

# Propylene Yield from Olefin Plant Utilizing Box-Cox Transformation in Regression Analysis

Mohamad Hafizi Zakria<sup>1,2,\*</sup>, Mohd Ghazali Mohd Nawawi<sup>1</sup>, and Mohd Rizal Abdul Rahman<sup>2,3</sup>

<sup>1</sup>School of Chemical and Energy Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, 81310, Johor, Malaysia

<sup>2</sup>Steam Cracker Complex, Pengerang Refining Company Sdn Bhd, PICMO B2, Pengerang Integrated Complex, 81600, Johor, Malaysia

<sup>3</sup>PETRONAS Refinery and Petrochemicals Corporation Sdn Bhd, The Intermark Tower, 55000 Kuala Lumpur, Malaysia

**Abstract.** Propylene yield is one of the key operating parameters that is monitored daily in the running olefin plant. This study was conducted in the actual world-scale olefin plant to measure the impact of identified controlled variables on the propylene yield. The Box-Cox data transformation was adopted in the Regression Analysis using Minitab Software Version 18 due to non-normal data were observed after normality and stability test were conducted using Box Plot, I-MR Chart, Run Chart, Graphical Summary, and Normality Plot tools. The model concluded that propylene yield in the studied plant was contributed by the factors of -0.000243 Hearth Burner Flow, 0.01332 Integral Burner Flow, and 0.08598 Naphtha Feed Flow. The Response Optimizer tool also suggested that the propylene yield from naphtha liquid feed can be maximized at 11.22% with the control setting at 10,993.86 kg/hr of Hearth Burner Flow, 604.61 kg/hr of Integral Burner Flow, and 63.50 t/hr of Naphtha Feed Flow.

## 1 Introduction

Propylene yield monitoring is essential in the steam cracker furnace as its value translates to the profit generation and sustainability of the olefin plant. This study was conducted with a focus on propylene yield at the newly commissioned olefin plant. The plant was designed to produce 645 KTA capacity of polymer grade propylene product from naphtha feed via pyrolysis cracking in the steam cracker furnace.

Pyrolysis cracking refers to a reaction that causes hydrocarbon bonds to break and form the smaller unsaturated molecule [1, 2]. The high valuable olefins such as ethylene and propylene are normally produced from the elevated temperature cracking reaction [3, 4] in the steam cracker furnace. Pyrolysis cracking was primarily developed in the early 1920s with a focus to enhance the quality and quantity of gasoline components for the refinery [5] and the technology was continuously developed for olefin production. It is also the heart of the olefin process where the main reaction in the olefin plant occurs via pyrolysis cracking in the steam cracker furnace.

Due to this, pyrolysis reaction is often regarded as the key factor for the smooth running of the olefin plant operation. Conducting the study on the pyrolysis reaction in the actual plant conditions is more challenging than the lab-scale experiments due to the extreme process fluctuation [6] in the olefin plant resulted from dynamic operation in the upstream processes.

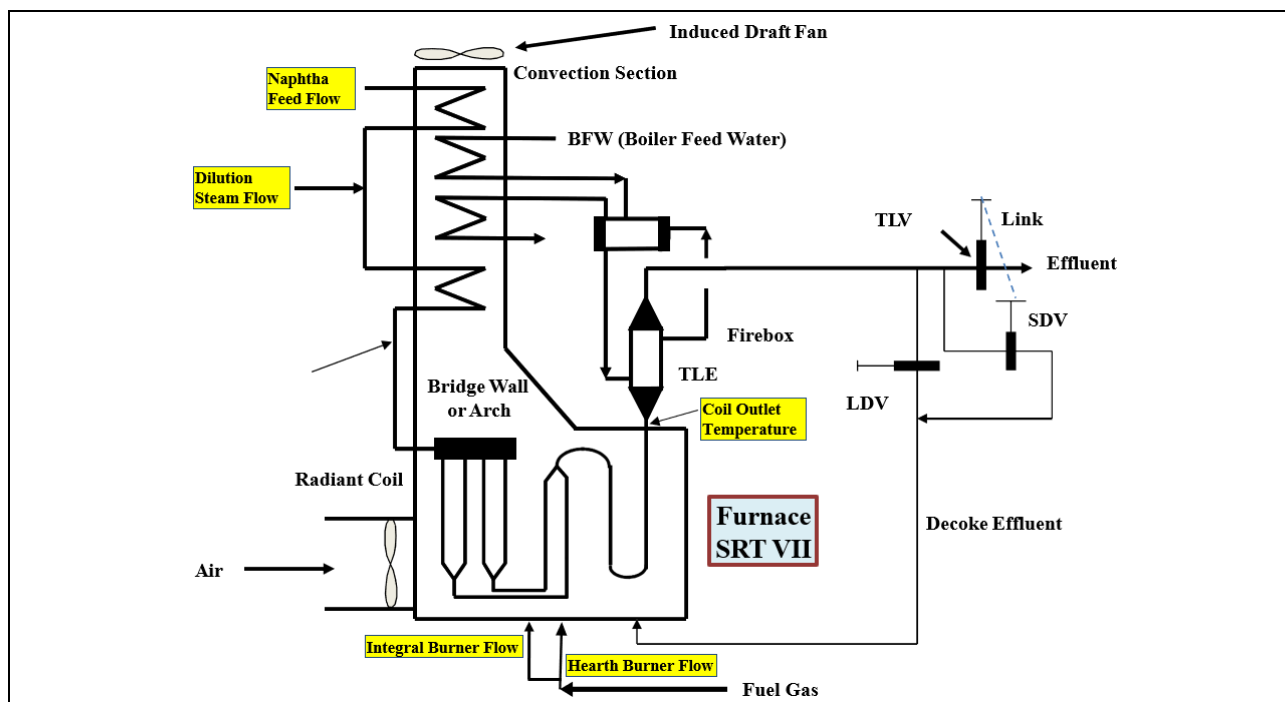
Olefin plant producing ethylene and propylene utilizing thermal cracking is often defined as the core of petrochemical manufacturing [7, 8] due to its significant impact on the industry. The furnace is the primary equipment in the olefin production industry [9] where its safe and stable operation is essential [10, 11] to determine the yield and quality of olefin [12] produced

Fig. 1 shows the configuration of the steam cracker furnace in the studied plant while Table 1 shows the naphtha feed specification utilized in the steam cracker furnace during the study.

The process starts from naphtha introduction into the steam cracker furnace at the first convection bank and mixes with Dilution Steam (DS) in the middle bank. The mixing is designed to improve olefin selectivity mainly by reducing the partial pressure of naphtha feed [13, 14]. The reduced partial pressure from the introduction of DS will favor the reversible reaction towards the olefin side following the Le' Chatelier's Principles [15, 16] and therefore will improve the Propylene Yield significantly.

The mixed feed of naphtha and DS will be further cracked in the furnace coils at the radiation section that is operated at elevated Tube Metal Temperature (TMT) of 1,050 °C - 1,180 °C, with the controlled short residence time (SRT) of <2 seconds. This condition will maximize light component yields such as ethylene and propylene product. This mixture which is now called a cracked gas continues being processed at the Transfer Line Exchanger (TLE) to rapidly cool the cracked gas effluent. The Boiler Feed Water (BFW) was utilized as a cooling medium in the TLE.

\*Corresponding author: [mohamadhafizi.zakri@prefchem.com](mailto:mohamadhafizi.zakri@prefchem.com)



**Fig. 1.** General arrangement for SRT VII Furnace in the studied plant with tags and descriptions of the chosen variables for the study

**Table 1.** Naphtha feed specification during the study

Parameter	Analysis Result	
	Unit	Value
Reid Vapour Pressure	kPa	44.5
Density	kg/L	0.7248
Paraffins	% vol	60.92
Olefins	% vol	1.02
Naphthenes	% vol	25.97
Aromatics	% vol	12.09
TBP Distillation Curve		
Initial Boiling Point	°C	34.1
30 vol - %	°C	84.4
50 vol - %	°C	105.1
70 vol - %	°C	125.8
Final Boiling Point	°C	166.1
Sulfur	ppm wt	298.5

The cracked gas that is quenched in the TLE will continue going to the downstream equipment for further quenching, compression, cooling, and product separation. Super High Pressure (SHP) steam generated from TLE will be used to drive Charge Gas Compressor (CGC) turbine at the downstream process.

Linear regression is a well-known statistical approach used to correlate the relationship between various input parameters towards achieving the target output systematically. It was proven successful from previous studies [6, 17-19] in showing the relationship of each input variable towards the output variable. It is also practical to be used for the specific process monitoring and planning including for the large-scale olefin plant.

A Box-Cox transformation is a way to transform non-normal dependent variables into a normal shape to give a broader number of tests. This model proposed by Box and Cox [20], aimed at reducing anomalies in data and ensuring the usual assumption for a linear model hold [20, 21]. This transformation results from modification of power transformation which account for discontinuity at  $\lambda = 0$  [22, 23]. Box-Cox Transformation is also intensively studied and useful to be used as a fundamental tool in the Regression analysis [22, 23].

## 2 Methods and Equipment

### 2.1. Equipment/Tools

The study was conducted utilizing the industrial scale SRT VII furnace and its auxiliaries including Steam Drum, Induced Draft Fan (IDF), Burners, and TLEs. The studied furnace was designed to process 93 t/hr of straight-run naphtha liquid from the upstream plant. The studied plant was carefully designed by an established Olefin Licensor, Lummus Technology Heat Transfer (LTHT) from the United States. The site construction was completed by Toyo Engineering, Japan.

The data collection was conducted using Process Information Management System (PIMS) Software, PI Process Book Version 2015. The analysis of these data was conducted using the Regression analysis in Minitab Software Version 18.

### 2.2 Methodology

5 controlled variables namely Hearth Burner, Integral Burner, Dilution Steam Flow, Naphtha Feed Flow, and

\* Corresponding author: [mohamadhafizi.zakri@prefchem.com](mailto:mohamadhafizi.zakri@prefchem.com)

Coil Outlet Temperature were chosen as the input to the Propylene Yield in the studied plant. The data for the analysis was extracted hourly (time-weightage) from PI Process Book. The date was selected on 24<sup>th</sup> January 2020, 1900 hrs until 2<sup>nd</sup> February 2020, 1200hrs (207 hrs total). This resulted in a total of 1,242 data records for the analysis (established from 1 output and 5 input variables).

The analysis was conducted based on actual plant conditions analyzed from these 1,242 historical plant data used for the Regression model. Therefore, the model established from this data was representing the current plant condition at the studied plant. Besides, this study also emphasized the model development for Propylene Yield with Naphtha Feed specification as shown in Table 1. Therefore the model established was limited to best represent the process condition with feed specification range close to Table 1. The huge variation of Paraffins, Olefins, Naphthenes, and Aromatics (PONA) composition from Table 1 may require further assessment to be conducted for the establishment of the new regression model at different PONA compositions.

The verification on the data stability was conducted using Box Plot, Individual-Moving Range (I-MR), and Run Chart while verification on data normality was established using Graphical Summary and Normality Test in Minitab Software. These steps were taken prior decision to conduct analysis utilizing Box-Cox Transformation. Box-Cox Transformation will be applied to the Regression analysis if the majority of these data tests were found unstable and non-normal.

This study was conducted without thorough outlier removal and data cleaning since it was intended to see the impact of Regression using Box-Cox transformation for the non-normal data. The P-Value for both stability and normality check should be below 0.05 for the analysis to proceed with Box-Cox Transformation.

Once the Regression analysis started, the analysis of the data was conducted to ensure the Variance Inflation Factor (VIF) value was lower than 10 and P-Value was less than 0.05 for all studied variables.

VIF is useful to quantify the severity of the multicollinearity relations in ordinary least squares of the Regression analysis. VIF value is recommended to be <10 [24] or <5 [25] to reduce the multicollinearity in the

final model. The high VIF in the Regression analysis may result in the unreliable final equation model.

The Regression was conducted a few times until satisfactory VIFs and P-Values were achieved. Once achieved, the residual removal was conducted to the latest Regression model utilizing the Normality Plot and Individual Moving Range (I-MR) chart produced from the identified residuals in the Regression analysis.

Final Regression was reassessed to all variables (including previously excluded variables) after the successful removal of all residuals in the latest Regression. It can be seen from the I-MR Chart (no red point outlier) and Normality Plot (P-Value >0.05).

The Response Optimizer tool was applied to the final model to predict the maximum value of Propylene Yield with its process setting utilizing actual plant data and model developed from the Box-Cox Transformation.

Besides, the Surface Plot in form of a three-dimensional graph was also applied to see the relationship of each significant variable in the final model. It was helpful to correlate the response values for two continuous significant variables based on the final model equation towards the fitted response value, Propylene Yield.

### 3 Results and Discussion

The results of stability and normality verification utilizing Box Plot, I-MR, Run Chart, Graphical Summary, and Normality Test are shown in Table 2. From the analysis, all data from studied variables found non-normal except for Dilution Steam Flow. Therefore, Regression analysis was conducted for all input variables utilizing the data transformation technique using Box-Cox Transformation in Minitab Software. No outlier removal and data cleaning were conducted before the Regression analysis except for normal data verification (to remove irrelevant data where required).

The Regression analysis utilizing Minitab Software was conducted a few times until all VIFs and P-Values reading were satisfactory to develop the reliable mathematical model for Propylene Yield in the studied plant. Table 3 shows the initial (1<sup>st</sup>) and final (9<sup>th</sup>) Regression analysis results conducted during the study.

**Table 2.** Analysis result of stability and normality verification

Input / Output	Stability			Normality		Normality Result
	Box Plot (Outlier)	Run Chart (Fail Part)	I-MR (Residual)	Normality Test	Graphical Summary	
Propylene Yield	0	Cluster	4	<0.005	<0.005	No
Hearth Burner Flow	1	Cluster & Trend	4	<0.005	<0.005	No
Integral Burner Flow	13	Cluster & Trend	6	<0.005	<0.005	No
Dilution Steam Flow	0	None	1	0.347	0.347	Yes
Naphtha Feed Flow	0	Cluster	2	<0.005	<0.005	No
Coil Outlet Temperature	2	None	4	<0.005	<0.005	No

**Table 3.** Regression analysis result

Tag and Description		Initial Regression (1 <sup>st</sup> )			Final Regression (9 <sup>th</sup> )		
		Coefficient	P-Value	VIF	Coefficient	P-Value	VIF
Constant		-1595	0.000		322518	0.000	
1	Hearth Burner Flow	-0.003017	0.001	2.72	-0.000243	0.000	4.59
2	Integral Burner Flow	0.3159	0.000	1.35	0.01332	0.000	1.79
3	Dilution Steam Flow	0.00001	0.990	1.02			
4	Naphtha Feed Flow	1.586	0.000	2.23	0.08598	0.000	3.43
5	Coil Outlet Temperature	1.804	0.000	1.14			

From the analysis, VIF and P-Value in the 2<sup>nd</sup> Regression already met the criteria VIF <10 and P-Value <0.05. However, the R-Square in the 2<sup>nd</sup> Regression was achieved at only 54.43%. This translated to only 54.43% of the data was accounted into the model. However, this value was not resulted from utilizing Box-Cox Transformation, it was mainly due to the variation of data collected during the study period.

Residual removal was conducted to improve the R-Square of the model. These residuals represented the difference in actual data and predicted data calculated from the Regression model. The Unusual Observations in Fit and Diagnostic Table in Minitab Session Window and Residuals in I-MR Plot were removed from 3<sup>rd</sup> Regression until 9<sup>th</sup> Regression (Final).

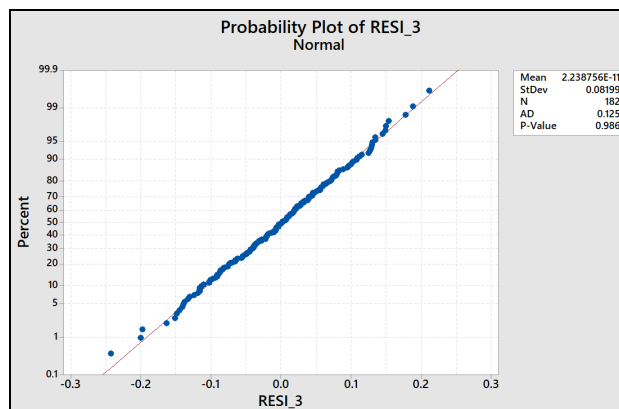
Fig. 2 and Fig. 3 show the Probability Plot and I-MR chart of residual for the 9<sup>th</sup> Regression (Final) respectively. The P-Value of 0.986 in Fig. 2 is higher than the minimum level of 0.05 and therefore this plot was accepted. Besides, there was also no residual detected outside of the Moving Range Plot in Fig. 3 which concluded that the final model was good in predicting the Propylene Yield in the studied plant.

The final model obtained from the 9<sup>th</sup> Regression is shown in Eq. (1).

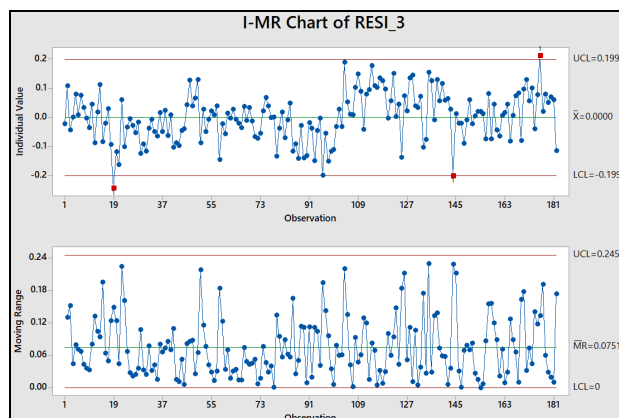
$$\begin{aligned}
 & (Propylene\ Yield^{\lambda-1})/(\lambda \times g^{\lambda-1}) = 322518 \\
 & - 0.000243 (HB\ Flow) + 0.01332 (IB\ Flow) \\
 & + 0.08598 (Naphtha\ Flow)
 \end{aligned}
 \tag{1}$$

Where,  $\lambda = -5$ ,  $g = 10.8290$  is the geometric mean of Propylene Yield

Table 4 shows the summary for the final model. Its R-square was found at 63.7%, which represented 63.70% of the data variability was accounted for in the final model. This value is lower than the advised 75% for experiments conducted in a controlled environment [26, 27]. However, there is no fixed rule for the acceptable R-Square value which depends on case to case basis. The value of 63.70 % was also good considering the study was conducted in the actual large-scale plant where process variations often occurred. The R-Square of this final model was also better compared to only 54.43% without residual removal in the 2<sup>nd</sup> Regression model.



**Fig. 2.** The Normality Test of residual for final Regression

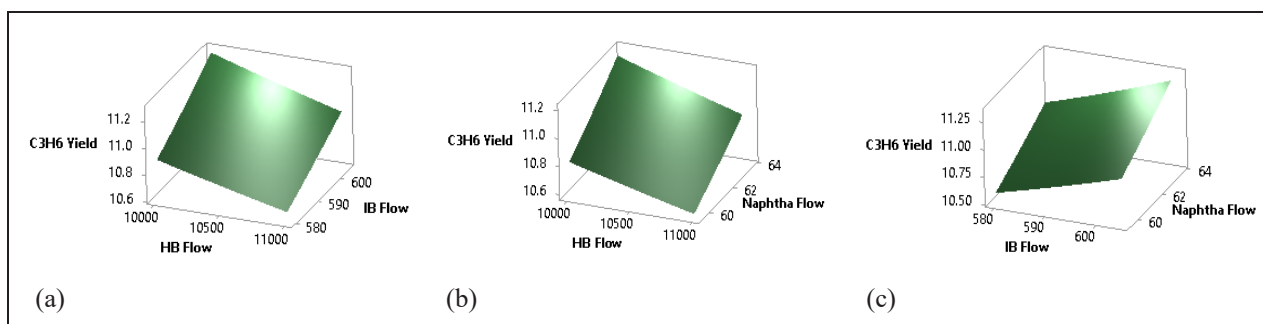


**Fig. 3.** The I-MR Chart for final Regression

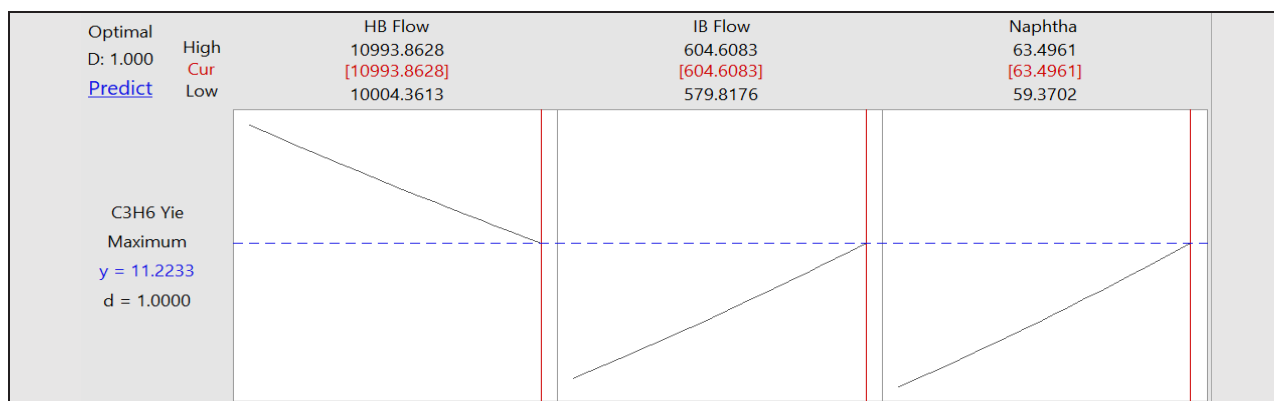
**Table 4.** Model summary

S	R-sq	R-sq (adj)	R-sq (pred)
0.0830881	63.70%	63.09%	61.99%

Fig. 4 shows the Surface Plot for the variables with the P-Value <0.05 obtained in the final regression model. The values for non-tested variables were held at mean value,  $\bar{x}$  of 10658.04, 585.96, and 61.26 for HB Flow, IB Flow, and Naphtha Flow respectively. The light on the Surface Plot also was set to the highest Propylene Yield achieved during the study.



**Fig. 4.** Surface Plot of Propylene Yield against significant variables of (a) HB Flow vs IB Flow, (b) HB Flow vs Naphtha Flow, (c) IB Flow vs Naphtha Flow



**Fig. 5.** Response Optimizer prediction for operating variables to achieve maximum Propylene Yield.

Fig. 4 (a) and Fig. 4 (b) show that manipulating lower operating HB Flow with higher IB Flow and Naphtha Flow may result in a higher Propylene Yield. Besides, Fig. 4 (c) shows that the higher IB flow combined with higher Naphtha Flow may also result in a higher Propylene Yield in the studied plant.

Overall, Fig. 4 suggested that manipulating IB Flow and Naphtha Flow are more favorable towards achieving a higher Propylene Yield in the studied plant. This analysis results also aligned with the model developed in Equation 1 earlier where IB Flow and Naphtha Flow contributed to the highest factor of 0.01332 and 0.08598 respectively towards Propylene Yield.

Fig. 5 shows the optimum process condition in achieving the maximum Propylene Yield from the identified significant variables while Table 5 shows the Multiple Response Prediction for the best Propylene Yield in the studied plant.

**Table 5.** Multiple Response Prediction

Response	Fit	Confidence	
		95% CI	95% PI
C <sub>3</sub> H <sub>6</sub> Yield	11.2233	(11.1188, 11.3340)	(11.0066, 11.4685)

The Response Optimizer result showed that Propylene Yield in the studied plant can be maximized at 11.22% with the optimized value of 10,993.86 kg/hr of Hearth Burner Flow, 604.61 kg/hr of Integral Burner Flow, and 63.50 t/hr of Naphtha Feed Flow. The range

of the process condition to maximize the Propylene Yield with a 95% confidence level also can be taken from the High and Low range settings in Fig. 5. This setting is helpful as a guide for Panel Operators and Operations Engineers in the studied plant to maximize the Propylene Yield.

## 4 Conclusion

The Propylene Yield model was successfully established in this study using the Box-Cox Data Transformation in the Regression Analysis. The maximized Propylene Yield that can be achieved from the model was identified at 11.22% which can be obtained by careful monitoring to the significant variables which were Hearth Burner Flow, Integral Burner Flow, and Naphtha Feed Flow.

The authors are grateful to the fellow process engineers at Steam Cracker Unit, Pengerang Refining Company Sdn Bhd for the strong technical supports. Besides, special thanks are also dedicated to the School of Chemical and Energy Engineering, Universiti Teknologi Malaysia for the academic advice in completing this study.

## References

1. S. Vangaever, P.A. Reyniers, S.H. Symoens, N.D. Ristic, M.R. Djokic, G.B. Marin, K.M. Van Geem, Chem. Eng. Res. Des. **153**, 380-390 (2020).
2. R. Van de Vijver, N. Vandewiele, P. Bhoorasingh, B. Slakman, F. Seyedzadeh Khanshan, H.H. Carstensen, M.F. Reyniers, B. Marin, R. West, K.

\*Corresponding author: [mohamadhafizi.zakri@prefchem.com](mailto:mohamadhafizi.zakri@prefchem.com)



- Van Geem, Int. J. Chem. Kinet. **47**, 199-231 (2015).
3. H. Shi, C. Su, J. Cao, P. Li, J. Liang, G. Zhong, Ind. Eng. Chem. Res. **54**, 1849-1860 (2015).
  4. H. Song, C.-l. Su, H. Shi, P. Li, J.-t. Cao, Meas. Control. **52**, 526-539 (2019).
  5. Z. Feli, A. Darvishi, A. Bakhtyari, M.R. Rahimpour, S. Raciissi, J. Taiwan Inst. Chem. Eng. **81**, 1-13 (2017).
  6. M.H. Zakria, A.A. Omar, M.A. Bustam, Procedia Eng. **148**, 561-567 (2016).
  7. T.-J. Fan, R. Luo, H. Xia, X. Li, Nat. Hazards. **75**, 319-332 (2015).
  8. I.K. Nikolaidis, L.F.M. Franco, L.N. Vechot, I.G. Economou, Fluid Phase Equilib. **470**, 149-163 (2018).
  9. Z. Junfeng, P. Zhiping, C. Delong, L. Qirui, H. Jieguang, Q. Jinbo, IEEE Access. **7**, 158643-158654 (2019).
  10. Z. Peng, J. Zhao, Z. Yin, Y. Gu, J. Qiu, D. Cui, Processes. **7**, 909 (2019).
  11. L. Shen, J. Gong, H. Liu, Appl. Mech. Mater. **750**, 192-197 (2015).
  12. Z. Wang, Z. Li, Y. Feng, G. Rong, Can. J. Chem. Eng. **94**, 1723-1739 (2016).
  13. H. Karimi, E. Cowperthwaite, B. Olayiwola, H. Farag, K. McAuley, Can. J. Chem. Eng. **96**, 33-48 (2017).
  14. M. Masoumi, S.M. Sadrameli, J. Towfighi, A. Niaei, Energy. **31**, 516-527 (2006).
  15. Y. Cai, S. Yang, S. Fu, D. Zhang, Q. Zhang, J. Mater. Sci. Technol. **33**, 580-586 (2017).
  16. L.G. Epstein, J. Econ. Theory. **19**, 103-122 (1978).
  17. G. Light, Global J. Sci. Front. Res. **20**, 1-6 (2020).
  18. D.S. Patnana, G. Hitesh, I.N.S. Kumar, J. Crit. Rev. **7**, 914-917 (2020).
  19. D. Takagi, T. Shimada, Front. Psychol. **10**, 2799 (2019).
  20. L. Broemeling, *Box-Cox Transformation - Introduction* (John Wiley & Sons, Inc., Hoboken, New Jersey, 2014).
  21. T. Zhang, B. Yang, Technometrics. **59**, 189-201 (2016).
  22. A. Atkinson *The box-cox transformation: review and extensions*. LSE Research Online Documents on Economics 103537, 2020.
  23. J. Osborne, Pract. Assess. Res. Evaluation. **15**, 1-9 (2010).
  24. J.F. Hair, R.E. Anderson, R.L. Tatham, W.C. Black, *Multivariate Data Analysis with Readings* (Prentice-Hall, New Jersey, United States, 1995).
  25. C. Ringle, S. Wende, J.-M. Becker *SmartPLS 3*. 2015.
  26. P.D. Haaland, *Experimental design in biotechnology* (Marcel Dekker, New York, United States, 1989).
  27. W.N.N. Wan Omar, N. Nordin, M. Mohamed, N.A. Saidina Amin, J. Appl. Sci. **9**, 3098-3103 (2009).