CHEMICAL ENGINEERING TRANSACTIONS
VOL. 89, 2021
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.cetjournal.it
Guest Editors: Jeng Shiun Lim, Nor Alafiza Yunus, Jiří Jaromír Klemeš

ISBN 978-88-95608-87-7; **ISSN** 2283-9216
DOI: 10.3303/CET2189026

Quality Prediction Improvement through Adaptive Nonlinear Principal Component Regression

Nor Adhihah Rashid[a], Azmer Shamsuddin[b], Wai Hong Khu[a], Muhammad Hisyam Lee[c], Norazana Ibrahim[a,*], Mohd Kamaruddin Abd Hamid[d]

[a] School of Chemical and Energy Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 Skudai. Johor Bahru, Johor, Malaysia.
[b] Lahad Datu Edible Oils Sdn. Bhd., KM 2, Jalan Minyak Off Jalan POIC, Locked Bag No. 16, 91109 Lahad Datu, Sabah, Malaysia.
[c] Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 Skudai. Johor Bahru, Johor, Malaysia.
[d] JMH Integrated Services Sdn. Bhd., 34, Jalan PI 10/3, Taman Pulai Indah, 81300 Johor Bahru, Johor Malaysia.
norazana@utm.my

The purpose of this paper is to present the predictor improvement for the refined palm oil quality based on the adaptive multivariate statistical process control. The time-varying behaviour of the palm oil refinery process has made it difficult for the industrial personnel to monitor and sustain the production of high quality refined palm oil. It will be very costly for the palm oil industries to repeat the refining process for the low quality refined palm oil, to meet the customer's preference in the market. Alternatively, the quality of the refined palm oil can be measured before the process through a systematic quality monitoring, where the information from the quality analysis and process condition is integrated to develop an efficient quality prediction tool. The hybrid of time-series expansion methods namely recursive window (RW) analysis and exponentially weighted recursive window (EWRW) analysis along with the nonlinear principal component regression based on the nonlinear iterative partial least squares algorithm (NIPALS-PCR) are proposed to develop the adaptive prediction model. The predictor coefficient is then used to predict the refined palm oil quality based on the input quality and process variables. Through the validation with the online data, both NIPALS-PCR RW and NIPALS-PCR EWRW perform better than the NIPALS-PCR static, where the prediction is improved by 95 %.

1. Introduction

In real industrial process, accurate prediction of quality variables is important for process control, decision making, safety and reliability of industrial processes (Xu et al., 2017). Due to the complexity of batch process, natural variation of raw material, process setting for different quality production and chemical interactions at different time-series may degrade the quality of the final product (Zhang et al., 2015). The weather and other natural factor (e.g., rain, soil acidity, etc.) may affect the quality of Crude Palm Oil (CPO), and consequently affecting the production of high quality Refined Bleached Deodorized Palm Oil (RBDPO).To maintain the RBDPO quality during the refinery process, an accurate quality control is needed.

In current practices, the RBDPO quality can only be determined via chemical analysis after the end product exit the refining process. If the RBDPO quality did not satisfied the specification, the refining plant opts to go for the recycle to reprocess the off-specification RBDPO which eventually increasing the processing cost and lead to processing delay. Alternatively, the recycle stream can be prevented using quality prediction tool where the RBDPO quality can be predicted at the beginning of the process. The plant operator can perform corrective action on the defect quality before the product exits the refining process and the RBDPO quality can be systematically guaranteed.

Multivariate statistical prediction model such as principal component analysis (PCA) is the most common method for modeling the relationships between variables. Due to the multicollinearity in the regression coefficient, PCA requires to combine with other regression method such as Principal Component Regression (PCR) to reduce the variance and covariance between the variables. Conventional PCR model is incapable to

Paper Received: 9 July 2021; Revised: 23 September 2021; Accepted: 16 November 2021
Please cite this article as: Rashid N.A., Shamsuddin A., Khu W.H., Lee M.H., Ibrahim N., Abd Hamid M.K., 2021, Quality Prediction Improvement through Adaptive Nonlinear Principal Component Regression, Chemical Engineering Transactions, 89, 151-156 DOI:10.3303/CET2189026

trace the nonlinear relation and time-varying behaviour of the refinery process, given it is a linear regression model and the predictor coefficient is developed with the assumption that the process mean and variance are constant over time-series (Rashid et al., 2017). The accuracy of static prediction model is reduced when different time-series data is fed to the model, due to the fact that the trained model is no longer represent the current process status (Jeng, 2010). The prediction model must be frequently updated over time to achieve more accurate prediction such that the trained model representing the current process status (Jungmittag, 2014).
This paper would like to improve the quality prediction through adaptive nonlinear PCR prediction model using time-series expansion method for designing better prediction tool's framework. The nonlinear PCR is developed through the combination of nonlinear PCA with the ordinary least square regression, where in this paper, nonlinear iterative partial least square algorithm (NIPALS) is used as nonlinear PCA. The nonlinear PCR is then integrated with the time series expansion methods namely recursive window (RW) analysis and exponentially weighted recursive window (EWRW) analysis to develop the adaptive prediction model.

2. Methodology

The main idea of this paper is to develop the enhanced framework through the integration of time-series expansion method with the nonlinear PCR model as illustrated as in Figure 1. The data is collected from a palm oil refinery located in Sabah, Malaysia for three months. The refining plant is producing the RBDPO according to the three major quality specifications which are Vietnam, China and Palm Oil Refiners Association Malaysia (PORAM).

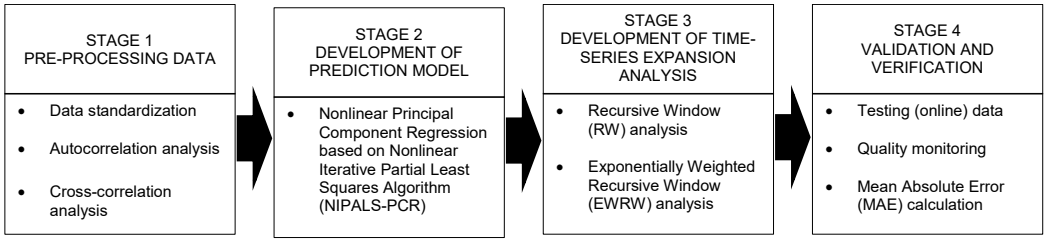

Figure 1: Flowchart of prediction model's framework

The requirement of statistical analysis is that, the data must be random and the processing delay must be augmented to ensure the data is competent during the development of prediction model. The data is pre-processed using autocorrelation analysis to ensure the data is random and exhibit zero correlation between previous and current data (Aue et al., 2014). The autocorrelation analysis is conducted by calculating the correlation of the same time series twice; once in its original form and once is a lagged version of itself over successive time interval. The cross-correlation analysis is performed to identify the processing delay, where the output data is staggered by the processing lag while the input variables remain intact, to ensure the input variables are aligned and matched to the output variables (Rosely et al., 2016). The cross-correlation analysis is conducted by identifying the relationship of the data series by calculating a set of correlation values at increasing time delay.
The predictor coefficient is generated using NIPALS-PCR model through the combination of nonlinear iterative partial least squares algorithm (NIPALS) and Multiple Linear Regression. NIPALS algorithm is one of the nonlinear Principal Component Analysis (PCA) which decomposed the input variables data into principal component (PC) through the iteration of score and loading. The dimension reduction has been utilized in this study by retaining only relevant PC with optimum amount of cumulative variance percentages which is 100 % and 95 % variation, to prevent the overfitting and underfitting of the prediction model's performance (Kracík and Strnadel, 2018). The retained PC from NIPALS algorithm is used as input to the multiple linear regressions to generate the predictor coefficients.
The adaptive prediction model is developed using time-series expansion method, where the new sample observation is included in the model, and the predictor coefficient is continuously updated over time-series (Zou et al., 2018). For Recursive Window (RW) analysis, the number of observation per window is expanding and increases over time-series since the new observation is included without removing the old observation (Pan and Lee, 2003). For example, in the first window, the first predictor coefficient, k_1 is generated from 1^{st} to n^{th} observation number to predict the next observation, $n+1^{th}$. Then, the data is updated in second window by including the $n+1^{th}$ observation without excluding the 1^{st} observation. The predictor coefficient generated from the second window is then used to predict the $n+2^{th}$ observation. The process is repeated until all out-of-sample observations are predicted. The prediction equation can be expressed as in Eq(1).

$$\hat{y}_i = k_{(i-1),1}x_{i,1} + k_{(i-1),2}x_{i,2} + k_{(i-1),3}x_{i,3} + \cdots + k_{(n-1),p}x_{n,p} \quad (1)$$

where $\hat{y}$ is the predicted output variable; x is the input variables; k is the predictor coefficient between input and output variables; i is the observation number ($i=1,2,3\ldots,n$); j is the input variables ($j=1,2,3,\ldots,p$). As expressed in Eq(1), the previous predictor coefficient is used to predict the next or future observation. The Recursive Window analysis can be improved via Exponential Weight Recursive Window (EWRW) analysis. Exploiting the advantage of Exponential Weight Moving Average (EWMA) statistic, the EWRW analysis is developed using the predictor coefficients from recursive window analysis, by assigning the weights to the coefficients such that the present predictor coefficient gets a larger weight, while previous predictor coefficient gets smaller weight. The EWRW predictor coefficient sign statistic can be expressed as in Eq(2).

$$k_{EWRW,i} = \lambda k_{RW,i} + (1-\lambda)k_{EWRW,(i-1)} \quad (2)$$

where λ is the smoothing constant between 0 and 1. Larger λ indicates little memory on the past predictor coefficient, giving more weight to the recent predictor coefficient, while smaller λ indicates the statistic gives high importance to past predictor coefficient.

Monitoring chart is a graph used to monitor the process or quality changes over time (Hrehova, 2016) and constructed based on the three RBDPO quality specifications which are China, Vietnam and Palm Oil Refiners Association Malaysia (PORAM). The actual and predicted output values are plotted on the chart to visually monitor the possible quality deterioration that is the point located beyond the specification limit. The reliability of the prediction models can be analyzed from the consistency of the prediction error from training to testing. The prediction error is calculated using Mean Absolute Error (MAE) as expressed in Eq(3).

$$MAE = \frac{\sum|y - \hat{y}|}{n} \quad (3)$$

where $\hat{y}$ is the predicted output variable; y is the actual output variables; n is the total number of sample. The improvement of adaptive prediction models from the static model is measured using Unscaled Mean Bounded Relative Absolute Error (UMBRAE) as expressed in Eq(4) (Chen et al., 2017). The percentage improvement is calculated as (1-UMBRAE)*100 %, where the positive percentage value indicates the adaptive model performs better than the static model and vice versa.

$$\text{UMBRAE} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\left(|\hat{y}_i - y_i|_{\text{adaptive}}\right)}{\left(|\hat{y}_i - y_i|_{\text{adaptive}}\right)+\left(|\hat{y}_i - y_i|_{\text{static}}\right)}\right]}{1-\left[\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\left(|\hat{y}_i - y_i|_{\text{adaptive}}\right)}{\left(|\hat{y}_i - y_i|_{\text{adaptive}}\right)+\left(|\hat{y}_i - y_i|_{\text{static}}\right)}\right]\right]} \quad (4)$$

3. Results and discussion

The data is statistically analyzed to ensure data is random and the input-output data is matched to each other. Figure 2a is a plot of autocorrelation value against the time lag, where the point falls inside the confidence band (blue horizontal line) is said to be statistically insignificant or random. In time-series, the data is sequentially taken in time order and cannot be solely used, since the current output quality is not the actual output value for the current input. Figure 2b is a plot of cross-correlation value against the time lag, where the three lags (i.e. first lag touches the zero correlation value) are determined as the processing lag or processing delay. The input-output data is matched to each other, when the output data is shifted by the three lags.

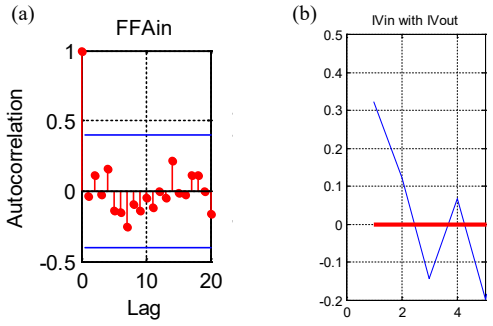

Figure 2: (a) Autocorrelation plot, (b) Cross-correlation plot

The performance of the prediction models is evaluated in terms of deviation error between the actual and predicted output quality using mean absolute error (MAE) as shown in Figure 3. A smaller MAE value, approaching to zero value indicates the predicted output is less deviated from the actual output value. The comparisons of MAE are made for the NIPALS-PCR static model, NIPALS-PCR RW model and NIPALS-PCR EWRW model. The performance of the prediction models was also compared in term of percentage of retained variance, by using 100 % and 95 % variation. The qualities of interest for RBDPO are Free Fatty Acid (FFA_{out}) content, Moisture content ($MOIST_{out}$), Iodine Value (IV_{out}) and Colour ($COLOR_{out}$).

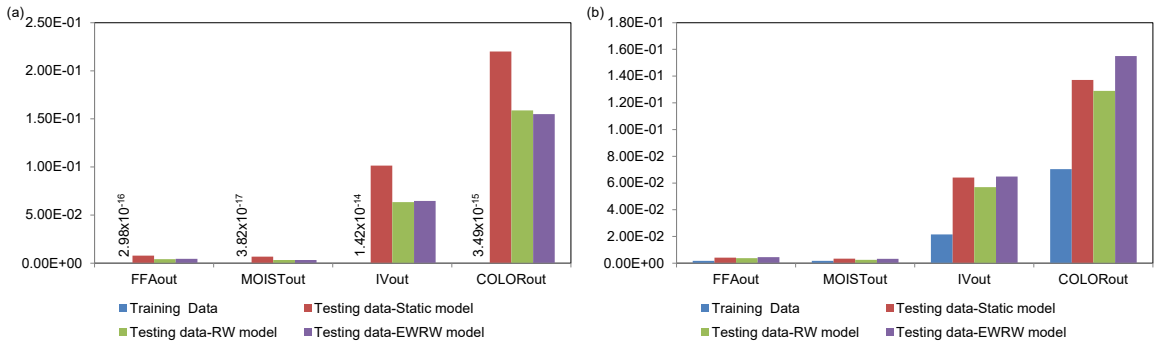

Figure 3: Mean Absolute Error of (a) NIPALS-PCR using 100 % PC Variation, (b) NIPALS-PCR using 95 % PC Variation

The performance consistency is described as the consistency of MAE range value from the training data to testing data. If the prediction model produced a consistent error range, the model is said to be well developed and does not suffer from overfitting or underfitting, and any type of data can fit into the prediction model. Figure 3a shows that the NIPALS-PCR with 100 % PC variation produced inconsistent error from training to testing data. The smallest MAE value (approximately to zero error value) during the training indicates that the predicted output value perfectly fit and follows the trend of the actual output value, but performed poorly when testing data is used as depicted by high MAE value for all prediction models. The NIPALS-PCR model using 100 % PC variation retained all the variability in the training data. The model learns the noise in the training data too well but does not generalized. The model is said to be 'overfitted' and negatively affects the prediction performance on new data (Kracík and Strnadel, 2018). The NIPALS-PCR with 95 % PC variation shows a consistent error range from training to testing data (Figure 3b), and is said to be well developed and can fit to any type of data. The NIPALS-PCR model using 95 % PC variation retained only 95 % of the variability in the training data. Based on the prediction error performance, the prediction model with reduced percentage variation is sufficient to produce a reliable prediction performance.

In comparison of prediction analysis performance, the adaptive prediction models which are NIPALS-PCR RW model and NIPALS-PCR EWRW model, is outperformed the NIPALS-PCR static model. Although the prediction performance for NIPALS-PCR with 100 % PC variation (Figure 3a) shows a significance improvement from static model to adaptive model, still there is a large gap of error range from the training, which indicates that the model's overfitting cannot be reduced using adaptive prediction model. For NIPALS-PCR with 95 % PC variation (Figure 3b), the performance of adaptive prediction model specifically the NIPALS-PCR RW model, improved slightly from the NIPALS-PCR static model, but still producing the error within the error range of training data. Table 1 shows the percentage improvement of adaptive prediction model from the static model for NIPALS-PCR using 95 % PC variations. The adaptive prediction models show significance improvement from static model for all RBDPO quality. Both NIPALS-PCR RW and NIPALS-PCR EWRW models give similar average percentage improvement that is 95 % which indicates that the two adaptive models have equivalence prediction performances.

Table 1: Percentage improvement of adaptive prediction performance using 95 % PC variation

Percentage Improvement	FFA_{out}	$MOIST_{out}$	IV_{out}	$COLOR_{out}$	Average Percentage Improvement
RW model	93.91 %	91.77 %	99.89 %	94.52 %	95.02 %
EWRW model	93.74 %	92.02 %	99.89 %	94.59 %	95.06 %

Monitoring chart is constructed for the actual and predicted value of the RBDPO quality, to observe the accuracy of the prediction, by comparing the trend and pattern of the actual and predicted value. The chart is also used to determine the specification of the predicted output quality. The quality is categorized in three specifications which are Vietnam, China and PORAM. These three specifications have similar quality value for MOIST(< 0.10 %), IV (50 - 55 Wijs) and COLOR (3 red colour), but for FFA quality, there are two quality value which are < 0.10 % (PORAM and Vietnam) and < 0.07 % (China). Figure 4 compares the actual and predicted output value of FFA quality between the NIPALS-PCR using 100 % PC variation 95 % PC variation.

As can be seen from Figure 4, the deviation of predicted value from actual value is very large for 100 % PC variation (Figure 4a) compared to 95 % PC variation (Figure 4b). The predicted value using 100 % variations shows a random trend where the predicted output from three prediction models exceeding the China specifications, and to be specific, the predicted output of NIPALS-PCR static model also exceeding the PORAM and Vietnam specification. By referring to this prediction performance, the refining plant can produce RBDPO quality with PORAM and Vietnam specification using NIPALS-PCR RW and NIPALS-PCR EWRW prediction models. Several monitoring and process adjustment actions need to be planned by the plant manager if the refining plant wants to produce the RBDPO quality with China specification. The findings show that the integration of time-series expansion method can improve the accuracy and reliability of the prediction performance.

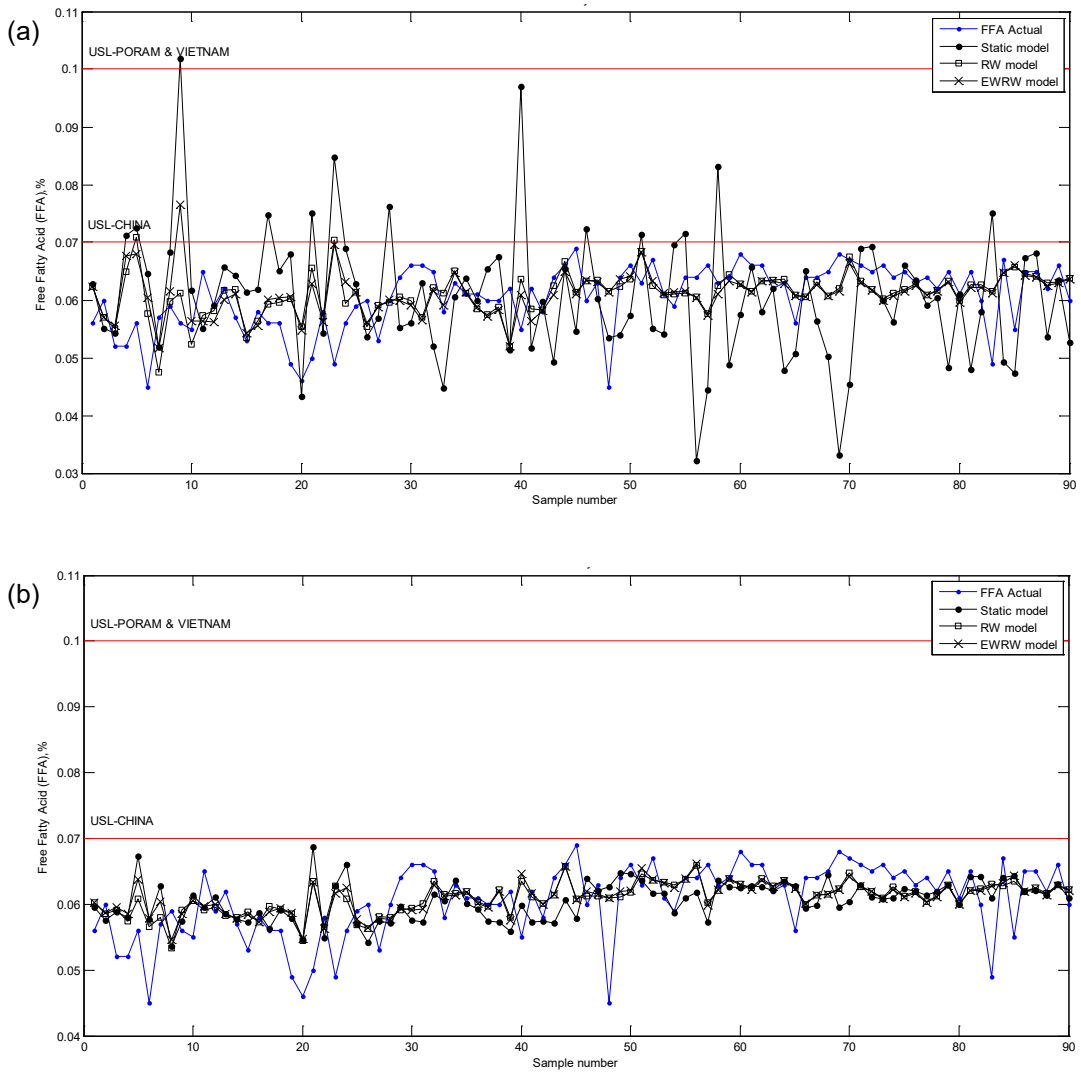

Figure 4: Monitoring chart of actual and predicted free fatty acid quality of (a)NIPALS-PCR using 100 % PC variation, (b)NIPALS-PCR using 95 % PC variation

The predicted value using 95 % variations shows a predictable trend where the predicted output from three prediction models are not exceeding the PORAM, Vietnam and China specifications. The NIPALS-PCR RW and NIPALS-PCR EWRW models show a better prediction trend which is close to the actual output value than the NIPALS-PCR static model. This excellent performance shown by the NIPALS-PCR using 95 % variation is important to help the refining plant in predicting the incoming product quality with no false alarm. By referring to this prediction performance, there is no critical action required during the production since the RBDPO quality is successfully predicted within all quality specification. The findings show that the prediction models can be improved through dimension reduction by retaining optimum percentage of variance. The monitoring chart acts as guidelines to the plant operators at the beginning of the refinery process. Action plans such as identifying the time where off-specification input is expected to come into the process and actions for adjustment on the off-specification products should be prepared, specifically during the production of high-quality specification end product such as China specification.

4. Conclusions

The prediction using static model shows random prediction trend, and is incapable to predict the actual quality specification, which can bring in false alarm to the process. Through the integration of MSPC prediction model with time-series expansion method, the process variation is reduced and the predicted output follows the trends of the actual output quality. The prediction performance is improved by 95 % which depicts that the adaptive prediction model is more capable to predict the actual quality specification of the RBDPO quality. The improvement of quality prediction can bring many benefits to the palm oil refinery plants in the form of monetary and energy savings through the prevention of the product recycle. This allows the refining plant to operate at optimum level and minimizing the process downtime. As data sample increases, the proposed models lead to a slower speed of model adaptation. This limitation can be overcome through the hybrid of two time-series expansion method such as recursive window analysis with moving window analysis.

Acknowledgments

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

References

Aue A., Norinho D.D., Hörmann S., 2014, On the prediction of stationary functional time series, Journal of The American Statistical Association, 110(509), 378-392.

Chen C., Twycross J., Garibaldi J.M., 2017, A new accuracy measure based on bounded relative error for time series forecasting, PLoS ONE, 12(3), 1–23.

Hrehova S., 2016, Predictive model to evaluation quality of the manufacturing process using MATLAB tools, Procedia Engineering, 149, 149-154.

Jeng J.C., 2010, Adaptive process monitoring using efficient recursive PCA and moving window PCA algorithms, Journal of the Taiwan Institute of Chemical Engineers, 41, 475-481.

Jungmittag A., 2014, Combination of forecasts across estimation windows: An application to air travel demand, Journal of Forecasting, 35(4), 373-380.

Kracík J., Strnadel B., 2018, A statistical model for lifespan prediction of large steel structures, Engineering Structures, 176, 20-27.

Pan Y., Lee J.H., 2003, Recursive data-based prediction and control of product quality for a PMMA batch process, Chemical Engineering Science, 58, 3215-3221.

Rashid N., Rosely N.M., Noor M.M., Shamsuddin A., Hamid M.M., Ibrahim K., 2017, Forecasting of refined palm oil quality using principal component regression, Energy Procedia, 142, 2977-2982.

Rosely N., Rashid N., Noor M., Hawi N., Sepuan S., Shamsuddin A., Abd. Hamid, M.K., 2017, Product sampling time and process residence time prediction of palm oil refining process, Chemical Engineering Transactions, 56, 1411-1416.

Xu Y., Mi C., Zhu Q.-X., Gao J.-Y., He Y.-L., 2017, An effective high-quality prediction intervals construction method based on parallel bootstrapped RVM for complex chemical processes, Chemometrics and Intelligent Laboratory Systems, 171, 161-169.

Zhang L., Li Y., Wang Q., Yan C., 2015, Prediction model for steel/slag interfacial instability in continuous casting process, Ironmaking & Steelmaking, 42(9), 705-713.

Zou M., Zhao L., Wang S., Chang Y., Wang F., 2018, Quality analysis and prediction for start-up process of injection molding processes, 10th IFAC Symposium on Advanced Control of Chemical Process ADCHEM 2018, 25th-27th July, Shenyang China, 51(18), 233-238.