

Refined Palm Oil Product Quality Predictor for Supporting Palm Oil Refinery Energy Management System

Azmer Shamsuddin^a, Nor Adhahah Rashid^b, Mohd Kamaruddin Abd Hamid^c, Norazana Ibrahim^{b,*}

^a Lahad Datu Edible Oils Sdn. Bhd., KM 2, Jalan Minyak Off Jalan POIC, Locked Bag No. 16, 91109 Lahad Datu, Sabah, Malaysia

^b School of Chemical and Energy Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia

^c JMH Integrated Services Sdn. Bhd., 34, Jalan PI 10/3, Taman Pulai Indah, 81300 Johor Bahru, Johor Malaysia. norazana@utm.my

As a part of the smart palm oil refining purported to be the factory for the future achieving Industry 4.0 targets, smart quality prediction tools have been developed to minimise current hourly manual sampling practice. The quick and excellent forecasted quality of the refined, bleached and deodorized palm oil (RBDPO) on daily basis will reduce the rework of the off-spec products. The study aims to develop RBDPO quality forecasting model. It began with data collection, followed by a pre-processing stage to acquire the optimum sampling time and the processing time of the refining process using statistical tools such as boxplots, histograms, autocorrelation and cross-correlation plots. Using the pre-processed data, the predictor coefficients are then developed using various multivariate statistical analysis methods such as Partial Correlation Analysis (PCorrA), Principal Components Analysis (PCA), Partial Least Square (PLS), and Principal Components Regression (PCR) algorithms with the help of MATLAB programming software, and the forecasted data are being plotted together with the actual real time data in control charts to assess the refining process performance of Lahad Datu Edible Oils Sdn. Bhd. (LDEO). For the 327 sample size data, the sampling frequency is reduced by 75 % as product sampling time carry out at every 4 h. The residence time selected at 8 h. Through mean squared error (MSE) computations, PCorrA showed consistently low MSE readings of 0.0000386, 0.000014, 0.0036 and 0.04531 for FFA, MC, IV and COL. With proper energy management system, energy saving of 9 %, 9.5 %, 10 % and 10 % were registered for steam, LNG gas, electricity and water with the implementation of PCorrA predicting model. PCorrA is selected as the best forecasting algorithm which enables a systematic refining process monitoring and raw materials planning as well as supporting the palm oil refinery energy management system.

1. Introduction

The variation in CPO quality and poor energy management can determine the production quality of Refined Bleached Deodorised Palm Oil (RBDPO) whereby the plant interruption may lead to frequent oil rejection or recycling. Currently, the quality analysis can only be done after product passed through deodorizing stage which total processing time lasted for up to 6 h. The corrective action at various processing stages will lead to downtime in and loss of production. The lengthy recycling of off-specs products will increase energy usage and processing cost which affected the refinery's profitability. In order to produce consistent quality product, real-time process monitoring should be conducted to identify and to rectify the unusual variability promptly.

In this study, much better than other predicting tools such as PLS, PCA and PCR. The intended objective for this study is to forecast RBDPO quality using multivariate statistical analysis helps in monitoring production planning and reliability process improvement. This enables a systematic refining process monitoring and raw materials planning as well as supporting the palm oil refinery energy management system. The implementation of the prediction of RBDPO quality ensures the refinery to stay competitive in years to come.

1.1 Partial Least Square (PLS)

PLS is a widely used technique which combines key features from PCA and multiple linear regressions (MLR) (Abdi, 2003). PLS is a quick and efficient regressing method based on covariance. It is recommended when high and correlated explanatory variables are present. Many applications of PLS regression have appeared over the years for prediction and estimation in the manufacturing industry including chemical, pharmaceutical, and semiconductor processes (Madanhire and Mbohwa, 2016). Moh (2017) recently has conducted a study on predicting RBDPO quality using PLS and the results show that PLS was proven to be the better prediction model against least squares regression (LSR) and average of range (AVE) methods. In a similar study, Lei (2017) proved PLS regression method performed better than Kernel PLS (KPLS) regression method.

1.2 Principal Components Analysis (PCA)

PCA is also usually used to expedite the computation by dimension reduction of data. PCA can improve the accuracy of data with high dimensionality and high correlated variables of one another (Jolliffe, 2002). PCA's two main objectives are identification of new meaningful variables and reduction of dimensionality of the problem as a prelude to further analysis of data (Onwuka, 2012). The objective of PCA is to decompose a data table with correlated measurements into a new set of uncorrelated variables. PCA is interpreted as an orthogonal decomposition of the variance of a data table (Abdi, 2003). The goal of PCA is to simplify the complexity of the data. If a large number of factors are needed to define the dimensionality of the data, then there is very little need for PCA. Hawi (2017) conducted RBDPO prediction study using Hotelling T²-PCA and the results show that Hotelling T²-PCA has better prediction performance compared to LSR and AVE regression methods. In a similar study, Nair (2017) proved PCA performed better as it has the minimum MSE against LSR and AVE regression methods.

1.3 Principal Component Regression (PCR)

PCR is a technique for analysing multiple regression data that suffer from multicollinearity. When multicollinearity occurs, least squares estimates are unbiased, but their variances are large so they may be far from the true value (Hernandez-Arteseros et al., 2000). PCA alone is not capable to solve multicollinearity of regressors. PCR performed better with the implementation of PCA which eliminates data multicollinearity. According to Fekedulegn et al. (2002), the independent variables are transformed into a new set of orthogonal or uncorrelated variables known as principal components (PC). This transformation ranks the orthogonal variables in order of their importance according to the eigenvalues and involving the elimination of less important PCs. Following the PC's elimination, multiple linear regression analysis of response variable against the reduced set of principal components is performed to obtain the regression coefficient (Ul-Saufie et al., 2011). Additional of the nonlinearity analysis in PCR improved the prediction performance. Lelavathy (2018) proved in her prediction study that kernel PCR (KPCR) method has better prediction than PCR method.

1.4 Partial Correlation Analysis (PCorrA)

PCorrA is a technique used to develop correlation coefficient between the process variables and selected predicted variables. The presence of deviation or outlier may give inaccurate results for the predictor (DeVor et al., 1992). Noslan (2017) has carried out RBDPO quality prediction using PCorrA technique. PCorrA showed the largest MSE against AVE and LSR regression methods due to outliers presented in the data.

2. Methodology

The research methodology including data collection method, data analysis, measure and statistical analysis. The methodology framework covers the major stages of the study listed in hierarchical order along with their respective subordinate procedures. In order to verify the performance of the four predicting engines, dimension reduction was carried out to avoid over-fitting of data which may contribute to inaccuracy of results. The optimized 23 variables out of total 75 predictor variables with eigenvalue more than one were retained as shown in Table 1. The stages listed in the framework include the Pre-Screening stage, Pre-Processing stage, Sampling Time Identification stage, Resident Time Identification stage, the Development stage of RBDPO Quality Forecasting Models, the Process Performance Analysis stage using Statistical Process Control (SPC) Charts, and ultimately, the Validation and Verification stage.

2.1 Pre-processing data

Data pre-processing was carried out in an iterative fashion which comprises of three main stages, namely data standardization, sample size and sampling time determination through boxplots, histograms (to achieve normality), autocorrelation plots (to achieve randomness), and processing time determination through cross-correlation plots. To establish the prediction model, four selected methods were implemented including the PCR,

PLS, PCorrA and PCA. The prediction models operated with predictor coefficient, k in predicting the values of RBDPO properties by only referring to the selected process variables of CPO. During the development of forecasting models to predict the properties of RBDPO, predictor coefficient k is used to generate forecasted response.

Table 1: PC-Dimension Reduction trimmed to 23 Variables

No.	Variables	No.	Variables
1	FFA	13	DEODORIZER D 302 (Temperature 5)
2	IV	14	PACK COLUMN (Level)
3	CPO FEED OIL (Pressure)	15	BPO FLOWRATE (ton/hr)
4	BIO TANK (Temperature)	16	BPO FLOWRATE (ton)
5	DRYER (Level)	17	PFAD (ton)
6	BIO TANK (Level)	18	DEODORIZER (Pressure)
7	CPO FLOWRATE (ton/hr)	19	VACUUM (Pressure)
8	NIAGARA FILTER (Pressure 4)	20	GEKA BOILER (Pressure)
9	NIAGARA FILTER (Pressure 5)	21	DEODORIZER (Pressure 2)
10	SPIRAL (Temperature 1)	22	DEODORIZER (Pressure 6)
11	RBDPO (Temperature)	23	DEODORIZER (Pressure 7)
12	DEODORIZER D 302 (Temperature 4)		

2.2 Development of quality prediction tool

After preliminary steps were done, predictor models are developed to attain the correlation constant between the independent input variables and the dependent output variables of a training data set. The mean squared errors and standard estimated errors of the respective prediction were calculated to determine the deviation of the estimated values with the actual values. The theory of computing PCorrA method, for example, is shown in Eq(1) and Eq(2) (multivariate regression).

$$k(X, Y : Z) = \frac{k(X, Y) - k(X, Z)k(Y, Z)}{\sqrt{[(1 - k^2(X, Z)][1 - k^2(Y, Z)]}} \quad (1)$$

$$\hat{y}_i = k_1x_{i,1} + k_2x_{i,2} + k_3x_{i,3} + \dots + k_px_{i,p} \quad (2)$$

where k is predictor coefficient vector; X and Y are the matrix of samples and predicted variables; Z is the matrix of intervening variables.

2.3 Mean Square Error (MSE)

The aim of applying different methods was to compare the results of respective models and the most effective models amongst the four selected methods were decided as the best predicting tool. The prediction models were established along with the control charts as the control chart is a significant tool in visualizing the trend of the predicted data comparing with the actual data and also providing the user information on the properties that are within or exceed the range of specifications according to the standards of China, PORAM and Vietnam. In all statistical prediction models, predicted variables relationship to independent variables is shown in Eq(3).

$$Y = kX \quad (3)$$

where, Y is predicted output data (FFA, MC, IV, COL); X is actual input data; k is correlation coefficient. As a comparison to find the best method, error of prediction is computed by using formula of Mean Square Error (MSE) as shown in Eq(4). The nearer the error to zero, the lower the error.

$$\text{MSE \%} = \frac{\sum (Y1 - y1)^2}{n} \times 100 \% \quad (4)$$

where, MSE is Mean Square Error; $Y1$ is Predicted output data (FFA, MC, IV, COL); $y1$ is actual output data (FFA, MC, IV, COL); n is sample size

3. Results

The initial data sets contained 327 data, with the interval of 30 min. A sample size of 25 was determined the best sample size after carried out the boxplot and histogram plots. As per Figure 1 for auto-correlation, the biggest number of lags is chosen as the best sampling time for the RBDPO process because it gives the best

repeating patterns over the various time intervals. The biggest lags identified for this study is eight lags as selected from FFA out. The proposed sampling time interval from this analysis is now set at 4 h. For cross-correlation, two (2) lags have been chosen as the actual lag to compute the optimum residence time. Since the previous optimum sampling time is 4 h, the residence time for this RBDPO process is doubled of the sampling time which is 8 h. The residence time selected in this study is 8 h. Rosely et al. (2017) in their study obtained sampling time of 2 h and residence time of 2.4 h. The lower value of residence time for cross-correlation explained that the data was taken when the plant was running under different operating capacity.

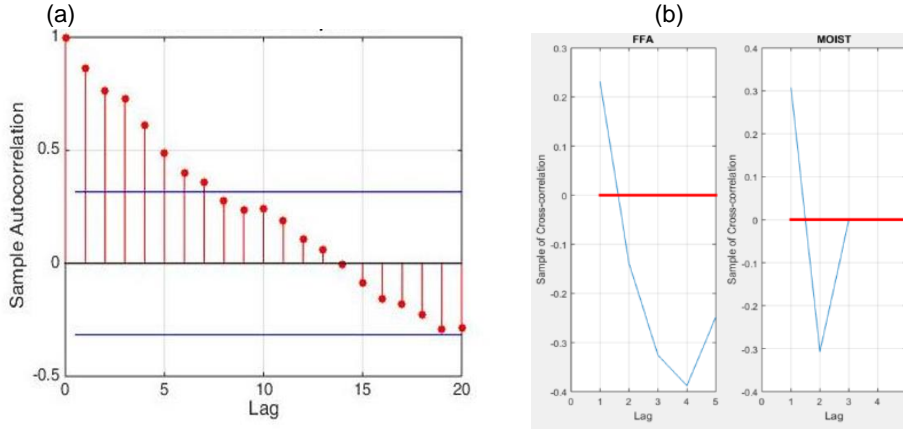


Figure 1: (a) Autocorrelation plot, and (b) Graphs of cross-correlation plot

The MSE values of prediction conducted using all four regression methods are in Figure 2. It can be clearly seen that the MSE of PCorrA method was much closer to zero in both training and testing data sets as compared to PCR, PLS, and PCA methods. This indicates that prediction of PCorrA had lesser deviation from the real data. PCorrA method can predict the quality of RBDPO the best in this case. PCA can be the alternative option to be used as predicting tool as it exhibited the second-best performance from both training and testing data sets. PCR and PLS predicting models are not reliable RBDPO quality predictors.

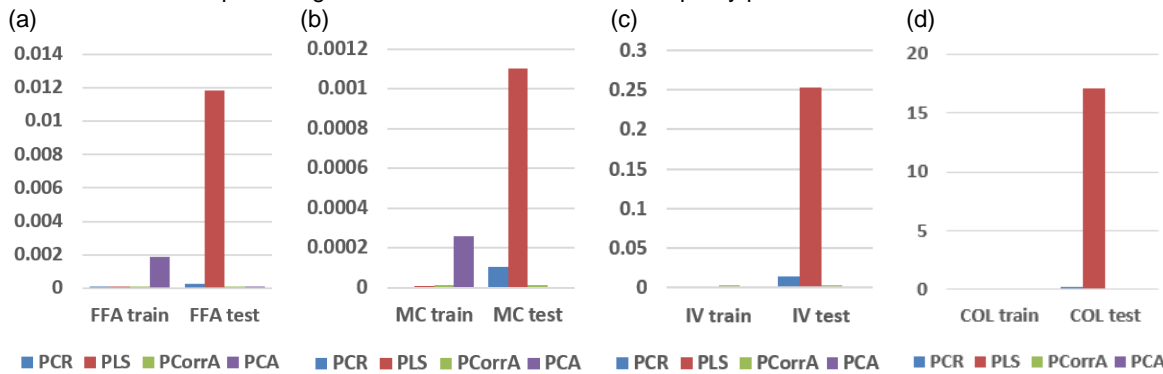


Figure 2: MSE comparison for four predictors based on training and testing data sets (a)FFA, (b)MC, (c)IV and (d)COL

One criteria that used to choose the best predictor model is the consistency of the MSE values for training and testing data sets. As can be clearly seen from Figure 2, for the PCorrA predicting model, the MSE values of the testing data are quite close to the training data, it can be concluded that the developed PCorrA predicting model is reliable and suitable in predicting quality parameters of RBDPO. The deviation of MSE between testing and training data is still comparable and consistent, the PCorrA prediction model is proven to be valid. PCorrA model has high sensitivity to the presence of outliers. The presence of outliers may lead to inaccurate analysis and conclusion. Further validations have been carried out to confirm and validate the best prediction model performance based on several new sets of testing data. The data testing for predictor performance of PCorrA is based on data with smooth operation that compiled with no plant interruption and no abnormality, data with present of outliers that compiled with some quality parameters exceeded the specification and data with plant interruption/recycling that compiled during plant interruption or rework of product. Figure 3 shows the calculated MSE values of the PCorrA predictor using three new testing data against the training data. The calculated MSE

values remain close to zero using new testing data from three different plant conditions discussed earlier. This validates that the prediction of PCorrA had lesser deviation from the real data, the best prediction model. The deviation of MSE between testing and training data is still comparable and consistent, the PCorrA prediction model is proven to be valid. This finding showed contradiction result against the earlier study done by Noslan (2017) that PcorrA did not perform as good as LSR and AVE due to outliers present in the data.

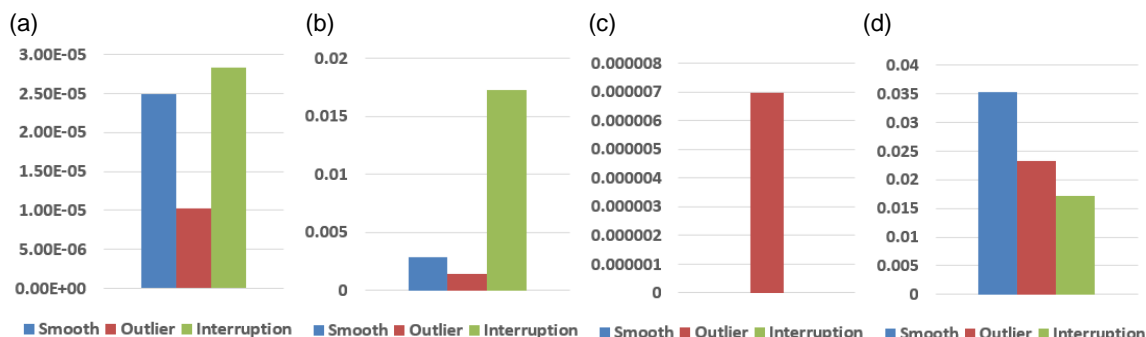


Figure 3: Calculated MSE of PCorrA predictor using new testing data against training data (a)FFA, (b)MC, (c)IV and (d)COL

In general, the predictor model developed using PCorrA method has a better capability in predicting the RBDPO quality parameters with or without the presence of outliers in the data. This shows that the predictor is robust and can be used as a prediction tool for RBDPO quality prediction in daily operation in LDEO. The setting process parameter can be adjusted to suit the required quality grades for various market including PORAM, China and Vietnam standards. By having the robust and effective quality predictor, the physical refining plant can now be operated smoothly and without the need of slowing down the plant. The recycling and rejection can now be monitored easily. Monitoring the product recycle and rejection is the important part in the palm oil refinery energy management. Various quality grades of oil are produced by varying the feed oil flowrate, steam sparging pressure, deodorising temperature setting and bleaching earth setting which will reflect the consumption of energy from steam, water, electricity and gas. With proper energy management system, as per Table 2, energy saving of 9.0 %, 9.5 %, 10.0 % and 10.0 % were registered respectively for steam, LNG gas, electricity and water with the implementation of PCorrA predicting model.

Table 2: Comparison of specific energy consumption before and after utilizing PCorrA predictor model

Parameter	Unit	Before	After	% Saving
Steam	kg/mt CPO	88.0	80.0	9.0
LNG Gas	MMBTU/mt CPO	4.2	3.8	9.5
Electricity	kWh/mt CPO	10.0	9.0	10.0
Water	kg/mt CPO	100.0	90.0	10.0

It should be noted that processing of bad quality of CPO causes high consumption of energy in steam and electricity as well as longer retention time. The frequent of oil rejection or recycling is higher due to disparity and variation in CPO quality. In the refinery, processing poor quality oil can be done by slowing down the plant throughput and by dosing higher chemical and increasing energy usage in order to achieve the required RBDPO quality. Predicting of RBDPO product quality and performance optimization are needed to overcome these limitations so that proper managing or blending can be done on CPO prior to processing. The usage of steam, bleaching earth, and phosphoric acid can be estimated and predicted as well prior to processing. With the earlier prediction of expected product quality, the refinery plant operating parameters can be monitored according to the quality of the feed oil in the storage tank. The optimal plant performance can be planned to achieve RBDPO with better quality and higher plant productivity. The oil recycling during production due to quality upset can be minimized to ensure higher throughput.

4. Conclusion

The best sample size has been successfully obtained from the raw quality and operation data from the refinery. The optimum sampling time and residence time of the refinery process can be obtained from the autocorrelation and cross-correlation plots. The optimized 23 variables were selected out of total 75 predictor variables through

dimensional reduction step from input quality properties and plant's process variables to fine-tune the analysis for the best results. The four statistical prediction models have been effectively developed based on the obtained optimum sampling time and residence time. The four developed statistical prediction models have been individually trained and tested using training and testing data sets, where the prediction performances in terms of prediction stability and quality trends tracking were compared based on plotted control charts and calculated MSE values. From all results, it can be confirmed that the developed PCorrA model is reliable and suitable in predicting quality parameters of RBDPO which enables a systematic refining process monitoring and raw materials planning as well as supporting the palm oil refinery energy management system. It is recommended for future research be conducted to study comprehensive energy saving and quality improvement in the refinery by using PCorrA as a predicting tool.

Acknowledgments

The authors would like to acknowledge the financial support by Universiti Teknologi Malaysia (R.J130000.7351.4B572).

References

- Abdi H., 2003, Multivariate Analysis, Program in Cognition and Neurosciences, The University of Texas at Dallas, USA, 1-4.
- DeVor R.E., Chang T.H., Sutherland J.W., 1992, Statistical Quality Design and Control: Contemporary Concepts and Methods, Macmillan, New York, USA.
- Fekedulegn D., Colbert J.J., Hicks R.R., Shukers M., 2002, Coping with Multicollinearity: An Example on Application of Principal Components Regression in Dendroecology, Northeastern Research Station, Research Paper NE-721, 1-48.
- Hawi A.N.D., 2017, Refined Bleached Deodorized Palm Oil Quality Prediction Using Hotelling T²-PCA, Bachelor Degree Thesis, Universiti Teknologi Malaysia, Johor, Malaysia.
- Hernandez-Arteseros J.A., Compano R., Ferrer R., Prat M.D., 2000, Application of principal component regression to luminescence data for the screening of ciprofloxacin and enrofloxacin in animal tissues, *The Analyst*, 125, 1155-1158.
- Jolliffe I.T., 2002, Principal Component Analysis (2nd edition), Springer-Verlag, New York, USA.
- Lelavathy S.M., 2018, Refined Bleached Deodorized Palm Oil Quality Prediction Using Non-Linear (Kernel) Principal Component Regression Technique, Bachelor Degree Thesis, Universiti Teknologi Malaysia, Johor, Malaysia.
- Lei T.Y., 2017, Predictor of Refined Bleached Deodorized Palm Oil Quality Using Kernel Partial least squares regression, Bachelor Degree Thesis, Universiti Teknologi Malaysia, Johor, Malaysia.
- Madanhire I., Mbohwa C., 2016, Application of Statistical Process Control (SPC) in Manufacturing Industry in a Developing Country, *Procedia CIRP*, 40, 580-583.
- Moh Y.H., 2017, Refined Bleached Deodorized Palm Oil Quality Prediction Using Partial Least Squares Regression Analysis Technique, Bachelor Degree Thesis, Universiti Teknologi Malaysia, Johor, Malaysia.
- Nair P., 2017, Refined Bleached Deodorized Palm Oil Quality Prediction Using PCA Technique, Bachelor Degree Thesis Universiti Teknologi Malaysia, Johor, Malaysia.
- Noslan M.H., 2017, Refined Bleached Deodorized Palm Oil Quality Prediction Using Partial Correlation Analysis Technique, Bachelor Degree Thesis, Universiti Teknologi Malaysia, Johor, Malaysia.
- Onwuka G.I., 2012, Hotelling T-square & Principal Components Analysis Approaches to Quality Control Sustainability, *International Journal of Computational Engineering Research*, 2, 211-217.
- Rosely N.A.M., Rashid N.A., Noor M.A.M., Hawi N.D.A., Sepuan S.Q., Shamsuddin A., Ibrahim K.A., Hamid M.K.A., 2017, Product sampling time and process residence time prediction of palm oil refining process, *Chemical Engineering Transactions*, 56, 1411-1416.
- Ul-Saufie A.Z., Yahya A.S., Ramli N.A., 2011, Improving multiple linear regression model using principal component analysis for predicting PM₁₀ concentration in Seberang Prai, Pulau Pinang, *International Journal of Environmental Sciences*, 2(2), 415-422.