Predicting Power Consumption of Cryogenic Compressors using Multiple Linear Regression in Machine Learning

Muhammad Fikri Hashim[a], Nur Izyan Zulkafli[a,*], Mohamad Fani Sulaima[b], Mohd Hafiz Jali[b], Tarmizi Ahmad Izzuddin[b], Nur Saleha Jayiddin[c], Azmi Md Lasin[c], M Tarmidzi Iskandar[c]

[a]Faculty of Mechanical Technology and Engineering, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100, Durian Tunggal, Melaka
[b]Faculty of Electrical Technology and Engineering, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100, Durian Tunggal, Melaka
[c]Petronas Group Technical Solutions, Projects, Technology and Health, Safety, Security and Environment, 50088, Kuala Lumpur
nurizyan@utem.edu.my

Compressor performance is being evaluated based on its power consumption and other operational parameters to meet load demand efficiently while consuming less power. Without proper correlation with other operational data, it is difficult to predict future power consumption that may lead to a low performance of compressors. The Multiple Linear Regression (MLR) analysis in Altair AI Studio software is being used as a model to predict power consumption for four compressors with two different models by considering mass flow rate, suction and discharge temperature, and pressure as its dependent variables. The set of data has been split into two, which are training and testing, at a ratio of 90:10, respectively. This study resulted in a low percentage difference between the predicted and actual power consumption of those four compressors, which are 1.46 %, 1.40 %, 2.00 %, and 2.25 % for Compressor 1, Compressor 2, Compressor 3, and Compressor 4, respectively. The MLR of the compressor power consumption model can be utilized to predict its future power consumption to move towards more sustainable and low-carbon emissions.

1. Introduction

Compressors are vital components in industrial plants, supplying compressed gas through pipings to meet the demands of production units. The compressed gases are used in various applications from process to utilities. However, compressors, which are commonly used in such settings, are highly energy-intensive, leading to higher electricity costs and increased carbon emissions. Before implementing optimization planning, it is important to first understand which operational parameters have a significant effect on the energy consumption of the compressor. Predicting future energy consumption can be challenging without an appropriate correlation, as the compressor operates under various conditions. The implementation of machine learning can be used to analyze the complexity of the power consumption correlation with its operational parameters. Hence, this study is being conducted to develop a power consumption predicted model based on the current condition of the compressor by using Multiple Linear Regression (MLR) analysis in machine learning. MLR belongs to a group of regression algorithms that fall into a subset of machine learning (Ansari and Nassif, 2022). Few studies have explored the prediction of the performance of compressors, such as Ding et al. (2023) explored a fuel air compressor, while Ossorio et al. (2024) focused on a variable-speed compressor. However, there is a lack of studies that focus on gas as the compressed fluid. This limitation is crucial as these compressed air models are unsuitable for gas application, as their chemical and physical properties are different. Moreover, the demand for gas transportation in the energy industry is rising, and a predictive performance model needs to be developed (Shui et al., 2023).

2. Methodology

2.1 Regasification Terminal Description

The compressor plant consists of five compressors that are divided into two types, which are three boil-off gas (BOG) compressors (Compressors 3, 4, and 5) and two regasification terminal export (RGTEC) compressors (Compressors 1 and 2). Compressor 5 is not included in this study as it was in a shutdown status for the whole time that data was being collected. During minimal send-out, boil-off gas (BOG) from the storage unit (FSU) at extremely low temperatures is initially compressed by BOG compressors, increasing its pressure and temperature. The gas then passes through air fin coolers to reduce its temperature before being further compressed by RGTEC compressors to high pressures and distributed directly to the PGU pipeline. During normal send-out mode, BOG discharged from the BOG compressor will be delivered to the regasification package unit. The number of active compressors depends on the operational mode, ensuring efficient send-out and minimizing flaring during low-demand periods. Figure 1 shows the arrangement of the compressors along with the sensors that are being used to collect operational parameters.
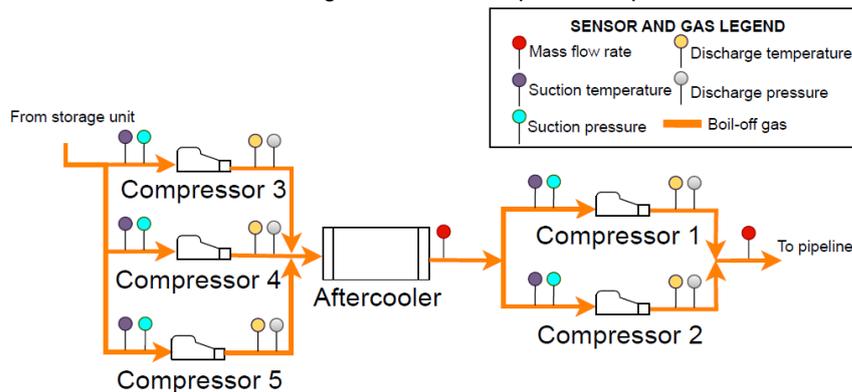

Figure 1: Compressors network arrangement in the plant

2.2 Data Collection

The data is being obtained from historical operational logs of the compressors that have been recorded hourly for three weeks. In this study, five independent variables are chosen, which are the mass flow rate of gas, suction and discharge of temperature and pressure, while the power consumption of the compressor is the dependent variable. The selection of those independent variables was due to the influence they have on the thermodynamic state of the compressor (Zhao et al., 2024).

2.3 Multiple Linear Regression Analysis setup in Altair AI Studio software

Altair AI Studio software is being used in this study to develop a power consumption predicted model based on the current condition. Historical data for power consumption, mass flow rate, suction, and discharge of temperature and pressure for each compressor are being retrieved and served as an input for this model. All dependent (power consumption) and independent variables (mass flow rate of gas, suction and discharge of temperature and pressure) are being selected in Select Attributes. While in Set Role, the dependent variable, which is power consumption, is being set as "label" for the model to recognize it as the parameter that is desired to predict.

2.4 Model Training and Testing

The set of data is being split into a ratio of 90:10 for training and testing, respectively. 90 % of the data is being used for training, while another 10 % is being used for testing. According to Sivakumar et al. (2024), the regular ratio that is used in train-test data split is 70:30, 80:20, or 90:10, depending on the size of the dataset. In training, the model uses the data to learn the correlation between the dependent variable and the independent variables. Upon completion of training, the model uses another 10 % of data to do the same task as in training.

2.5 Multiple Linear Regression (MLR) analysis

To find the relationship between dependent and independent variables, a statistical analysis called Multiple Linear Regression can be used (Dhaval and Deshpande, 2020). The relationship of compressor power consumption with mass flow rate, suction and discharge temperature, and pressure is acquired through a linear function for each compressor. As the assumption of MLR is linear, the general equation of MLR can be seen as

in Eq(1), with Y as the dependent variable, x_n as the independent variables, β_n as the coefficients, and β_0 as the intercept.

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \tag{1}$$

Goodness of the model is being evaluated through the correlation coefficient, R^2. It ranges from zero to one, which greater than 0.7 considered a good correlation (Saleem et al., 2024). The final stage of MLR analysis is to use the generated intercept and coefficients for independent variables that form a linear function to calculate the predicted power consumption. The predicted power consumption is being used to be compared in terms of percentage difference with the actual one. The selection of MLR is based on its interpretability compared to other methods. It quantifies excellently the influence of each independent variable on the dependent variable through the generated coefficients. It represents the change in the dependent variable when there is a change of one unit of the independent variable. This clear interpretation not only can predict power consumption, but also identify which operational parameters have the most influence to aid in decision making when it comes to real applications.

3. Results and Discussion

3.1 Performance of Power Consumption Predicted Model based on Current Condition

MLR analysis is being carried out to find the metrics that will be used to evaluate the model's performance. Correlation coefficient, R^2, p-value, and coefficients are the outcomes from MLR analysis that have been observed to evaluate the performance of the compressor power consumption model. R^2 value indicates how much the independent variables explain the dependent variable in percentage. A good R^2 value should be more than 0.7 in the engineering field. P-value indicates the significance of the independent variables towards the dependent variable. A good p-value should be less than 0.05, indicating statistically significant results for that independent variable towards the dependent variable (Ng et al., 2018).

Table 1: Statistical metrics of the training model

Compressor	Compressor 1		Compressor 2		Compressor 3		Compressor 4	
Independent variables	p-value	R^2	p-value	R^2	p-value	R^2	p-value	R^2
$\dot{m}$	0.00		0.00		0.00		0.00	
T_{suc}	0.85		0.09		0.00		0.00	
T_{dis}	0.05	0.998	0.02	0.977	0.00	0.960	0.00	0.999
P_{suc}	0.00		0.00		0.00		0.00	
P_{dis}	0.00		0.00		0.00		0.00	

The correlation coefficient, R^2, for every compressor is presented in Table 1. All four models portray a very high correlation between power consumption and the operational parameters. This demonstrates how well the selected operational parameters are taken into account for power consumption, which exceeds 95 %. The p-value for each independent variable concerning its compressor is also shown in Table 1. It is visible that the mass flow rate of gas, $\dot{m}$, has a significant effect on power consumption, resulting in a zero p-value for every compressor. This can be proved by the coefficient of mass flow rate for each compressor, as in Table 2, which has a positive value, indicating higher mass flow rate means a higher quantity of gas that needs to be compressed. Hence, compressors need to do more work to achieve the desired pressure. Eq(2) illustrates theoretical power consumption, where n is the polytropic exponent, η_{poly} is the polytropic efficiency and Z_a is the compressibility factor. P_s, T_s and q_s represent the standard pressure, temperature and volumetric flow rate, while T_{suc}, P_{suc} and P_{dis} represent the suction temperature, suction pressure and discharge pressure, respectively. Both suction and discharge pressures portray a strong influence on power consumption that can be explained by using the theoretical expression as in Eq(2) (Moshfeghian, 2014). It demonstrates that higher discharge pressure combined with low suction pressure amplifies the power consumption (Farzaneh-gord and Niazmand, 2014)

Suction and discharge of pressure also gave a high significance to power consumption. It can be explained in detail through theoretical calculation of power consumption, as shown in Eq(2) (Moshfeghian, 2014) illustrates that a high discharge pressure with low suction pressure can increase power consumption (Farzaneh-gord and Niazmand, 2014). It also aligns with theory as higher discharge pressure needs more work for the compressor to do. The suction and discharge temperature for Compressors 3 and 4 shows a high significance, vice versa for Compressors 1 and 2. This is due to Compressors 3 and 4 taking gas from storage, which has a high variability that MLR deems to be significant. Moreover, since Compressors 3 and 4 take in gas at a very low

temperature, the power being used is higher and more significant for the compression process to reach the discharge temperature at the desired pressure. This huge difference between suction and discharge temperature in Compressors 3 and 4 also explains why its models consider the significance of discharge temperature. Contrast with Compressors 1 and 2, which receive gas at warmer and more stable temperatures, which is not very significant towards power consumption. The low significance of discharge temperature for Compressors 1 and 2 follows the theory as it is not included in Eq(2) in calculating power consumption.

$$\dot{W} = \frac{n}{n-1} \frac{Z_a}{\eta_{poly}} \frac{P_s}{T_s} q_s T_{suc} \left[\left(\frac{P_{dis}}{P_{suc}}\right)^{\frac{n-1}{n}} - 1 \right] \tag{2}$$

3.2 Actual vs. Predicted Power Consumption

After the MLR is performed, the goal is to find the percentage difference between the actual and predicted power consumption. The actual power consumption is from historical data that serves as the input of the MLR analysis. The intercepts and coefficients in Table 2 are being used to form a linear function to calculate the predicted power consumption as shown in Eq(3) for each compressor. The intercept represents the baseline value of power consumption when all the independent variables are zero, while $\alpha, \beta, \gamma, \delta$ and σ are the coefficients that are being associated with mass flow rate ($\dot{m}$), suction and discharge of temperature (T_{suc}, T_{dis}), and pressure (P_{suc}, P_{dis}) respectively.

$$P_{c,n} = \text{intercept} + \alpha(\dot{m}) + \beta(T_{suc}) + \gamma(T_{dis}) + \delta(P_{suc}) + \sigma(P_{dis}) \tag{3}$$

Table 2: Intercept and coefficients for each compressor

Compressor Item	Compressor 1	Compressor 2	Compressor 3	Compressor 4
intercept	-15,211.9	-482,730.0	-14,130.3	-19,505.0
α	1277.2	3590.3	48.1	74.4
β	57.8	-711.5	3,805.7	395.0
γ	482.1	-946.5	-1,573.3	365.4
δ	-12,033.8	-29,811.2	152.2	-31,402.7
σ	11,088.5	18,154.9	152,161.0	20,159.4

Figure 2 shows the trend comparison of actual and predicted power consumption according to time series for Compressor 1, Compressor 2, Compressor 3, and Compressor 4. Compressor 1 and Compressor 2 in Figure 2a and Figure 2b, portray a high correlation between the actual and predicted power consumption. These models are also able to predict accurately during the fluctuations that occur at the early stage of start-ups until the time point 111. This fluctuation is due to the Compressor 1 is about to start up, shut down, and restart while Compressor 2 is delivering a low gas demand during that period. The high reliability of these models can also be seen when they manage to keep on track for a long period with the actual one from time point 111 until the end of operation. It is due to the operational parameters having been stabilized, which are mass flow rate, suction, and discharge of temperature and pressure. Figures 3a and 3b also support the reliability of the models by showing that the predicted versus actual power consumption are being plotted near the trendline along with high R^2 values, which are 0.998 and 0.977 for Compressor 1 and Compressor 2, respectively. For Compressor 3 in Figure 2c, it can be observed that the trend difference between the actual and predicted power consumption is significant. The model underpredicted the actual power consumption at time points 46, 114, and 443 while overpredicting at time points 300 and 450. Based on Figure 3c, it can be observed that the model tends to underpredict the power consumption at a high load, which is 1,400 kW and above. Inconsistency of this model in predicting near the actual one explains why 0.960 of its correlation coefficient, R^2, is the lowest compared to the other models, as visualized in Table 1 and Figure 3c. For Compressor 4 in Figure 2d, most of the time it is in a shutdown status, time points 15 to 70 and 100 to 500, and operates for a while, time points 0 to 14 and 69 to 99. The high reliability of this model is evident when it handles a sudden change in power consumption. During the shutdown status, especially at the early stage, it can be observed that the model predicted a very small value compared to when it is in operation, but not zero. This is due to the sensor still capturing a small value of those operational parameters even when it is in shutdown status. When all the small values of operational parameters are being plugged into Eq(3), the predicted power consumption will produce a non-zero value. Overall, all the power consumption predicted model based on current conditions show a strong and good pattern recognition with the actual trend of power consumption, resulting from a very high correlation coefficient, R^2, as in Table 1. However, the accurate prediction of fluctuations and trends, along with a very high R^2 value,

may indicate overfitting. It occurs when the models perform very well on the training dataset but poorly on unseen data as they learn specific details, including the outliers and noise, rather than capturing the general trend of the data. This can reduce the reliability of these models when they are being introduced to new and unseen data. To evaluate overfitting, the performance on the testing set should be observed. High performance through the value of R^2 for both datasets might indicate that overfitting is unlikely to happen. However, the similarity of input data in both datasets should be carefully considered, as the testing performance might replicate the same patterns of data learning. Hence, a robust way to evaluate overfitting is by introducing a truly new unseen dataset. A good model can be considered if it maintains the same performance as in training.
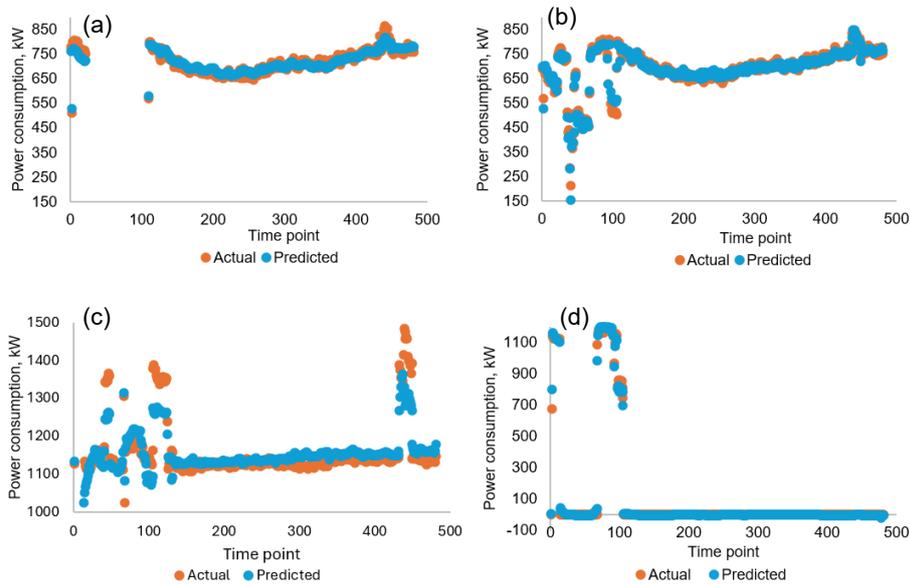

Figure 2: Comparison of trend between actual and predicted power consumption for a) Compressor 1, b) Compressor 2, c) Compressor 3, and d) Compressor 4
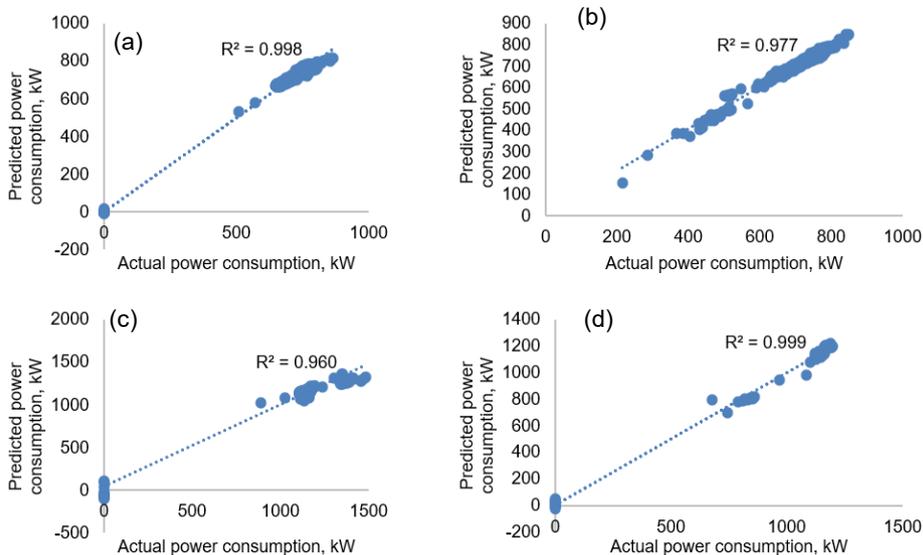

Figure 3: Overall deviation of predicted from actual power consumption for a) Compressor 1, b) Compressor 2, c) Compressor 3 and d) Compressor 4

3.3 Percentage Deviation Analysis of Power Consumption Models

To summarize the deviation between actual and predicted power consumption for each compressor, the average percentage difference is calculated as in Eq(4), where the predicted value is the outcome of the model in Eq(3) while the actual value is obtained from raw data. The percentage difference for all compressors is low,

demonstrating a high accuracy of the model in predicting power consumption. This indicates that the high performance of the model does not only rely on the R^2 value but is also good at predicting power consumption.

$$\text{Percentage difference \%} = \left|\frac{\text{Actual} - \text{Predicted}}{\text{Actual}}\right| \times 100\ \% \tag{4}$$

Table 3: Average percentage difference between actual and predicted power consumption

Compressor	1	2	3	4
Percentage difference, %	1.40	1.46	2.00	2.25

4. Conclusion

Chapter 2 A power consumption prediction model for cryogenic compressors was developed through Multiple Linear Regression (MLR) analysis in Altair AI Studio software based on the correlation between power consumption and operational parameters, which are mass flow rate, suction and discharge temperature and pressure. The historical data being split into a ratio of 90:10 for training and testing, respectively. The outcomes illustrate a high correlation between power consumption and operational parameters, as the correlation coefficient, R^2, for four compressors is greater than 0.95, with most p-values below 0.05. The average percentage difference has been calculated for each compressor to observe the deviation of the models from the actual power consumption. This percentage difference resulted in 1.40 %, 1.46 %, 2.00 %, and 2.25 % for Compressors 1-4, respectively, signifying high model reliability. However, the limitation of this study lies in its potential for overfitting due to a very high R^2 value and being too sensitive to fluctuations, outliers, and noise. Future studies may explore the utilization of deep learning algorithms, to predict the power consumption of the compressor and compare its reliability by setting up the actual power consumption as the reference.

Acknowledgments

References

Ansari, S., Nassif, A.B., 2022, A Comprehensive Study of Regression Analysis and the Existing Techniques. Advances in Science and Engineering Technology International Conferences, ASET 2022, Institute of Electrical and Electronics Engineers, Dubai, United Arab Emirates, 1–10.

Dhaval, B., and Deshpande, A., 2020, Short-term load forecasting with using multiple linear regression. International Journal of Electrical and Computer Engineering, 10(4), 3911–3917.

Ding, C., Xia, Y., Yuan, Z., Yang, H., Fu, J., Chen, Z., 2023, Performance prediction for a fuel cell air compressor based on the combination of backpropagation neural network optimized by genetic algorithm (GA-BP) and support vector machine (SVM) algorithms. Thermal Science and Engineering Progress, 44, 102070.

Farzaneh-gord, M., Niazmand, A., 2014, Thermodynamic analysis of natural gas reciprocating compressors based on real and ideal gas models. International Journal of Refrigeration, 56, 186-197.

Moshfeghian M., 2014, How to Estimate Compressor Efficiency?, John M. Campbell and Co., <https://www.jmcampbell.com/tip-of-the-month/2015/07/how-to-estimate-compressor-efficiency/> accessed 04.11.2025.

Ng, S. F., Chew, Y.M., Chng, P.E., and Ng, K.S., 2018, An Insight of Linear Regression Analysis. Scientific Research Journal, 15(2), 1-16.

Ossorio, R., Marchante-Avellaneda, J., Navarro-Peris, E., 2024, Development of compact empirical models for variable-speed compressors for the prediction of energy consumption, mass flow and discharge temperature. Applied Thermal Engineering, 244, 122666.

Saleem, S.N.A.N., Zulkafli, N.I., Tee, B.T., Sukri, M. F., Tahir, M M., Muhajir, A., Sulaima, M.F., Hanak, D.P., 2024, Multiple Linear Model Analysis of Indoor Air Quality for Air Conditioning System in Office Building. Chemical Engineering Transactions , 113, 127–132.

Shui, C., Zhou, D., Wu, Z., Yu, W., Zhang, L., Xing, T., Wang, C., Du, Z., 2023, Short-term operation optimization for natural gas pipeline considering line-pack: A perspective of optimal transport. Gas Science and Engineering, 117(February), 205075.

Sivakumar, M., Parthasarathy, S., Padmapriya, T., 2024, Trade-off between training and testing ratio in machine learning for medical image processing. PeerJ Computer Science, 10, e2245.

Zhao, Z., Wang, G., He, J., Tian, Y., 2024, Thermodynamics and dynamic investigation of ultra-high-pressure diaphragm compressor. Frontiers in Energy Research, 12, 1–6.