



# Statistical Quality Control of NPK Fertilizer Production Process using Mixed Dual Multivariate Cumulative Sum (MDMCUSUM) Chart based on Multioutput Least Square Support Vector Regression (MLS-SVR)

Margaretha Gracia Hotmatua Silalahi<sup>1</sup>, Muhammad Ahsan<sup>1</sup>[0000-0003-3444-2766] and Muhammad Hisyam Lee<sup>2</sup>[0000-0002-3700-2363]

<sup>1</sup> Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

<sup>2</sup> Department of Mathematical Sciences, Universiti Teknologi Malaysia, Johor Baru, Malaysia  
margarethahgs@gmail.com

**Abstract.** Each year, the price of non-subsidized fertilizer has increased. The price increases may reduce demand. However, quality also has an impact on demand. To maintain demand, the quality of non-subsidized fertilizer must be monitored. Using the Mixed Dual Multivariate Cumulative Sum (MDMCUSUM) chart, this study will monitor the quality of non-subsidized NPK (Nitrogen, Phosphorus, and Potassium) fertilizer. The MDMCUSUM chart integrates two different types of Multivariate Cumulative Sum (MCUSUM) charts into a single chart to detect a specific shift in the process mean. Two types of MDMCUSUM, the CP (CMCUSUM and PRMCUSUM) chart and the PC (PRMCUSUM and CMCUSUM) chart, will be used, and the performance will be compared with the assumption that the mean shift is  $[0.75, 1.5]$ . However, autocorrelation in the data has led to an increase in false alarms. To overcome the issue, the Multioutput Least Squares Support Vector Regression (MLS-SVR) model is used. The MLS-SVR algorithm, which utilizes RBF (Radial Basis Function) kernel functions and grid search methods to find the optimal hyperparameters, will be used to generate the residuals. Those residuals are then used to construct the MDMCUSUM chart. Using the optimal hyper-parameters, the MLS-SVR model has successfully reduced autocorrelation in the residuals. When the MDMCUSUM chart is used to monitor the residuals in phase I, out-of-control observations are detected. The optimal hyperparameters obtained from phase I are used in phase II. Further, in phase II, the process is already in-control. Based on monitoring results, the PC chart can detect out-of-control signals faster than the CP chart.

**Keywords:** Control Charts, MDMCUSUM, MLS-SVR, NPK Fertilizer.

## 1 Introduction

Fertilizer is a natural or synthetic material that is applied to plants to help them thrive [1]. However, each year, price increases have occurred for non-subsidized fertilizers.

This is due to the 2019 Corona Virus Disease (COVID-19) and rising commodity prices in the global market. Another factor that contributes to the price increases is the rising cost of distribution [2]. This price increase will decrease the product's demand. This is due to the inverse relationship between demand and price, which states that as price increases, demand decreases, and vice versa [3]. Demand is not only influenced by price but also by quality [4]. Therefore, to maintain demand, companies must be able to maintain or improve the quality of their products. One way to maintain or improve product quality is through a control chart, one of the seven major tools in statistical process control (SPC) [5].

Control chart plots the average measurements of a quality characteristic in samples taken from a process against time [5]. Abdul Haq, Tahir Munir, and Michael B.C. Khoo proposed a control chart that integrates two different types of Multivariate Cumulative Sum (MCUSUM) charts, Crosier MCUSUM (CMCUSUM) and Pignatiello-Runger MCUSUM (PRMCUSUM), into a single chart called Mixed Dual Multivariate Cumulative Sum (MDMCUSUM) to monitor the mean of the multivariate normal process [6]. This control chart has the advantage of detecting the mean shift size that lies within an interval, and merging two MCUSUM charts is used to detect shifts from both small to moderate shift and moderate to large shift. Two types of MDMCUSUM that mix the CMCUSUM chart and PRMCUSUM chart, the CP chart and the PC chart, will be used, and the performance will be compared with the assumption that the mean shift size lies within  $[0.75, 1.5]$ .

In this study, MDMCUSUM charts are used to monitor the non-subsidized NPK fertilizer. However, autocorrelation in the data has led to an increase in false alarms. Autocorrelation is a situation where the errors are correlated across time [7]. It usually happens in time series data because the same variables are measured repeatedly over time. Hence, the MDMCUSUM charts proposed by Abdul Haq, Tahir Munir, and Michael B.C. Khoo cannot be used directly. To overcome the issue, the Multioutput Least Squares Support Vector Regression (MLS-SVR) model is used to reduce or eliminate the autocorrelation in the data. In the previous study, MLS-SVR also brought some benefits when used in combination with multivariate control chart for detecting shifts in the mean, such as MEWMA. The MLS-SVR based MEWMA control chart captures actual additive outlier [8] and becomes an early warning [9]. Therefore, the MLS-modeling and MDMCUSUM charts are used together in this study to monitor the quality of the non-subsidized NPK fertilizer. The performance of the MDMCUSUM charts, the CP chart and the PC chart, will also be compared in this study. The remaining parts of this paper are composed as follows: Section 2 presents the MLS-SVR algorithm. Section 3 shows the MDMCUSUM procedures. Section 4 presents the result and discussion. Finally, Section 5 displays a summary of this study.

## **2 Multioutput Least Square Support Vector Regression (MLS-SVR)**

In this section, the MLS-SVR algorithms are presented. Multioutput regression seeks to discover mapping between multivariate input space and multivariate output space.

Let  $\mathbf{Y} = [y_{i,j}] \in R^{n \times m}$ , where  $i = 1, 2, \dots, n$  is the sample size and  $j = 1, 2, \dots, m$  is the number of output. MLS-SVR algorithm uses a kernel function to perform a non-linear mapping to a higher dimension where  $h$  is a higher dimension [10].

MLS-SVR solves the mean vector  $\mathbf{w}_0 \in R^h$ ,  $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m) \in R^{h \times m}$ , and  $\mathbf{b} = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_m) \in R^m$  simultaneously, one can minimize the following constraint:

$$\text{minimize } J(\mathbf{w}_0, \mathbf{V}, \mathbf{\Xi}) = \frac{1}{2}(\mathbf{w}_0^T \mathbf{w}_0) + \frac{\gamma''}{2m} \text{trace}(\mathbf{V}^T \mathbf{V}) + \frac{\gamma'}{2} \text{trace}(\mathbf{\Xi}^T \mathbf{\Xi}) \quad (1)$$

$$\text{With constraint } \mathbf{Y} = \mathbf{Z}^T \mathbf{W} + \text{repmat}(\mathbf{b}^T, n, \mathbf{1}) + \mathbf{\Xi} \quad (2)$$

Turn the optimization in the previous equation into a Lagrange function as follows:

$$L(\mathbf{w}_0, \mathbf{V}, \mathbf{b}, \mathbf{\Xi}, \mathbf{A}) = J(\mathbf{w}_0, \mathbf{V}, \mathbf{\Xi}) - \text{trace}(\mathbf{A}^T (\mathbf{Z}^T \mathbf{W} + \text{repmat}(\mathbf{b}^T, n, \mathbf{1}) + \mathbf{\Xi} - \mathbf{Y}) \quad (3)$$

The function of the MLS-SVR decision can be obtained as follows:

$$\begin{aligned} \hat{f}(\mathbf{x}) &= \varphi(\mathbf{x})^T \tilde{\mathbf{W}} + (\tilde{\mathbf{b}})^T = \varphi(\mathbf{x})^T \text{repmat}(\tilde{\mathbf{w}}_0, 1, m) + \varphi(\mathbf{x})^T \tilde{\mathbf{V}} + (\tilde{\mathbf{b}})^T \\ &= \varphi(\mathbf{x})^T \text{repmat}(\sum_{j=1}^m \mathbf{Z} \tilde{\alpha}_j, 1, m) + \frac{m}{\gamma''} \varphi(\mathbf{x})^T \mathbf{Z}(\tilde{\mathbf{A}}) + \tilde{\mathbf{b}}^T \\ &= \text{repmat}(\sum_{j=1}^m \sum_{i=1}^n \tilde{\alpha}_{ij} K(\mathbf{x}, \mathbf{x}_i), 1, m) + \frac{m}{\gamma''} \sum_{i=1}^n \tilde{\alpha}_i K(\mathbf{x}, \mathbf{x}_i) + \tilde{\mathbf{b}}^T. \end{aligned} \quad (4)$$

Using the grid search method with Mean Square Error (MSE) criterion, the optimal hyperparameters  $(\gamma', \gamma'', \sigma)$  are chosen. This study uses the Radial Basis Function (RBF) as its kernel function [11].

### 3 Mixed Dual Multivariate Cumulative Sum (MDMCUSUM)

In this section, the MDMCUSUM chart procedures are presented. As a way of detecting the mean shift size that lies within an interval,  $\delta \in [c, d]$ , this chart has two reference values,  $k_1$  and  $k_2$  [12]. The reference values that can be calculated as follows:

$$k_1 = \frac{(3c+d)}{8} \quad (5)$$

$$k_2 = \frac{(c+3d)}{8} \quad (6)$$

Let  $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m$  is the sample characteristics to be monitored that satisfy multivariate data with autocorrelation, where  $\mathbf{y}_j = (\mathbf{y}_{1j}, \mathbf{y}_{2j}, \dots, \mathbf{y}_{nj})^T$ . Each output is assumed to have a significant PACF up to lag  $p_1, p_2, \dots, p_m$  so that the input variables for the MLS-SVR model can be written as follows:

$$\mathbf{X} = (\mathbf{y}_{1,(i-1)}, \dots, \mathbf{y}_{1,(i-p_1)}, \dots, \mathbf{y}_{j,(i-1)}, \dots, \mathbf{y}_{j,(i-p_j)}, \dots, \mathbf{y}_{m,(i-1)}, \dots, \mathbf{y}_{m,(i-p_m)}, \dots) \quad (7)$$

Then, the residual  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_j, \dots, \mathbf{e}_m$  can be computed with  $\mathbf{e}_j = \mathbf{y}_j - \hat{f}(\mathbf{x})_j$ .

### 3.1 MDMCUSUM CP Chart

The MDMCUSUM CP chart is an integration of the CMCUSUM chart and the PRMCUSUM sub-chart, so that it can be expressed as follows:

$$\left. \begin{aligned} \mathbf{A}_{1,i} &= (\mathbf{A}_{1,i-1} + \mathbf{e}_i - \boldsymbol{\mu}_e) \left(1 - \frac{k_1}{B_{1,i}}\right) & \text{if } B_{1,i} > k_1 \\ \mathbf{A}_{1,i} &= \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0_m \end{pmatrix} & \text{if } B_{1,i} \leq k_1 \end{aligned} \right\} \quad (8)$$

where the  $B_{1,i}$  is defined as follows:

$$B_{1,i} = \sqrt{(\mathbf{A}_{1,i-1} + \mathbf{e}_i - \boldsymbol{\mu}_e)' \boldsymbol{\Sigma}_0^{-1} (\mathbf{A}_{1,i-1} + \mathbf{e}_i - \boldsymbol{\mu}_e)} \quad (9)$$

$$\mathbf{A}_{1,0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0_m \end{pmatrix} \quad (10)$$

The CMCUSUM chart is plotted based on the following statistics:

$$E_{1,i} = \sqrt{\mathbf{A}'_{1,i} \boldsymbol{\Sigma}_0^{-1} \mathbf{A}_{1,i}} \quad (11)$$

And PRMCUSUM sub-chart is plotted based on the following statistics:

$$D_{2,i} = \max \left[ 0, \sqrt{\mathbf{C}'_{2,i} \boldsymbol{\Sigma}_0^{-1} \mathbf{C}_{2,i}} - k_2 p_{2,i} \right] \quad (12)$$

where the  $C_{2,i}$  is defined as follows:

$$\mathbf{C}_{2,i} = \sum_{t=i-p_{2,i}+1}^i (\mathbf{e}_t - \boldsymbol{\mu}_e) \quad \text{with} \quad \begin{cases} p_{2,i} = 1 & \text{if } D_{2,i-1} \leq 0 \\ p_{2,i} = 1 + p_{2,i-1} & \text{if } D_{2,i-1} > 0 \end{cases} \quad (13)$$

The CP chart works with two different plotting statistics,  $E_{1,i}$  and  $D_{2,i}$ . An out-of-control signal initiated by the CP chart whenever  $E_{1,i} > h_1$  or  $D_{2,i} > h_2$  [6].

### 3.2 MDMCUSUM PC Chart

The MDMCUSUM PC chart is an integration of the PRMCUSUM chart and the CMCUSUM sub-chart, so that it can be expressed as follows:

$$\mathbf{C}_{1,i} = \sum_{t=i-p_{1,i}+1}^i (\mathbf{e}_t - \boldsymbol{\mu}_e) \quad \text{with} \quad \begin{cases} p_{1,i} = 1 & \text{if } D_{1,i-1} \leq 0 \\ p_{1,i} = 1 + p_{1,i-1} & \text{if } D_{1,i-1} > 0 \end{cases} \quad (14)$$

The points for the PRMCUSUM chart will be plotted based on the following statistics:

$$D_{1,i} = \max \left[ 0, \sqrt{\mathbf{C}'_{1,i} \boldsymbol{\Sigma}_0^{-1} \mathbf{C}_{1,i}} - k_1 p_{1,i} \right] \quad (15)$$

And the points for the CMCUSUM sub-chart will be plotted based on the following statistics:

$$E_{2,i} = \sqrt{\mathbf{A}'_{2,i} \boldsymbol{\Sigma}_0^{-1} \mathbf{A}_{2,i}} \quad (16)$$

where  $\mathbf{A}_{2,i}$  is defined as follows:

$$\left. \begin{aligned} \mathbf{A}_{2,i} &= (\mathbf{A}_{2,i-1} + \mathbf{e}_i - \boldsymbol{\mu}_e) \left( 1 - k_2 / B_{2,i} \right) && \text{if } B_{2,i} > k_2 \\ \mathbf{A}_{2,i} &= \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0_m \end{pmatrix} && \text{if } B_{2,i} \leq k_2 \end{aligned} \right\} \quad (17)$$

and  $B_{2,i}$  is defined as follows:

$$B_{2,i} = \sqrt{(\mathbf{A}_{2,i-1} + \mathbf{e}_i - \boldsymbol{\mu}_e)' \boldsymbol{\Sigma}_0^{-1} (\mathbf{A}_{2,i-1} + \mathbf{e}_i - \boldsymbol{\mu}_e)} \quad (18)$$

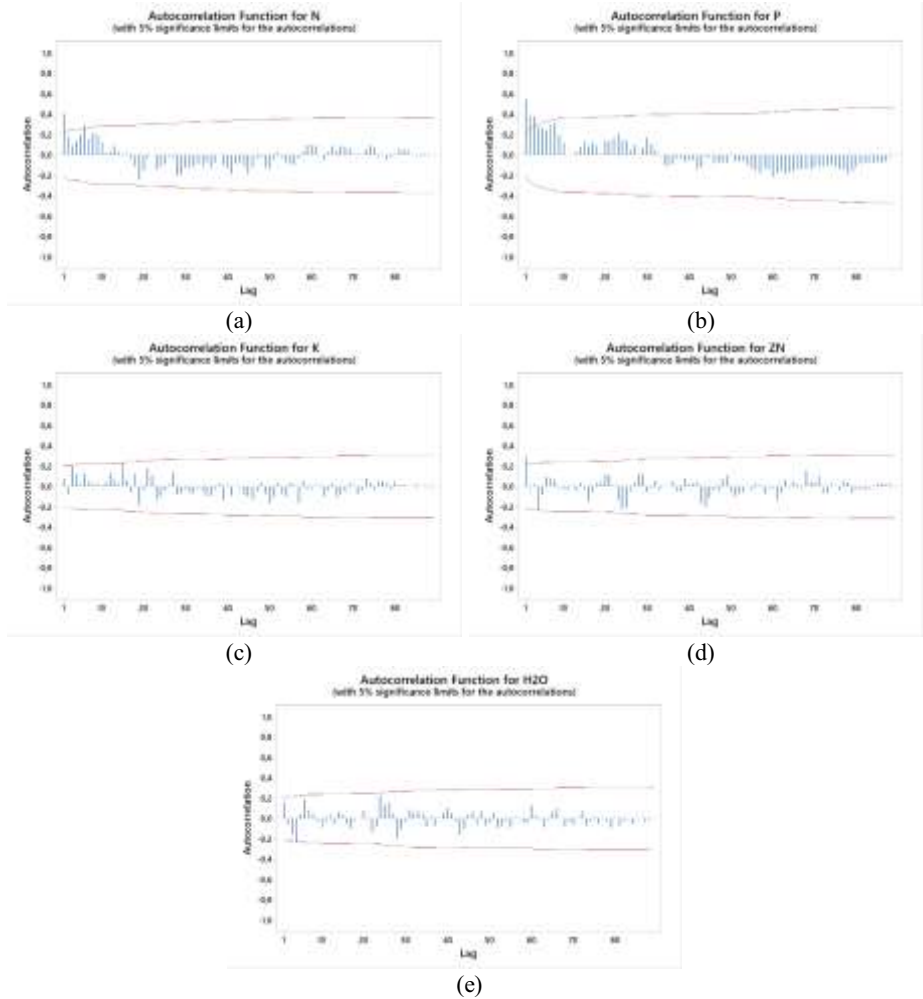
$$\mathbf{A}_{2,0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0_m \end{pmatrix} \quad (19)$$

The PC chart works with two different plotting statistics,  $D_{1,i}$  and  $E_{2,i}$ . An out-of-control signal initiated by the CP chart whenever  $D_{1,i} > h_1$  or  $E_{2,i} > h_2$  [6].

## 4 Result and Discussion

### 4.1 Non-subsidized NPK Fertilizer Data

The data used was obtained from a fertilizer company in Indonesia between September 7 and September 28, 2022. The data used is individual and will be divided into two phases. Phase I is from September 7 to September 21, 2022, and Phase II is from September 22 to September 28, 2022. Five variables are analyzed: Nitrogen (N), Phosphor (P), Potassium (K), Zink (Z), and Water Level (H<sub>2</sub>O). According to Figure 1, all variables analyzed exhibit autocorrelation. To reduce or remove the autocorrelation, the MLS-SVR algorithm is implemented in Phase I.



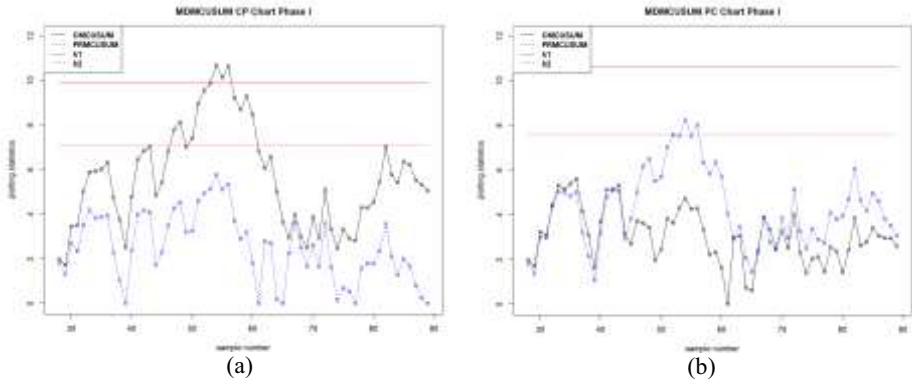
**Fig. 1.** ACF plot of full data for: a) N, b) P, c) K, d) Zn, and e) H<sub>2</sub>O

The optimum hyperparameter of the MLS-SVR model is  $\gamma' = 2^{-20}$ ,  $\gamma'' = 2^{-18}$  and  $\sigma = 2^{-6}$  with an MSE of 0.804. The MLS-SVR algorithm used succeeded in removing or reducing the autocorrelation in the data. From all the variables analyzed, three variables (N, P, and Zn) do not have any autocorrelation and two variables (K and H<sub>2</sub>O) have their autocorrelation reduced. Then, the residual from the MLS-SVR model can be used to plot the MDMCUSUM charts.

## 4.2 Monitoring results for Phase I

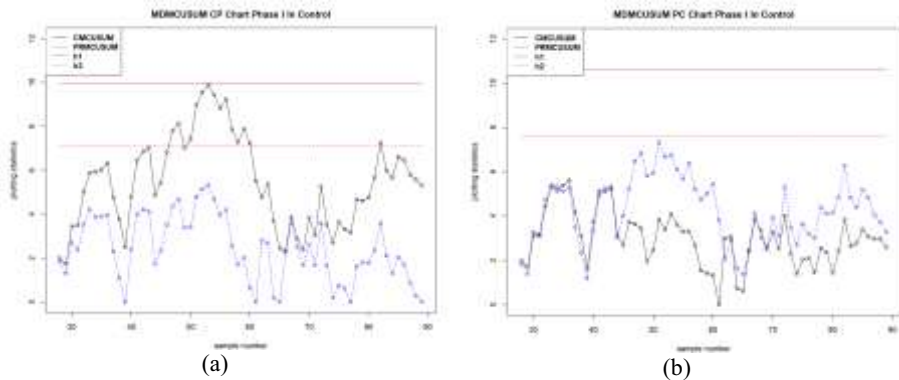
Phase I data was treated as training data to find the optimal hyperparameter. As we assumed that the mean shift is [0.75, 1.5], the two reference values for both the MDMCUSUM charts are  $k_1 = 0.469$  and  $k_2 = 0.656$ . The control limits for the

MDMCUSUM CP chart are  $h_1 = 9.928$  and  $h_2 = 7.093$  and for the MDMCUSUM PC chart are  $h_1 = 10.620$  and  $h_2 = 7.587$ . These values are obtained using simulation based on the paper written by Abdul Haq, Tahir Munir, and Michael B.C. Khoo when the  $ARL_0 = 200$  [6].



**Fig. 2.** Monitoring results for Phase I from: a) MDMCUSUM CP chart and b) MDMCUSUM PC chart

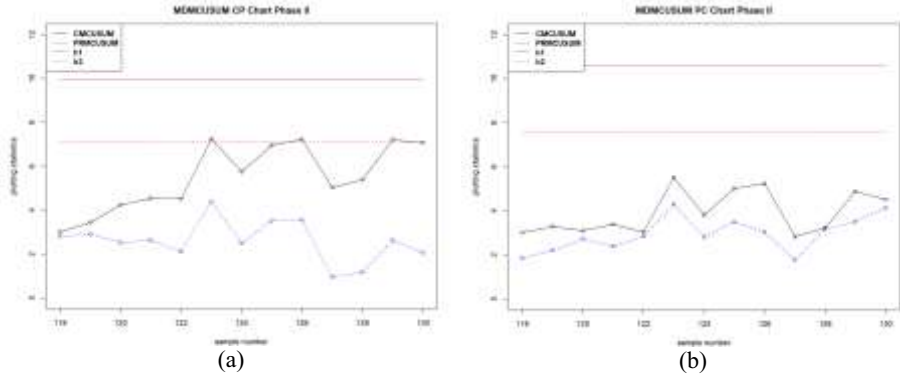
According to Figure 2, both the MDMCUSUM charts issued three out-of-control signals to indicate that there is a shift in the in-control process mean. Therefore, it is necessary to find a way to overcome the existence of out-of-control observations. Since the residual data used in this study is time series data, each time period must have a value, and out-of-control observations cannot be directly deleted as in conventional control charts. Thus, the farthest out-of-control observation from the residual needs to be replaced. The residual value is replaced with a zero value. Because the predicted value is anticipated to be near to the actual value, it is anticipated that the residual value will approach zero. Figure 3 depicts that all observations in the MDMCUSUM charts are statistically controlled, and the monitoring could continue to Phase II.



**Fig. 3.** Monitoring results for Phase I in control from: a) MDMCUSUM CP chart and b) MDMCUSUM PC chart

### 4.3 Monitoring results for Phase II

Phase II data was treated as testing data. The reference values and the control limits used in Phase II are the same as those used in Phase I.



**Fig. 4.** Monitoring results for Phase II from: a) MDMCUSUM CP chart and b) MDMCUSUM chart

According to Figure 4, all observations are statistically controlled.

### 4.4 Comparisons between the MDMCUSUM Charts

In this subsection, the performance between the two types of MDMCUSUM charts are compared.

**Table 1.** Comparison of out-of-control signal in Phase I

Types of MDMCUSUM Chart	Phase	Number of out-of-control signals	Out-of-control observations
CP Chart	I	3	54, 55, 56
	II	0	-
PC Chart	I	3	52, 54, 56
	II	0	-

According to Section 4.2, both MDMCUSUM charts can detect three out-of-control observations. From Table 1, it can be concluded that the MDMCUSUM PC chart detects the first out-of-control signal earlier than the MDMCUSUM CP chart.

## 5 Conclusion and Future Works

In this study, the non-subsidized NPK fertilizer data is monitored using the Mixed Dual Multivariate Cumulative Sum (MDMCUSUM) chart. One advantage of utilizing this control chart is its ability to identify the mean shift size within a specified interval, which is assumed to be  $[0.75, 1.5]$  in this study. Autocorrelation in the data precludes the use of conventional multivariate charts. Then, the MLS-SVR model is utilized to handle the data. Despite its potential utility, the MLS-SVR model requires multiple



iterations to find the optimal hyperparameter. After finding the optimal hyperparameter ( $\gamma' = 2^{-20}$ ,  $\gamma'' = 2^{-18}$ , and  $\sigma = 2^{-6}$ ), the residuals are computed. In Phase I, both the MDMCUSUM charts detect three out-of-control observations. Furthermore, in Phase II, both the MDMCUSUM charts cannot detect any out-of-control observations. Compared to the MDMCUSUM CP, the MDMCUSUM PC detects out-of-control observations faster. The residuals generated using the MLS-SVR model proved that the MLS-SVR model successfully reduced the autocorrelation, even though it took some time. Therefore, other methods of machine learning, like Long Short-Term Memory (LSTM), to generate the residual data can also be considered for future development of this control chart.

## References

- [1] Cambridge University Press, "Cambridge Academic Content Dictionary," Cambridge University Press and Assessment, [Online]. Available: <https://dictionary.cambridge.org/dictionary/english/fertilizer>. [Accessed 6 October 2023].
- [2] Renta, "Demfarm," Demfarm, 28 December 2021. [Online]. Available: <https://www.demfarm.id/check-this-out-prices-of-non-subsidized-fertilizers-in-2021>. [Accessed 6 October 2023].
- [3] S. B. Mohanty, "Revisiting Human Behaviour Through Demand Analysis: Why Demand Curve May Not Be Downward Sloping," *International Journal of Economics, Commerce, and Management*, vol. 2, no. 5, 2014.
- [4] B. H. Brata, S. Husani and H. Ali, "The Influence of Quality Products, Price, Promotion, and Location to Product Purchase Decision on Nitchi At PT. Jaya Swarasa Agung in Central Jakarta," *Saudi Journal of Business and Management Studies*, vol. 2, no. 4B, pp. 433-445, 2017.
- [5] D. C. Montgomery, Introduction to Statistical Quality Control, Sixth Edition, New Jersey: John Wiley & Sons, Inc., 2009.
- [6] A. Haq, T. Munir and M. B. Khoo, "Dual multivariate CUSUM mean charts," *Computers & Industrial Engineering*, vol. 137, pp. 5-8, 2019.
- [7] J. M. Woolridge, Introductory Econometrics: A Modern Approach, Fifth Edition, Douth-Western: Cengage Learning, 2012.
- [8] H. Khusna, M. Mashuri, Suhartono, D. D. Prastyo and M. Ahsan, "Multioutput Least Square SVR Based Multivariate EWMA Control Chart," *Journal of Physics: Conference Series*, 2018.
- [9] H. Khusna, M. Mashuri, Suhartono, D. D. Prastyo and M. Ahsan, "Multioutput least square SVR-based multivariate EWMA control chart: The performance evaluation and application," *Cogent Engineering*, vol. 5, no. 1, 2018.
- [10] S. Xu, X. An, X. Qiao, L. Zhu and L. Li, "Multi-output least-squares support vector regression machines," *Pattern Recognition Letters*, vol. 34, pp. 1078-1084, 2013.
- [11] M. Mashuri, H. Khusna, Wibawati and F. D. Putri, "Mixed Multivariate EWMA-CUSUM (MEC) Chart based on MLS-SVR Model for Monitoring Drinking Water Quality," *Journal of Physics: Conference Series*, vol. 2123, 2021.
- [12] Y. Zhao, F. Tsung and Z. Wang, "Dual CUSUM control schemes for detecting a range of mean shifts," *IIE Transactions*, vol. 37, pp. 1047-1057, 2005.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

