maintenance; Technische Universität Darmstadt, 2015; p 19.

(13) Raschka, S. Predictive Modeling, Supervised Machine Learning, and Pattern Classification—The Big Picture; Sebastian Raschka, 2014.

(14) Jackson, J. E. A User’s Guide to Principal Components; John Wiley & Sons, 2005; Vol. 587.

(15) Gupta, S.; Saputelli, L.; Nikolau, M. Applying big data analytics to detect, diagnose, and prevent impending failures in electric submersible pumps. SPE Annual Technical Conference and Exhibition, Dubai, UAE, September 2016.
(16) Ratcliff, D.; Gomez, C.; Cetkovic, I.; Madogwe, O. Maximizing Oil Production and Increasing ESP Run Life in a Brownfield Using Real-Time ESP Monitoring and Optimization Software: Rockies Field Case Study. *SPE Annual Technical Conference and Exhibition*, New Orleans, Louisiana, USA, September 2013.

(17) Gupta, S.; Nikolaou, M.; Saputelli, L.; Bravo, C. ESP Health Monitoring KPI: A Real-Time Predictive Analytics Application. *SPE Intelligent Energy International Conference and Exhibition*, Aberdeen, Scotland, UK, September 2016.

(18) Bimani, A.; Kulkarni, R.; Lee, C. Y.; Giuliani, V.; Musallami, M.; Awaad, A.; Kamyani, N.; Kumar, N.; Gala, R. Case Study Toward Digital Oil Field: How the ESP Operation is Changing by Using Automatic Well Models in PDO’s ESP Fields. *SPE Gulf Coast Section Electric Submersible Pumps Symposium*; The Woodlands: Texas, USA, May 2019.

(19) Alhanati, F.; Solanki, S.; Zahacy, T. ESP failures: can we talk the same language? *SPE Gulf Coast Section Electric Submersible Pump Workshop*, Houston, Texas, April 2001.