http://journal.uir.ac.id/index.php/JGEET

E-ISSN : 2541-5794
P-ISSN : 2503-216X

Journal of Geoscience,
Engineering, Environment, and Technology
Vol 10 No 04-2/2025

RESEARCH ARTICLE

Performance Comparison of Isolation Forest and COPOD Algorithms for Anomaly Detection in Electrical Submersible Pumps

Hygiano Paksi Widya Asmara[1], Dara Ayuda Maharsi[1,*] and Rian Rachmanto[2]
[1] Universitas Pertamina, Department of Petroleum Engineering, Jakarta, Indonesia
[2]Subsurface Data Science Department, Pertamina Hulu Kalimantan Timur, Balikpapan, Indonesia

* Corresponding author : dara.maharsi@universitaspertamina.ac.id
Tel.:+62-852-3636-0930
Received: Jul 10, 2025; Accepted: Jan 6, 2026.
DOI: 10.25299/jgeet.2025.10.1.1.23849

Abstract

The Electrical Submersible Pump (ESP) is a crucial technology in enhancing oil production, yet its performance can be compromised by anomalies that lead to operational disruptions and financial losses. Early detection of these anomalies is vital for minimizing risks and optimizing ESP lifespan. This study compares the performance of two machine learning algorithms—Isolation Forest and Copula-Based Outlier Detection (COPOD)—in identifying anomalies in ESP operational data. The study uses both long-term historical data and short-term period data from a well in Field X, focusing on key operational parameters such as amperes, frequency, voltage, discharge pressure, motor temperature, vibration, and gross rate. The results indicate that Isolation Forest outperforms COPOD in detecting anomalies, particularly in the presence of missing data. Short-term data detection yields clearer correlations between anomalies in different features, highlighting its advantage over long-term historical data. The findings underscore the importance of utilizing short-term operational data and demonstrate how anomaly detection algorithms can enhance ESP monitoring for improved performance and cost-efficiency.

Keywords: Electrical Submersible Pump (ESP), Anomaly Detection, Isolation Forest, Copula-Based Outlier Detection (COPOD), Machine Learning

1. Introduction

The Electrical Submersible Pump (ESP) is a widely used technology for enhancing production in oil wells. Currently, ESP is the most extensively installed artificial lift system and significantly contributes to total global oil production. The reliability and efficiency of ESPs are crucial, as they represent items with substantial CAPEX and OPEX costs (Al Maghlouth et al., 2013). Therefore, monitoring ESP performance and detecting anomalies in its operation at an early stage are essential to minimize potential failures and maximize the pump's lifespan.

Monitoring ESPs through conventional methods involves visiting well sites to collect ESP-related parameters and manually evaluating them using ampere (Amps) charts (Shuwaikhat et al., 2017). These charts are typically created weekly or monthly at the well site. Currently, all ESPs are equipped with downhole sensors that measure inlet and outlet pressure, temperature, and vibration. However, with conventional monitoring methods, anomalies or failures in ESPs may go undetected, increasing the risk of production decline. Therefore, routine early anomaly detection efforts are essential to minimize the likelihood of ESP failures.

Anomalies in ESPs can lead to operational disruptions, ESP damage, and significant financial losses. Machine learning-based anomaly detection algorithms offer a solution by learning patterns and identifying deviations that may indicate failures caused by anomalies. The algorithms used in this study are Isolation Forest and Copula-Based Outlier Detection (COPOD). The purpose of this paper is to compare the performance of these two algorithms in detecting anomalies in ESPs and to analyze the anomaly detection results using long-term historical data and short-term period data.

The scope of this study is limited to operational data such as amperes, frequency, voltage, discharge pressure, motor temperature, vibration, and gross rate. Since the study is unsupervised and lacks ground truth data, it does not provide an evaluation or validation of the model's accuracy. Additionally, cross-checks were performed using comments from company data.

2. Literature Study

2.1 Electrical Submersible Pump

The Electrical Submersible Pump (ESP) is one of the most widely used artificial lift systems in the oil and gas industry to enhance oil production when wells face low reservoir pressure issues. ESP operates by deploying a submerged electric pump into the oil well. It consists of several key components: the pump, motor, protector, and electrical cables that deliver power from the surface to the motor. In essence, the pump transfers fluids from the bottom of the well to the surface by converting electrical energy into mechanical energy, which is driven by the motor.

The conventional installation of ESPs typically does not include underground flow meters as part of the equipment configuration. However, many ESPs are now equipped with assemblies containing sensors to monitor their performance and health. Generally, ESP monitoring involves connecting downhole sensors to a SCADA (Supervisory Control and Data Acquisition) system. Currently, numerous Exploration and Production (E&P) companies use web-based monitoring platforms for real-time surveillance of ESP and well data. By acquiring and innovatively processing ESP operational data in real-time, it becomes easier to identify patterns in the data, distinguish between safe and unsafe operations, perform diagnostics to detect issues, implement corrective actions to address problems, and take precise measures (Gupta et al., 2016).

When monitoring ESPs with downhole sensors, several key parameters are measured to ensure proper operation. These include operating current (amperage), which flows through the motor and indicates the load, with changes in flow rate potentially leading to impeller blockage and overheating (Song et al., 2024; Guo et al., 2015). Operating voltage is crucial for motor stability and must align with motor specifications, as overvoltage, undervoltage, and voltage imbalances can disrupt performance (Sherif et al., 2019). The motor speed depends on the alternating current's frequency, and adjusting this frequency is a common solution for controlling speed (Takacs, 2009). Operating frequency itself is essential, as any instability can affect motor speed and pump performance (Sherif et al., 2019). Pump discharge pressure (PDP) and pump intake pressure (PIP) are indicators of ESP efficiency, with high PDP signaling blockages and low PIP suggesting issues with fluid supply or intake obstructions (Sherif et al., 2019; Takacs, 2009). Monitoring motor temperature (TM) is critical as higher current, or voltage increases temperature, and cooling failure can exacerbate this, potentially damaging the motor. Finally, motor vibration is tracked to detect abnormalities, as excessive vibration can indicate component failure and lead to rapid wear or system damage (Takacs, 2009).

2.2 Use of Machine Learning in ESP

In recent years, machine learning (ML) technology has been widely adopted across various industries, including the oil and gas sector. In the petroleum and gas industry, ML is used to enhance efficiency and reduce operational costs. Machine learning is applied to Electrical Submersible Pumps (ESPs) for effective monitoring and diagnosis of their operation by utilizing real-time operational data (**Table 1**). Additionally, the digitalization of oil fields is seen as a new opportunity to further optimize production in the oil and gas industry (Abdalla et al., 2022). Some ML applications used in ESP operations include the following.

ESP Failure Prediction: ML can predict ESP failures by analyzing historical data trends. This prediction is

Table 1. Summary of studies on Machine Learning applications for monitoring and predicting failures in Electrical Submersible Pumps (ESP)

Author, Year	Problem Statement; Objective(s)	Supervised/ Unsupervised	Algorithms	Feature	Findings
Song et al., 2024	Traditional methods using ammeter charts and nodal analysis require manual labor and are difficult for real-time failure diagnosis; Diagnose failures using PCA	Unsupervised	PCA	Vibration, MT, PDP, PIP, Volt	93.3% accuracy in diagnosing ESP failures
Bhardwaj et al., 2019	ESP maintenance requires significant resources for reactive monitoring of multivariate sensor data; Real-time failure identification alarm system	Unsupervised	PCA, XGBoost	MT, Pressure Head, Current, Volt	PCA: >12 mins alarm active for failure, XGBoost: >195 mins alarm active for failure
Guo et al., 2015	Need to predict failures to reduce production loss; Develop a data-driven ML-based approach	Supervised	SVM	Current, Volt, Frequency, Runtime	SVM classification is promising for predicting ESP failures
Sherif et al., 2019	ESP failure causes production delays; Use ML with real-time data to predict ESP failure	Unsupervised	PCA	PIP, Ti, MT, Current, Vibration, PDP	Identified correlations in ESP parameters and successfully detected data anomalies
Gupta et al., 2016	Changes in ESP parameters often lead to failure; Create a platform to monitor ESP trends and patterns	Unsupervised	PCA	WHP, PIP, MT, Current, PDP, Vibration, WC	Innovative real-time processing of ESP operational data successfully detects anomalies
Sneed, 2017	Unpredictable and recurring ESP failures reduce ESP lifespan and increase costs; Identify key failure factors and develop lifespan prediction models	Supervised	Random Forest	MT	Successfully predicted ESP lifespan within 5 days. 90% of model's prediction errors were within 30 days.

enhanced by combining historical data with real-time data from downhole sensors. Various techniques, such as Principal Component Analysis (Song et al., 2024; Bhardwaj et al., 2020), XGBoosting (Abdalla et al., 2022), Support Vector Machine (SVM) (Guo et al., 2015), and Long-Short Term Memory (LSTM) (Liu et al., 2024), are used for failure prediction.

ESP Anomaly Detection: Machine learning can accurately detect anomalies in historical ESP operation data. Detecting anomalies helps prevent failures during ESP operation and optimizes operational parameters for better production. By training models on operational data, systems can identify significant deviations from normal patterns, signaling issues with the ESP (Gupta et al., 2016). Techniques such as SVM (Guo et al., 2015), Principal Component Analysis (PCA) (Sherif et al., 2019), Sparse Autoencoder (Liang et al., 2020), and Isolation Forest (Ding and Fei, 2013) are commonly used for anomaly detection.

ESP Lifespan Optimization: After detecting anomalies and predicting failures, ML can also optimize the operational lifespan of ESPs. Using models like Random Forest (HP Forest), ML can analyze data to predict the optimal operation period for ESPs (Sneed, 2017). Knowing the optimal operating time makes it easier to optimize production processes during that period.

Operational Disturbance Monitoring: At a higher level, ML can monitor ESP operations and reduce disturbances. This can be achieved by integrating IoT (Internet of Things) systems with AI to monitor data from sensors (Jansen van Rensburg, 2018). This monitoring system aims to reduce operational disruptions in ESPs.

2.3 Anomaly Detection Using Machine Learning

Anomaly detection is the process of identifying patterns in data that do not conform to the expected or usual behavior. Patterns in data that deviate from this behavior are called anomalies or outliers, which, although rare, can significantly impact errors or failures (Chandola et al., 2009).

According to Chandola et al. (2009), anomalies can be categorized into three types. Point anomalies are the simplest, where an individual data point does not fit the normal range or pattern. Contextual anomalies are based on specific conditions or contexts and depend on both contextual and behavioral attributes, which influence anomaly detection. Finally, collective anomalies occur when multiple data points show abnormal behavior together, often due to interactions or combinations of several instances.

In anomaly detection, various machine learning approaches—unsupervised, supervised, and semi-supervised learning—are used, each with its own advantages and limitations. Unsupervised learning is applied when data lacks predefined labels, assuming that only a small portion of the data is anomalous (Amer et al., 2013). Supervised learning, on the other hand, requires labeled data to train the model, which helps improve accuracy but is less commonly used due to the difficulty of obtaining labeled data. Semi-supervised learning combines both approaches, using a small amount of labeled data to guide training and leveraging unlabeled data to enhance performance (Zhu et al., 2014). The performance of these algorithms depends on computational speed, minimal parameters, and the ability to handle diverse data distributions, with the ideal model detecting anomalies accurately and quickly in various types of data.

Isolation Forest and COPOD are two popular algorithms used for anomaly detection. Isolation Forest works by isolating anomalies instead of profiling normal data, which makes it efficient for handling large datasets. It isolates outliers by randomly selecting a feature and then randomly selecting a split value between the maximum and minimum values of that feature. The process continues until the data point is isolated. The more easily a point is isolated, the more likely it is an anomaly. COPOD, or Contextual Outlier Detection, focuses on identifying contextual anomalies in data, especially in high-dimensional datasets. It uses the concept of contextual density, where outliers are detected based on the local density of data points compared to their neighbors. COPOD performs well in datasets with complex and varied structures, where the relationships between data points are important. Both algorithms are effective in detecting anomalies but are suitable for different types of data and contexts.

3. Methodology

This study uses a quantitative approach with machine learning to detect anomalies in the parameters of Electrical Submersible Pumps (ESP) based on operational production data. The focus of this research is to develop and evaluate models for detecting anomalies using different algorithms. The main objective is to compare the performance of the Isolation Forest and Copula-Based Outlier Detection (COPOD) algorithms in identifying anomalies. Additionally, the study aims to identify the optimal input parameters for each algorithm in detecting anomalies in ESP parameters and to analyze the results of anomaly detection across historical ESP data and specific ESP installation periods.

Table 2. Hyperparameter setup for Isolation Forest model

Hyperparameter	Value Range	Detail
n_estimators	50-500	Determines how many trees are constructed. More trees improve robustness but increase computation time.
max_samples	auto	This parameter specifies the number of data samples used to train each tree. If set to 'auto', the algorithm uses the entire dataset.
contamination	0.01-0.5	Defines the proportion of anomalies in the data, helping the model identify how many points are anomalous.
bootstrap	False/True	Indicates whether bootstrapping (sampling with replacement) is used when constructing trees, promoting diversity in the forest.

The study uses real historical ESP data obtained from the real-time operation of ESP sensor parameters. The collected data includes various parameters such as Well, Date, Hours Online, Surface Choke, Ampere, Volt, Bottom Hole Pressure, Discharge Pressure, Tubing Head Pressure, Frequency, CSG, Base Sediment Water, Gross Rate, Net Oil, Form Gas, Slip Velocity, Motor Temperature (TM), Vibration, and Comments. All variables are reported using standard SI or industry units (A, V, Hz, °C), and feature names are used consistently throughout the manuscript. The data covers 43 wells in Field X, with Well AA-14 selected for anomaly detection based on both long-term historical data and short-term periods.

The historical data of ESP sensors can be categorized into short-term and long-term data, with long-term data spanning up to 14 years and providing a comprehensive overview of ESP's performance, although it may have gaps due to maintenance or other reasons. Short-term data, covering a period of 2 to 3 years, offers a more recent perspective. The research will focus on seven specific features selected for anomaly detection: Ampere, Frequency, Voltage, Discharge Pressure, Motor Temperature (TM), Vibration, and Gross Rate. These features are based on the monitoring parameters that influence ESP's operation. Time filtering will be applied to detect anomalies in both long-term and short-term data from each well. To handle missing data, the imputation method applied in this study involves removing data entries with missing values.

Missing values were handled by removing incomplete records after temporal filtering. This approach preserves data integrity for algorithms that do not explicitly model temporal dependence. However, we acknowledge that removing rows in long-term time-series data may disrupt temporal continuity. To mitigate this effect, anomaly detection was also performed on short-term continuous periods without missing values, allowing more reliable interpretation of trend- and correlation-based anomalies.

In detecting anomalies using the Isolation Forest algorithm, several parameters can be applied during the model creation process. These parameters are configured within a function called hyperparameters, which help determine the most suitable and accurate settings for the model. **Table 2** shows the hyperparameters that can be used to evaluate and select the best parameters for anomaly detection, with the goal of improving the accuracy of the Isolation Forest algorithm. These hyperparameters can be adjusted based on the available data or customized to suit specific needs in anomaly detection for each feature. Meanwhile, in the COPOD (Copula-based Outlier Detection) algorithm, the primary hyperparameter is "contamination", which defines the expected proportion of outliers in the dataset. This parameter helps the model identify and distinguish anomalous points from the normal ones by adjusting the sensitivity of the anomaly detection process. In this study, the author uses a value range of 0.01 to 0.5 for contamination hyperparameter for COPOD algorithm.

4. Results and Discussion

This section presents the results of anomaly detection using two models: Isolation Forest and Copula-Based Outlier Detection (COPOD). The analysis compares both models based on the entire historical data and specific time periods for each well. This comparison helps evaluate how well each model detects anomalies in large datasets and targeted time ranges. It also examines how anomalies in one feature may affect others. The detected anomalies are cross-checked with comments from the data to verify if they are real anomalies or related to maintenance activities. By comparing the two models, this study highlights the performance of each in detecting anomalies outside the normal data distribution, aiding in the analysis of feature relationships.

4.1 Effect of Contamination Hyperparameter

Before comparing the performance of COPOD and Isolation Forest, this section examines the impact of the contamination hyperparameter of Isolation Forest algorithm on anomaly detection. **Fig. 1** illustrates that higher contamination settings can lead to a false

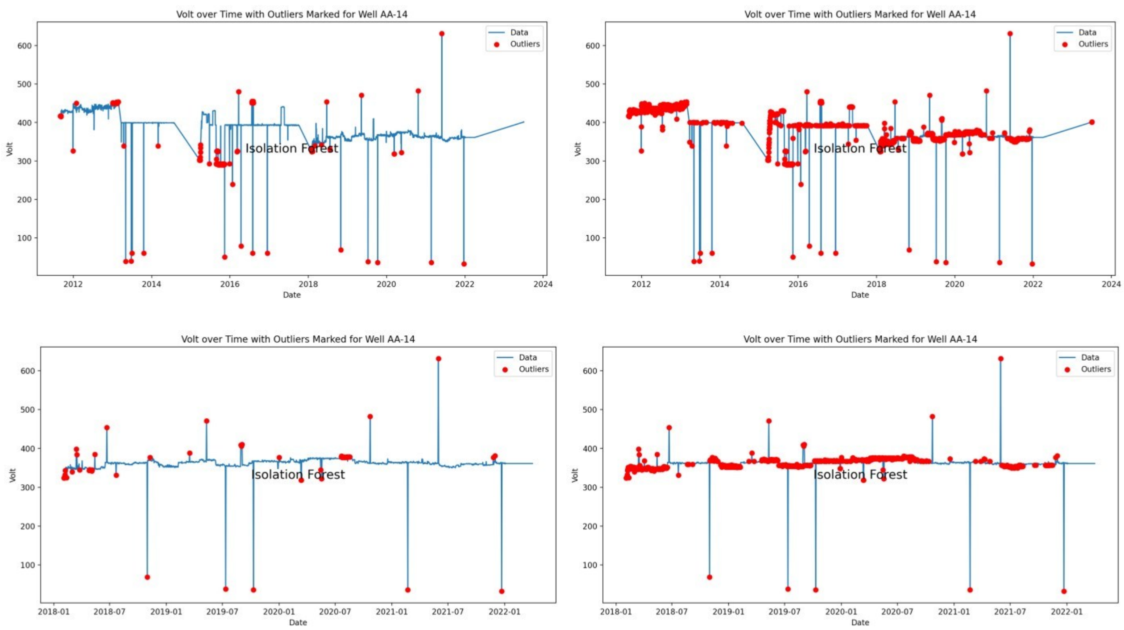

Fig.1 Comparison of anomaly detection in the voltage feature of Well AA-14 long-term (upper) and short-term (lower) history using Isolation Forest contamination 0.05 (left) and 0.5 (right).

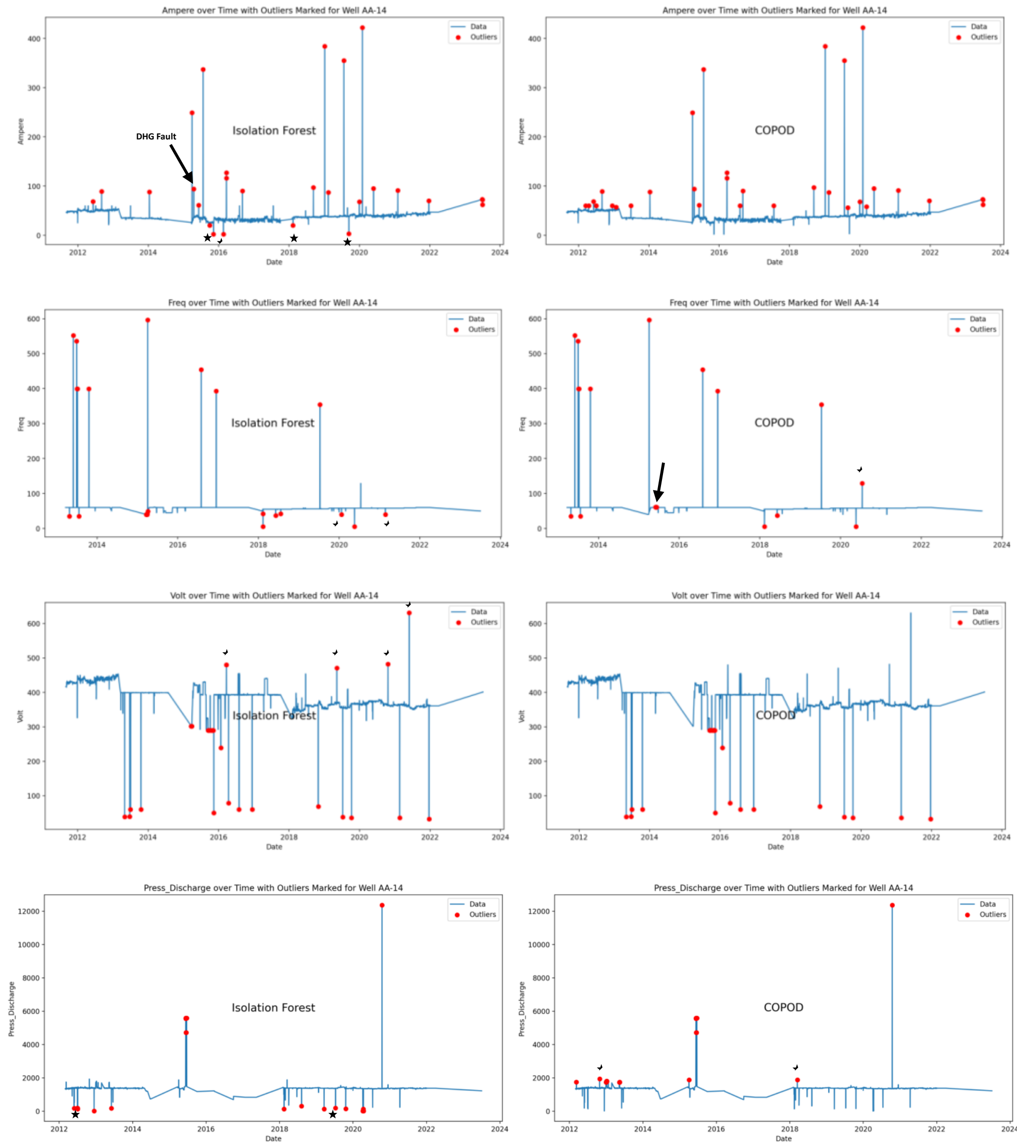

Fig.2 Comparison of anomaly detection on the Ampere, Frequency, Voltage and Discharge Pressure features of Well AA-14 based on long-term history data using the Isolation Forest (left); Copula-Based Outlier Detection (right).

interpretation of contamination, where non-anomalous data points are incorrectly labeled as anomalies, resulting in false alarms. This indicates that the detection results are highly sensitive to the chosen contamination value.

For this study, the contamination level is set to 0.01, ensuring a conservative assumption that the predicted anomaly labels correspond to actual anomalies. It should be noted that this contamination value does not represent the true anomaly frequency in the field, as ground-truth labels are unavailable. If the actual proportion of anomalous events is higher than 1%, the model may under-detect real anomalies. This limitation is inherent to unsupervised anomaly detection. Future work may incorporate semi-supervised approaches or data-driven statistical estimation methods, such as interquartile range (IQR) analysis on historical data, to derive a more realistic contamination value.

4.2 Long-Term History Anomaly Detection

This section presents the performance results of the Isolation Forest and COPOD models in detecting anomalies in the long-term historical data. Both models were treated similarly, with a contamination hyperparameter of 0.01 (i.e., 1% of the data consists of anomalies).

The anomaly detection results for Well AA-14 show that the Isolation Forest algorithm outperforms Copula-Based Outlier Detection (COPOD) in detecting anomalies across selected features such as ampere, frequency, voltage, discharge pressure, motor temperature, vibration, and gross rate, as shown in **Fig. 2** and **Fig. 3**. Isolation Forest effectively isolates data that deviates from the normal distribution, even with missing data, due to its hyperparameters like the number of trees, maximum sample size for training, and random state. In contrast, COPOD relies on assumptions about anomalies within the entire dataset, making it less accurate, especially when handling missing values. Detailed comparisons of both models for each parameter will be presented in the following paragraphs.

Ampere. Isolation Forest is more sensitive to anomalies in Well AA-14's ampere feature, detecting significant drops (e.g., 2016–2020) that COPOD misses. COPOD primarily

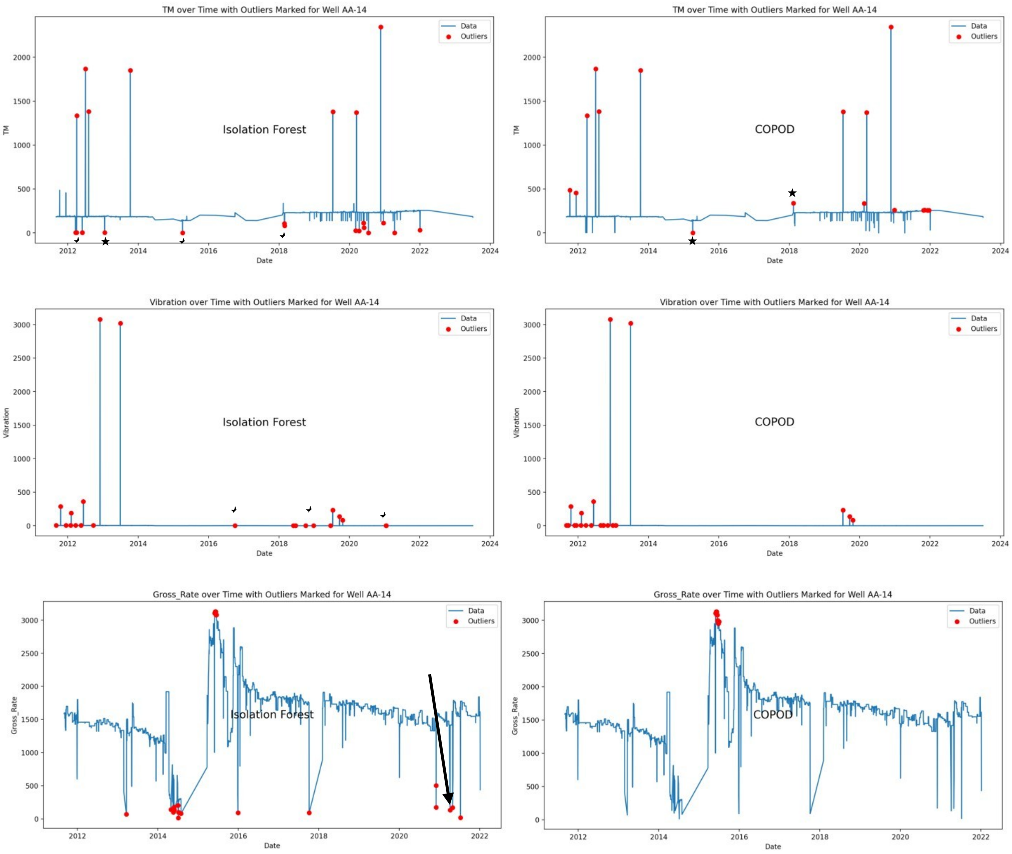

Fig.3 Comparison of anomaly detection on the Motor Temperature, Vibration and Gross Rate features of Well AA-14 based on long-term history data using the Isolation Forest (left); Copula-Based Outlier Detection (right).

Table 3. Summary of detected anomalies across individual sensors and their agreement during events flagged by COPOD and Isolation Forest in long-term historical data.

Algorithm	Ampere (A)	Voltage (V)	Frequency (Hz)	Discharge Pressure (Pa)	Motor Temperature (C)	Vibration $\left(\frac{m}{s^2}\right)$	Gross Rate (bbl/d)	Number of Anomalous Event Detected
COPOD	60	-	536	-	-	-	-	1
	-	60	399	-	-	-	-	4
	-	-	596	1889	-	-	-	1
	-	-	61	-	-	-	3105	8
	61	-	61	-	-	-	3126	1
	-	-	-	5591	-	-	3079	6
	-	-	354	-	1378	234	-	1
Isolation Forest	-	302	40	-	-	-	-	6
	249	-	42	-	-	-	-	2
	61	-	-	-	-	-	3126	1
	-	-	-	5591	-	-	3079	5
	127	480	-	-	-	-	-	3
		-	354	187	1378	234	-	1

detects increases in ampere. A cross-check on April 18, 2015, revealed a downhole gauge fault with an anomalous ampere value of 94 ampere given that the mean reading of ampere is 39.25 ampere.

Frequency. Isolation Forest identifies small, recurring anomalies, while COPOD detects sharp changes. Although COPOD detects an anomaly linked to the gross rate near the mean (59.1 Hz), Isolation Forest excels in detecting periodic anomalies.

Voltage. Isolation Forest detects a wider range of voltage anomalies, including extreme increases and decreases, while COPOD identifies only sharp drops. Voltage anomalies correlate with ampere anomalies.

Discharge Pressure. Isolation Forest consistently detects extreme variations in discharge pressure, whereas COPOD identifies increasing anomalies. COPOD's inability to detect decreases suggests a limitation in capturing certain anomaly patterns when strong physical correlations exist among ESP parameters. **Table 3** confirms anomalies in other features during undetected pressure anomalies, indicating that COPOD performance may be affected by correlated variables rather than data dimensionality.

Motor Temperature. There is no difference in anomaly detection between Isolation Forest and COPOD, indicating comparable performance for this feature.

Gross Rate. Isolation Forest detects both lower and upper anomalies, while COPOD recognizes only extremely high values as anomalies. A cross-check on lower gross rate anomalies reveals operational events: on April 7, 2021, the ESP operated for only 2 hours; on April 30, 2021, an Emergency Condition led to manual platform shutdown; and on July 10, 2021, the ESP ran for just 30 minutes. These findings highlight the effectiveness of Isolation Forest in capturing operational anomalies missed by COPOD.

Based on the results summarized in **Table 3**, both Isolation Forest and COPOD highlight correlations between anomalies across different features. Isolation Forest frequently detects anomalies in voltage and frequency, indicating electrical issues in ESP operations, where voltage and frequency show an inverse relationship. Additionally, it identifies strong correlations between discharge pressure and gross rate anomalies, which may suggest equipment wear or leakage. On July 12, 2019, anomalies in frequency, discharge pressure, motor temperature, and vibration point to potential issues with the ESP, motor, well conditions, or disruptions in the VSD control system. COPOD, while detecting more anomalies overall and identifying relationships between features that Isolation Forest misses (e.g., frequency with discharge pressure, and gross rate with both frequency and discharge pressure), tends to flag anomalies close to the normal distribution, limiting its effectiveness on broader anomaly patterns. Both algorithms reveal electrical-related issues in Well AA-14, particularly in voltage and frequency, and COPOD provides additional insights into potential equipment problems through its ability to correlate anomalies across features.

4.3 Short-Term History Anomaly Detection

This section presents the performance of Isolation Forest and COPOD in detecting anomalies over a short-term period from 2018/02/01 to 2022/04/30. This timeframe was chosen because it coincides with the ESP installation, ensuring no missing data. Both models were applied with the same hyperparameter, setting the contamination level

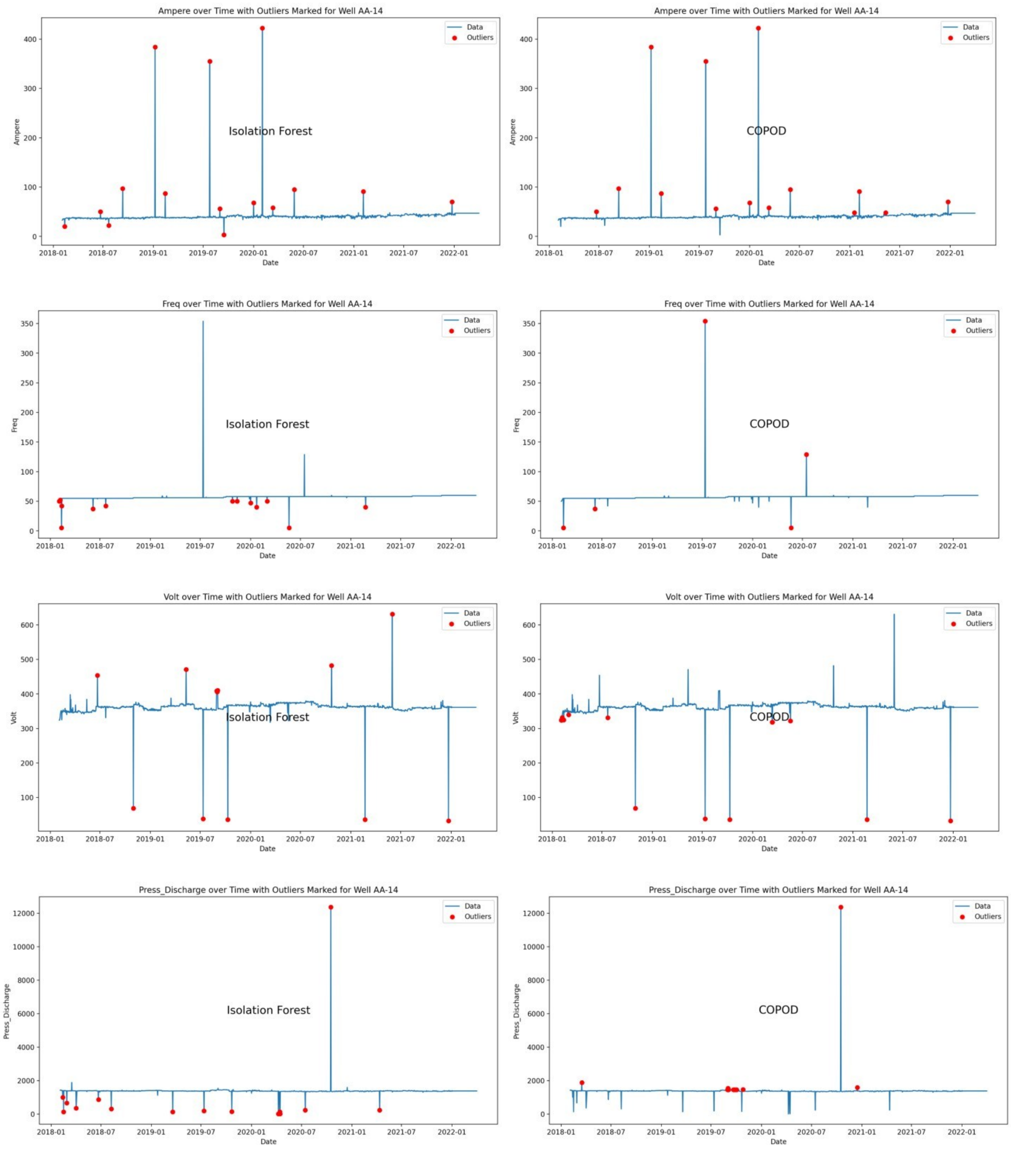

Fig.4 Comparison of anomaly detection on the Ampere, Frequency, Voltage, and Discharge Pressure features of Well AA-14 based on short-term history data using the Isolation Forest (left); Copula-Based Outlier Detection (right).

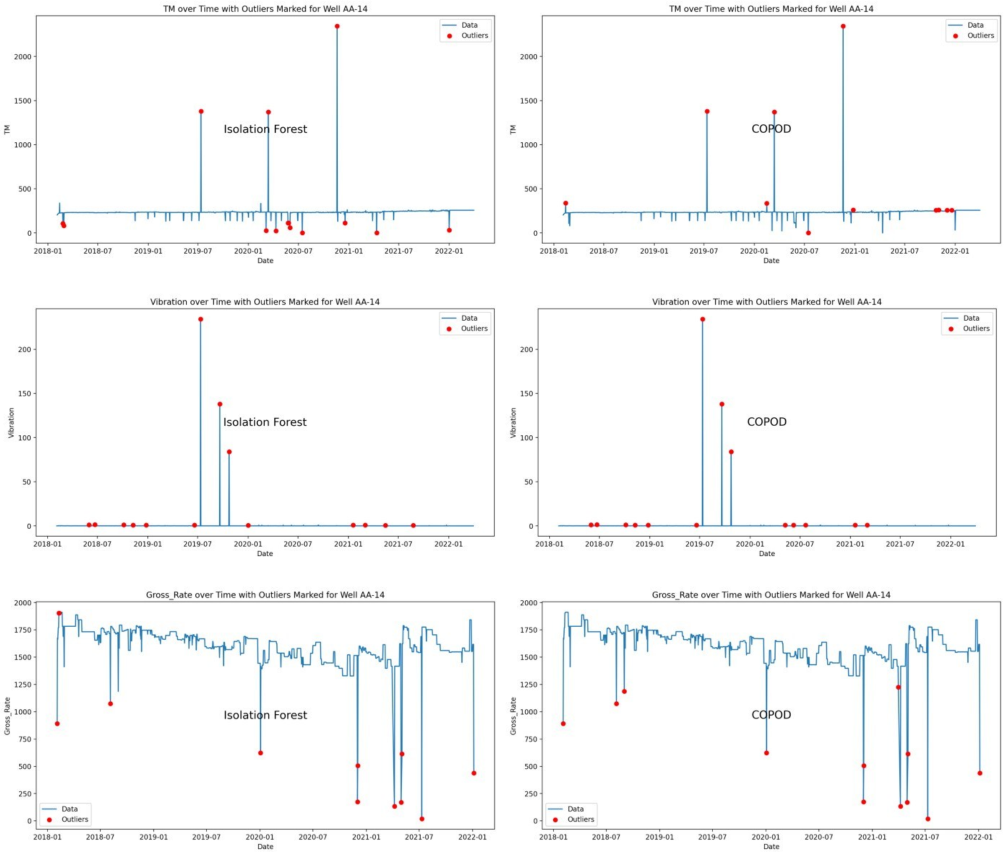

Fig.5 Comparison of anomaly detection on the Motor Temperature, Vibration and Gross Rate features of Well AA-14 based on short-term history data using the Isolation Forest (left); Copula-Based Outlier Detection (right)

Table 4. Summary of detected anomalies across individual sensors and their agreement during events flagged by COPOD and Isolation Forest in short-term historical data.

Algorithm	Ampere (A)	Voltage (V)	Frequency (Hz)	Discharge Pressure (Pa)	Motor Temperature (C)	Vibration $\left(\frac{m}{s^2}\right)$	Gross Rate (bbl/d)	Number of Anomalous Event Detected
COPOD	-	324	-	-	-	-	891	1
	50	454	-	869	-	1.5		1
	-	38	-	187	1378	234		1
	-	-	-	1548	-	-	1185	1
Isolation Forest	-	-	50	-	-	-	891	1
	20	-	42	1009	-	-	-	1
	50	454	-	869	-	1.5	-	1
	22	-	42	-	-	-	-	1
	-	38	-	187	1378	234	-	1
	68	-	47	-	-	0.5	-	1
	-	-	-	236	0.123	-	-	1

to 0.01, assuming 1% of the data is anomalous. A cross-check with dataset comments was also conducted to verify whether detected anomalies correspond to actual anomalies or scheduled maintenance.

As shown in **Fig. 4** and **Fig. 5**, the performance of Isolation Forest remains consistent in detecting anomalies, similar to its long-term history detection. Despite the expectation that COPOD would perform better without missing data in shorter period, it still struggles with accuracy in anomaly detection. In the analysis of Well AA-14's data from the ESP installation period, Isolation Forest showed strong performance in detecting anomalies across features like ampere, voltage, discharge pressure, motor temperature, vibration, and gross rate, though it underperformed with frequency, requiring further analysis with different hyperparameters. COPOD, on the other hand, consistently lacked sensitivity to data anomaly changes, suggesting a need for better hyperparameter tuning and further research to improve its performance.

Anomaly detection based on a short-term period or specific time frame allows for clearer identification of changes or fluctuations in the data for each feature. This results in data distributions appearing more normal compared to long-term history data. As a result, anomalies detected in one feature tend to correlate with anomalies in other features, unlike in long-term history detection. As shown in **Table 4**, short-term anomaly detection reveals how anomalies in one feature can impact others. Therefore, using short-term data, with no missing values, provides more accurate anomaly correlations, making it easier to analyze the underlying causes of the anomalies.

5. Conclusion

Isolation Forest demonstrates superior performance compared to COPOD in detecting anomalies across features in both long-term and short-term data. It effectively identifies extreme deviations from the normal distribution, while COPOD's reliance on feature correlation makes it less effective with high-dimensional data. Based on performance criteria, Isolation Forest is the most suitable model.

Anomaly detection results also reveal that anomalies in one feature often correlate with anomalies in others. Short-term data provides more accurate and diverse correlations compared to long-term data, enabling more precise identification of anomaly points. This underscores the interconnectedness of ESP feature anomalies.

References

Abdalla, R., Samara, H., Perozo, N., Carvajal, C.P. and Jaeger, P., 2022. Machine learning approach for predictive maintenance of the electrical submersible pumps (ESPs). *ACS Omega*, 7(21), pp.17641–17651. doi:10.1021/acsomega.1c05881

Al Maghlouth, A., Cumings, M., Al Awajy, M. and Amer, A., 2013. *ESP surveillance and optimization solutions: Ensuring best performance and optimum value*. Presented at: SPE Middle East Oil and Gas Show and Conference, Manama, Bahrain. doi:10.2118/164382-MS

Amer, M., Goldstein, M. and Abdennadher, S., 2013. Enhancing one-class support vector machines for unsupervised anomaly detection. *Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description*, pp.8–15. doi:10.1145/2500853.2500857

Bhardwaj, A.S., Saraf, R., Nair, G.G. and Vallabhaneni, S., 2020. *Real-time monitoring and predictive failure identification for electrical submersible pumps*. Presented at: SPE Russian Petroleum Technology Conference. doi:10.2118/201881-MS

Chandola, V., Banerjee, A. and Kumar, V., 2009. Anomaly detection. *Computing in Science & Engineering*, 14(1), pp.1–22. doi:10.1145/1541880.1541882

Ding, Z. and Fei, M., 2013. An anomaly detection approach based on isolation forest algorithm for streaming data using sliding window. *IFAC Proceedings Volumes*, 3(PART 1). doi:10.3182/20130902-3-CN-3020.00044

Guo, D., Raghavendra, C.S., Yao, K.T., Harding, M., Anvar, A. and Patel, A., 2015. *Data driven approach to failure prediction for electrical submersible pump systems*. Presented at: SPE Western Regional Meeting, pp.967–972. doi:10.2118/174062-MS

Gupta, S., Saputelli, L. and Nikolaou, M., 2016. *Big data analytics workflow to safeguard ESP operations in real-time*. Presented at: SPE North America Artificial Lift Conference and Exhibition, 25–27 October. doi:10.2118/181224-MS

Jansen van Rensburg, N., 2018. *Usage of artificial intelligence to reduce operational disruptions of ESPs by implementing predictive maintenance*. Presented at: Abu Dhabi International Petroleum Exhibition & Conference. doi:10.2118/192610-MS

Liang, X., Duan, F., Bennett, I. and Mba, D., 2020. A sparse autoencoder-based unsupervised scheme for pump fault detection and isolation. *Applied Sciences*, 10(19), p.6789. doi:10.3390/app10196789

Liu, D., Feng, G., Feng, G. and Xie, L., 2024. Hybrid long short-term memory and convolutional neural network architecture for electric submersible pump condition prediction and diagnosis. *SPE Journal*, 29(5), pp.2130–2147. doi:10.2118/218418-PA

Sherif, S., Adenike, O., Obehi, E., Funso, A. and Eyituoyo, B., 2019. *Predictive data analytics for effective electric submersible pump management*. Presented at: SPE Nigeria Annual International Conference and Exhibition. doi:10.2118/198759-MSs

Shuwaikhat, H.A., Ramos, M., Aifan, A.-R. and Al-Sadah, A.A., 2017. *Innovative approach to prolong ESP run life using algorithmic models*. Presented at: Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE. doi:10.2118/188807-MS

Sneed, J., 2017. *Predicting ESP lifespan with machine learning*. Presented at: SPE/AAPG/SEG Unconventional Resources Technology Conference, pp.863–869. doi:10.15530/urtec-2017-2669988

Song, Y., Jun, S., Nguyen, T.C. and Wang, J., 2024. Experimental data-driven model development for ESP failure diagnosis based on the principal component analysis. *Journal of Petroleum Exploration and Production Technology*, 14(6), pp.1521–1537. doi:10.1007/s13202-024-01777-9

Takacs, G., 2009. *Electrical Submersible Pumps Manual*. Oxford: Gulf Professional Publishing. doi:10.1016/B978-1-85617-557-9.X0001-2

Zhu, D., Wang, X., Chen, H. and Wu, R., 2014. Semi-supervised support vector machines regression. *Proceedings of the 2014 9th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pp.2015–2018. doi:10.1109/ICIEA.2014.6931500