2024 European Control Conference (ECC)
June 25-28, 2024. Stockholm, Sweden

Data-Driven Forecasting Based Anomaly Detection: A Reciprocating Compressor Application

Aycan Sapmaz, Aslı Yasmal, Gizem Kusoglu Kaya, Barış Akgün, *IEEE*

Equipment reliability is a critical aspect of petrochemical refineries. Timely anomaly (due to equipment failures, sensor faults, wear and tear, unexpected inputs etc.) detection is essential to keep equipment running safely, improve performance, and have an efficient and effective maintenance strategy. Advances in machine learning and availability of large amounts of process data makes it possible to build data-driven models for monitoring complex processes and equipment in real-time. These models can guide operators, maintenance and process engineers in identifying faults and isolating their root causes. This study's primary contribution is the demonstration of a methodology utilizing field data of a reciprocating compressor in a petrochemical refinery and final implementation in a real-time environment.

In the literature, Charoenchitt et al. used an autoencoder based method to detect anomalies in reciprocating compressors, incorporating their thermodynamic equations. Additionally, they used vibration spectrums to train the model [1]. In another study, Palacín et al. worked on anomaly detection for centrifugal compressors. Their approach consists of performing principal component analysis (PCA) on process data and utilizing the Mahalanobis distance on these to detect anomalies [2]. To appear our previous study, it was also tested on recycle compressor on real case. Its results are similar, and it will be published soon [3].

In contrast, we develop a fully data-driven approach. This involves learning a model of regular compressor operation and comparing its outputs to measurements, flagging large differences as anomalies. We employ deep learning models with Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs) layers to forecast future sensor outputs given past measurements. We use real historical data to train and select the model. After training, large differences between forecasts and measurements are treated as anomalies. The idea is that the model trained from regular operation will make forecasts close to regular operation and if there are any discrepancies between the forecasts and measurements, then there may be an anomaly in the system.

We tested our approach on both historical and real-time data. Three years of historical process data was collected starting from a periodic maintenance of the reciprocating compressor. This sensor data was gathered at 15-minute intervals and there were some stoppages, leading to 73968 data points. We chose the compressor outlet as the Key Process Indicator (KPI) for monitoring compressor health and detecting anomalies. The compressor outlet temperature effectively captures the dynamics of the process and mirrors fluctuations in other process variables. 18 process variables affecting the outlet temperature are selected with the help of process engineers: stage suction temperatures (°C), cylinder and crank case vibration values (rpm), gas flow rate (m^3/h), stage suction pressure (kg/cm^2) and stage outlet pressure (kg/cm^2). We divided the dataset into a training set (anomaly free) for learning the model and a testing set for evaluating the model. The test set encompasses anomalies, shutdowns, interventions, instances when the compressor operates independently, and other operational changes. From the model's standpoint, these events deviate from the regular operation and thus identified as anomalies.

When training our model, we employed the early stopping technique to address the issue of overfitting and ensure optimal generalization. To implement early stopping, we utilized a validation dataset that was separate from the training and test datasets. This validation dataset allowed us to assess the model's performance on unseen data during the training process. Early stopping played a crucial role in preventing overfitting by monitoring the performance metric, such as accuracy or error, on the validation dataset. If the performance metric failed to show improvement over a defined patience period, early stopping was triggered, halting the training process. This approach helped us strike the right balance between model complexity and generalization, as it allowed us to stop the training at the point where the model exhibited the best performance on unseen data.

We trained multiple deep learning models using GRU and LSTM layers, with hidden unit sizes of 64 and 128. Model performance was evaluated using the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) metrics. The RMSE and MAE values for the train and test datasets

Manuscript received February 14, 2024. This work was supported in part by the Turkish Petroleum Refineries Corporations.

A.Sapmaz is with the Turkish Petroleum Refineries Corporations, Körfez, Kocaeli, 41780, Türkiye (corresponding author to provide phone: +90 0262 316 38 64; e-mail: aycan.sapmaz@tupras.com.tr).

A.Yasmal is with the Turkish Petroleum Refineries Corporations, Körfez, Kocaeli, 41780, Türkiye (e-mail: asli.yasmal@tupras.com.tr).

G.Kuşoğlu Kaya is with the Turkish Petroleum Refineries Corporations, Körfez, Kocaeli, 41780, Türkiye (e-mail: Gizem.Kusoğlu@tupras.com.tr).

B.Akgün is with the Dept. of Computer Engineering, Koç University, Sarıyer, Istanbul, 34450, Türkiye (e-mail: BAAKGUN@ku.edu.tr).

are presented in Table 1. The performance metrics were carefully analyzed for the training set to identify the most suitable model. In this context, we selected the GRU 128 model based on its lower RMSE and MAE values. As this study focused on anomaly detection, a visual examination of the RMSE and MAE values for the test dataset was also conducted. Figure 1 depicts the graph of the model that yielded the most promising results.

TABLE I
PERFORMANCE METRICS

Train	GRU (64)	GRU (128)	LSTM (64)	LSTM (128)
RMSE	0.56	0.37	0.55	0.47
MAE	0.37	0.23	0.35	0.31
Test	GRU (64)	GRU (128)	LSTM (64)	LSTM (128)
RMSE	0.74	0.61	0.96	0.80
MAE	0.48	0.39	0.60	0.58

The control chart displayed in Figure 1 showcases the difference between the observed compressor outlet temperature and the predictions made by the model. By collaborating with process engineers and referring to relevant literature, upper and lower control limits (horizontal lines) were established at ±2.0. The region enclosed by these control limits is referred to as the control region, while the data points falling outside this region are identified as anomalies. While the numbered vertical lines indicate the occurrence of errors within the specified dates, the two unnumbered vertical lines represent predictive maintenance periods. The model has achieved success in providing accurate alerts for 4 out of 5 anomalies in the system for two years. It has a precision rate of 80% and a recall rate of 100%.

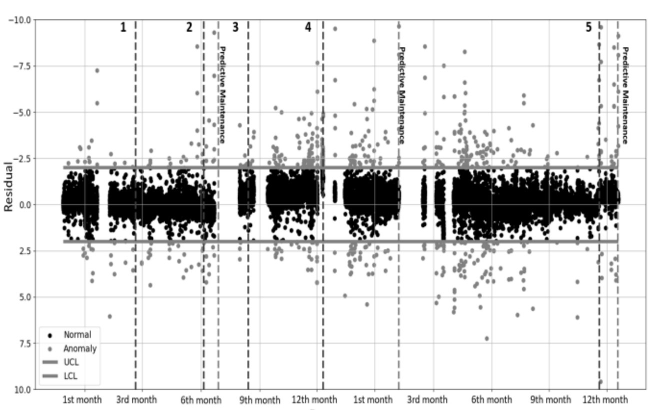

Fig 1. Residual of compressor outlet temperature GRU128

The real-time outcomes of the study can be examined by maintenance and process engineers, who serve as the primary users. The study's findings hold significance for these users, as they rely on them to make informed decisions and take preventive measures prior to the occurrence of anomalies. The study's results serve as a valuable decision-support system for proactive action.

References

[1] [1] Charoenchitt, C., & Tangamchit, P. (2021, January). Anomaly detection of a reciprocating compressor using autoencoders. In 2021 Second International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP) (pp. 1-4). IEEE.

[2] Palacín, I., Gibert, D., Planes, J., Arena, S., Orrù, P. F., Melis, M., & Annis, M. (2021). Anomaly detection for diagnosing failures in a centrifugal compressor train. In Artificial Intelligence Research and Development (pp. 217-220). IOS Press.

[3] A.Sapmaz, A.Yasmal, G.Kusoğlu Kaya, Y. Utar, B. Akgün, Data-Driven Forecasting for Anomaly Detection in a Compressor Unit, In Computer Aided Process Engineering, to be published.