Vector Autoregressive-Based Maximum MCUSUM Control Chart for Monitoring the Quality of White Crystal Sugar

H Khusna 1*, M Mashuri 1, and Wibawati 1
1 Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

*Email: hidayatul@its.ac.id

Abstract. The white crystal sugar which is widely consumed sugar has two critical to qualities, namely the index of solution colour and the level of sulphur dioxide. These quality characteristics have small mean and variability shifts, as well as autocorrelation pattern. This research aims to propose residual-based Maximum Multivariate Cumulative Sum (Max-MCUSUM) control chart, one of the single control charts to monitor small shifts of mean and variability simultaneously, for monitoring the quality of white crystal sugar. The vector autoregressive (VAR) model is utilized to model the daily solution colour index and the daily sulphur dioxide level, then the residuals are monitored using Max-MCUSUM chart. The VAR-based Max-MCUSUM chart employs bootstrap, one of the nonparametric resampling methods, to estimate the control limit. The results of white crystal sugar quality control show that the processes in the last week of August 2020 need to be improved. Monitoring the white crystal sugar data using conventional control chart leads to many false alarm signals. Furthermore, the proposed control chart is more sensitive than the residual-based MEWMA and residual-based Hotelling’s $T^2$ charts in case of monitoring the quality of white crystal sugar.

Keywords: autocorrelated, control chart, Max-MCUSUM, quality, VAR, white crystal sugar

1. Introduction
Indonesia has established sugar as a special commodity (special product) along with rice, corn, and soybeans in the World Trade Organization (WTO) negotiations [1]. This determination gives meaning to the role of sugar commodities which are very influential in people's lives. Sugar becomes the most common sweetener consumed by the Indonesian people. As the main sweetener, the use of sugar still cannot be completely replaced by sweeteners such as honey, corn syrup, or others. The need for sugar consumption as one of the staples Indonesian societies is increasing along with the increasing of population [2]. The level of sugar consumption by households in Indonesia tends to decrease until in 2018 it reached 6,607 kg/capita/year. However, the need for sugar in Indonesia continues to increase every year due to the development of the food and beverage industry [3]. This condition leads in a competitive atmosphere between the sugar industry to satisfy the desires of consumers.

The white crystal sugar is one of the widely consumed sugars in Indonesia. According to the National Standardization Agency (BSN), the parameters that determine the quality of white crystal sugar include solution colour, grain size, drying shrinkage, polarization, conductivity ash, sulphur dioxide, lead, copper, and arsenic. The main parameters determining the quality of white crystal sugar are solution colour and sulphur dioxide [4]. These parameters have negative relationship, which is the higher the index of solution colour, the lower the level of sulphur dioxide. Therefore, the quality of white crystal sugar can be monitored using multivariate control chart, one of the seven tools statistics which is useful for improving quality through reducing variability [5].
Some researchers have pointed out that the utilization of simultaneous control chart, that jointly monitored the mean and variability of a process, is more reasonable than the conventional control chart [6,7]. Since the control limits of conventional mean chart are affected by the shift of variability process, both parameters are more effective to be monitored simultaneously [8]. Some simultaneous control charts proposed for monitoring multivariate data such as multivariate Max chart [9,10], multivariate Max-half chart [11,12], Maximum Multivariate Exponentially Weighted Moving Average (Max-MEWMA) chart [13], and Max-MCUSUM chart [14,15]. If the variability and mean of a process are shifted no more than 1.5 sigma, then the memory type chart, such as Max-MEWMA and Max-MCUSUM, has better performance than Shewhart type chart, such as multivariate Max chart and multivariate Max-half chart.

This research aims to develop a multivariate simultaneous control chart for monitoring the quality of white crystal sugar. Since the quality characteristics of white crystal sugar show autocorrelation pattern and tend to have small process shifts, the residual-based simultaneous-memory type chart is utilized. The vector autoregressive (VAR) is harnessed to analyse the time series pattern in white crystal sugar characteristics, which are the level of solution colour and the level of sulphur dioxide, then the residuals are monitored using Max-MCUSUM control chart. VAR is one of the conventional time series models, based on Box–Jenkins’s algorithm, that is useful for analyse multivariate linear data [16]. The bootstrap-based Max-MCUSUM chart is chosen due to its advantages in predetermining the level of tightness and its flexibility in following certain distribution as stated in Khusna et al. [15]. Hence, the Max-MCUSUM chart based on the residuals of VAR model is proposed to monitor the level of solution colour and the level of sulphur dioxide in white crystal sugar production process.

2. The proposed control chart

The Vector Autoregressive (VAR) model is a development of the Autoregressive (AR) model for multivariate time series data [17]. The advantage of applying residuals from the VAR model on the control chart is its effectiveness in monitoring autocorrelation of multivariate processes [18]. The general form of the VAR model with order \( p \) can be written as follows:

\[
X_i = \mu + \Phi_1 X_{i-1} + \Phi_2 X_{i-2} + \cdots + \Phi_k X_{i-k} + \cdots + \Phi_p X_{i-p} + e_i, \quad (1)
\]

where \( \mu \) is the \( m \times 1 \) vector consisting of the constant of VAR model, \( X_{i-k} \) is the \( m \times 1 \) vector of quality characteristics \( X \) at time \( i-k \) \((k = 1,2, \ldots, p)\), \( \Phi_k \) is \( m \times m \) matrix of VAR parameter at order \( k \), the \( m \times 1 \) vector of residuals at time \( i \) is denoted as \( e_i \), and \( m \) shows the number of quality characteristics [16].

The identification VAR order can be carried out by observing the Matrix Partial Cross Correlation Function (MPCCF) pattern [16]. The parameters of VAR model are estimated using least square method [19]. The VAR model in equation (1) can be written as follows:

\[
X = L \beta + e, \quad (2)
\]

where

\[
X = \begin{bmatrix} X_{p+1}' \\ X_{p+2}' \\ \vdots \\ X_n' \end{bmatrix}_{(n \times m)}, \quad \beta = \begin{bmatrix} \mu' \\ \Phi_1' \\ \vdots \\ \Phi_p' \end{bmatrix}_{((mp+1) \times m)}, \quad e = \begin{bmatrix} e_{p+1}' \\ e_{p+2}' \\ \vdots \\ e_n' \end{bmatrix}_{(n \times m)}, \quad L = \begin{bmatrix} 1 & X_p' & \cdots & X_1' \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n-1}' & \cdots & X_{n-p}' \end{bmatrix}_{(mp+1) \times (mp+1)}.
\]

By minimizing the sum of square error in least square method, we can obtain the equation to estimate the parameters of VAR model as:

\[
\hat{\beta} = [L' L]^{-1} L' X. \quad (3)
\]

The residuals of VAR model for \( j \)-th quality characteristics \((j = 1,2, \ldots, m)\) with \( n \) observations are denoted as \( e_{1j}, e_{2j}, \ldots, e_{ij}, \ldots, e_{nj} \). In the other word, the \( m \times 1 \) vector of residual for \( i \)-th observation can be written as \( e_i = (e_{i1}, e_{i2}, \ldots, e_{ij}, \ldots, e_{im}) \). The residuals \( e_i \) have the mean vector, either denoted as \( \mu_{e(g)} \) (when obtained from the in-control process) or noted as \( \mu_{e(0)} \) (when resulted from the out-of-control process). If \( \Sigma_{e(g)} \) is the known in-control covariance matrix of \( e_i \), then the statistics of Max-
MCUSUM based on the residuals of VAR model can be obtained by forming two random variables as in equation (4) and (5).

\[
Z_i = \frac{(\mu_{(e)} - \mu_{(e_{(t)})})^T \Sigma^{-1}_{(e)} (e_i - \mu_{(e_{(t)})})}{\sqrt{\sum_{i=1}^n (\mu_{(e)} - \mu_{(e_{(t)})})^T \Sigma^{-1}_{(e)} (\mu_{(e)} - \mu_{(e_{(t)})})}}, \tag{4}
\]

\[
W_i = \Phi^{-1} \left( H \left[ (e_i - \mu_{(e_{(t)})})^T \Sigma^{-1}_{(e)} (e_i - \mu_{(e_{(t)})}); m \right] \right), \tag{5}
\]

where \( \Phi(z) = P(Z \leq z) \) for \( Z \sim N(0,1) \). The function \( \Phi^{-1}(.) \) specifies the inverse of standard normal cumulative distribution function and \( H(x; m) = (X \leq x; m) \), for \( X \sim \chi^2(m) \). The shifted covariance matrix is written as \( \Sigma_{e(B)} \).

The residual based MCUSUM statistics for monitoring the mean vector and the covariance matrix (which are written in equation (6) and (7), respectively) are formulated by transforming the random variables in equation (4) and (5), respectively.

\[
C^{+}_{(r)} = \max[0, Z_{i} - k + C^{+}_{(r_{i-1})}], \tag{6}
\]

\[
C^{-}_{(r)} = \max[0, -k - Z_{i} + C^{-}_{(r_{i-1})}], \tag{7}
\]

\[
S^{+}_{(r)} = \max[0, W_{i} - k + S^{+}_{(r_{i-1})}], \tag{7}
\]

\[
S^{-}_{(r)} = \max[0, -k - W_{i} + S^{-}_{(r_{i-1})}], \tag{7}
\]

where \( C_{r0} = 0 \) and \( S_{r0} = 0 \) explain the starting point, and \( k > 0 \) is the reference value that need to be predetermined in Phase I monitoring process. Then, the residual based MCUSUM statistics are transformed as in equation (8) because we are monitoring the significance of the shift magnitude using multivariate quality control technique.

\[
C_{(r)} = \max[C^{+}_{(r)}, C^{-}_{(r)}], \tag{8}
\]

\[
S_{(r)} = \max[S^{+}_{(r)}, S^{-}_{(r)}]. \tag{8}
\]

Hence, the statistics of Max-MCUSUM control chart based on the residuals of VAR model can be written as in equation (9). The process is said to be out of control if the statistic \( M \) is greater than the Upper Control Limit (UCL). The bootstrap method to calculate the UCL is explained in Algorithm 1.

\[
M_{(r)} = \max[C_{(r)}, S_{(r)}]. \tag{9}
\]

**Algorithm 1.** The Bootstrap Method for Calculating the Upper Control Limit of VAR-based Max-MCUSUM Chart

1. Predetermine the reference value and significance level.
2. Analyse the quality characteristics in phase I monitoring process using VAR model as in equation (1), where the parameters are estimated using equation (3).
3. For \( i = 1, 2, \ldots, n \) observations, follow these steps:
   a. Calculate the residuals of VAR model \( e_i = (e_{i1}, e_{i2}, \ldots, e_{im}) \), which is the difference between the original data and the fitted value of VAR model, as stated in equation (2).
   b. Compute the VAR-based Max-MCUSUM statistics \( M_{(r)} \) by following equation (4) until equation (9).
4. For \( \ell = 1, 2, \ldots, N \) replications, follow these steps:
   a. Resample \( M_{(r)} \) for \( B \), certain large number, times.
   b. Calculate the \((100 \times (1 - \alpha))-\)th percentile of bootstrap samples \( M_{(r)}^{100 \times (1 - \alpha)} = M_{(B \times (1 - \alpha))}^{\ell} \), where \( M_{(r)}^{\ell} \) denotes order statistics in \( \ell \)-th replication.
5. Obtain the UCL, which is the average of \((100 \times (1 - \alpha))-\)th percentile over \( N \) replications.
3. Results and discussion
This section performs the application of the proposed control chart to monitor the quality of white crystal sugar. The index of solution colour and the level of sulphur dioxide are measured every hour with a sample of 100 gram of white crystal sugar taken from a conveyor. This research uses the daily quality characteristics obtained from the average of data gathered in every hour. The index of solution colour indicates the purity associated with the colour of sugar in a solution, and it is measured according to the International Commission for Uniform Methods of Sugar Analysis (ICUMSA) standard. The higher the index, then the darker the solution colour, and vice versa. Meanwhile, the level of sulphur dioxide as a food additive can blanch the colour to produce sugar crystals with a bright white colour. The index of solution colour and the level of sulphur dioxide are measured in UI and mg/kg units, respectively [4].

![Figure 1](https://example.com/figure1.png)

**Figure 1.** Time series plot comparison between actual and fitted value of (a) index of solution colour and (b) level of sulphur dioxide

This research utilizes the white crystal sugar data recorded from 4th of June until 31st of August 2020. As displayed in the time series plot in figure 1, there are 89 observations of solution colour index as well as sulphur dioxide level (see the black plot for original data). Based on the MPCCF of the original data displayed in Table 1, we can identify the significance of MPCCF at lag 1, 4, and 6 which will be a possible order of VAR. From Table 2, it can be known that the minimum AIC is found in lag AR 4 with MA 0, then we fit the VAR (4) model. Since the parameters of VAR (4) model at lag 2 and lag 3 are not significant, we need to re-estimate the VAR model with order 1 and 4, which also correspond to the significance of MPCCF lag. The forecast values of VAR([1,4]) model with all significant parameters are displayed in figure 1 (see the red plot). However, the residuals of this VAR model violence the normal distribution assumption.

### Table 1. The MPCCF of white crystal sugar quality characteristics

| Variable/Lag | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|--------------|---|---|---|---|---|---|---|---|---|----|----|----|
| Solution colour index | + | . | . | . | + | . | . | . | . | . | . | . |
| Sulphur dioxide level | .+ | . | . | . | . | . | . | . | . | . | . | . |
| + is > 2*std error, - is < -2*std error, . is between |

### Table 2. Minimum information criterion based on akaike’s information criterion (AIC)

| Lag | MA 0   | MA 1   | MA 2   | MA 3   | MA 4   | MA 5   |
|-----|--------|--------|--------|--------|--------|--------|
| AR 0 | 10.59325 | 10.19118 | 10.19581 | 10.17362 | 10.09831 | 10.10176 |
| AR 1 | 9.91237 | 9.94618 | 9.99270 | 9.98641 | 9.98119 | 10.03171 |
| AR 2 | 9.92583 | 9.95836 | 10.04807 | 10.06876 | 10.06014 | 10.07368 |
| AR 3 | 9.98753 | 9.97175 | 9.91437 | 9.95228 | 10.02059 | 10.03104 |
| AR 4 | **9.90703** | 9.92440 | 9.96237 | 9.98561 | 10.06895 | 10.11889 |
| AR 5 | 10.00889 | 9.96159 | 10.00418 | 9.96159 | 10.00418 | 9.96159 |

Table 2 shows the possible outlier in solution colour index which is recorder at 14th of July 2020 leads to the violation of normal distribution assumption in the residuals of VAR model. Since in a control
chart we need to build normal profile for calculating the control limit, then the solution colour index observation at 14th of July 2020 is replaced with the forecast value (see the black plot in figure 2 for the revised original data). Furthermore, these revised data are analysed using VAR([1,4]) model. The forecast values using the recent model with all significant parameters are exhibited as the red plot in figure 2. From the parameter significance test for VAR model shown in Table 3, we can understand that the parameters are significant after some variables are restricted. The residuals of VAR([1,4]) model already satisfy the normal distribution assumption. Moreover, Table 4 shows the portmanteau test indicating that the residuals of VAR([1,4]) model follow white noise condition.

![Figure 2](image)

**Figure 2.** Revised time series plot comparison between actual and fitted value of (a) index of solution colour and (b) level of sulphur dioxide.

The VAR([1,4]) model the index of solution colour ($X_1$) and the level of sulphur dioxide ($X_2$) can be written as follows:

$$X_t = \mu_1 + \phi_{1,11} X_{t-1} + \phi_{1,12} X_{t-4} + \epsilon_{1,t},$$

$$X_2 = \mu_2 + \phi_{2,11} X_{2,t-1} + \phi_{2,22} X_{2,t-4} + \epsilon_{2,t}.$$  

**Table 3.** The parameter significance test of VAR model

| Equation   | Parameter | Estimate | Standard Error | t Value | P-value | Variable |
|------------|-----------|----------|----------------|---------|---------|----------|
| $X_1$ (Index of solution colour) | $\mu_1$ | 114.4856 | 32.3026 | 3.54 | 0.0007 | 1 |
|            | $\phi_{1,11}$ | 0.3679 | 0.0922 | 3.99 | 0.0001 | $x_{1,t-1}$ |
|            | $\phi_{1,12}$ | 0.0000 | 0.0000 | - | - | $x_{2,t-1}$ |
|            | $\phi_{1,11}$ | 0.3291 | 0.0847 | 3.89 | 0.0002 | $x_{1,t-4}$ |
|            | $\phi_{1,12}$ | -1.9217 | 0.8422 | -2.28 | 0.0252 | $x_{2,t-4}$ |
| $X_2$ (Level of sulphur dioxide) | $\mu_2$ | 7.1709 | 1.2950 | 5.54 | 0.0001 | 1 |
|            | $\phi_{2,11}$ | 0.0000 | 0.0000 | - | - | $x_{1,t-1}$ |
|            | $\phi_{2,22}$ | 0.4458 | 0.0959 | 4.65 | 0.0001 | $x_{2,t-1}$ |
|            | $\phi_{2,21}$ | 0.0000 | 0.0000 | - | - | $x_{1,t-4}$ |
|            | $\phi_{2,22}$ | 0.0000 | 0.0000 | - | - | $x_{2,t-4}$ |

**Table 4.** The portmanteau test for cross correlations of residuals

| Up to Lag | DF | Chi-Square | P-value |
|-----------|----|------------|---------|
| 5         | 4  | 8.23       | 0.0835  |
| 6         | 8  | 13.29      | 0.1023  |
| 7         | 12 | 17.34      | 0.1372  |
| 8         | 16 | 21.97      | 0.1443  |
| 9         | 20 | 26.76      | 0.1421  |
| 10        | 24 | 34.71      | 0.0729  |
| 11        | 28 | 35.53      | 0.1551  |
| 12        | 32 | 37.54      | 0.2301  |
The residuals of VAR([1,4]) model are monitored using Max-MCUSUM chart using reference value $k = 0.5$. The UCL of the proposed chart for significance level $\alpha = 0.0027$ is calculated using Algorithm 1 which is equal to 4.248 (see figure 3.d). As a comparison, the authors also monitor the residuals of VAR([1,4]) model using Hotelling’s $T^2$ (Shewhart type chart) and MEMMA (memory type chart). The MEWMA chart for the original data exhibited in figure 3.a indicating that monitoring autocorrelated data using conventional control chart results in many false alarm signals. If the data have autocorrelation pattern, then the residual-based control charts such as displayed in figure 3.b until 3.d become the proper choice. The residual-based control charts (see figure 3.b until 3.d) are starting from 8th of June 2020 because we have VAR model with maximum lag at order 4. Both MEWMA and residual based MEMMA control charts employ the smoothing parameter $\lambda = 0.5$ and in-control Average Run Length ($ARL_0$) equal to 370. The VAR-based MEWMA chart (see figure 3.b) is more sensitive than the VAR-based $T^2$ chart (see figure 3.c). From the VAR-based Max-MCUSUM chart shown in Figure 3.d, we can understand that the quality of white crystal sugar starting from 26th of August 2020 need to be improved.

![Figure 3](image-url)

**Figure 3.** Monitoring result of white crystal sugar quality using (a) MEWMA chart, (b) VAR-based MEWMA chart, (c) VAR-based $T^2$ chart, and (d) VAR-based Max-MCUSUM chart

### 4. Conclusion and future research

This research proposes the Max-MCUSUM control chart based on the residuals of VAR model, one of the simultaneous charts for joint monitoring the mean and variability of multivariate autocorrelated data. The bootstrap resampling method is utilized to estimate the control limit of the proposed chart. The proposed chart is applied to monitor the quality of white crystal sugar. The residuals of VAR([1,4]) model for monitoring the quality of white crystal sugar satisfy white noise and normal distribution. Some false alarm signals are found when monitoring the autocorrelated white crystal sugar data using conventional MEWMA chart. The proposed VAR-based Max-MCUSUM chart shows that the quality of white crystal sugar in the last week of August 2020 need to be improved. This study can be extended by evaluating the performance of the proposed control chart using ARL.
References

[1] Arifin B 2008 Ekonomi swasembada gula indonesia Economic Review 211 1–12
[2] Sugiyanto C 2007 Permintaan gula di Indonesia Jurnal Ekonomi Pembangunan: Kajian Masalah Ekonomi dan Pembangunan 8 113–27
[3] Trikuntari D, Permadhi D and Putra L K 2020 Analisis Kinerja dan Prospek Komoditas Gula Opini Dan Analisis Perkebunan 1 1–10
[4] Cahyanı O E 2021 Pengendalian Kualitas GKP di PG Kebon Agung menggunakan Peta Kendali MEWMV dan MEWMA Berbasis Model Time Series Tugas Akhir, Departemen Statistika, Institut Teknologi Sepuluh Nopember
[5] Montgomery D C 2009 Introduction to statistical quality control (John Wiley & Sons (New York))
[6] Mccracken A K and Chakraborti S 2013 Control charts for joint monitoring of mean and variance: An overview Quality Technology and Quantitative Management 10 17–36
[7] Thaga K and Sivasamy R 2015 Single Variables Control Charts: A Further Overview Indian Journal of Science and Technology 8 518
[8] Khusna H, Mashuri M, Suhartono, Prastyo D D, Lee M H and Ahsan M 2019 Residual-based maximum MCUSUM control chart for joint monitoring the mean and variability of multivariate autocorrelated processes Production & Manufacturing Research 7 364–94
[9] Khoo M B C 2004 A new bivariate control chart to monitor the multivariate process mean and variance simultaneously Quality Engineering 17 109–18
[10] Thaga K and Gabaitiri L 2006 Multivariate Max-Chart Economic Quality Control 21
[11] Kruba R, Mashuri M and Prastyo D D 2021 Max-Half-Mchart: A Simultaneous Control Chart Using a Half-Normal Distribution for Subgroup Observations IEEE Access 9 105369–81
[12] Kruba R, Mashuri M and Prastyo D D 2021 The effectiveness of Max-half-Mchart over Max-Mchart in simultaneously monitoring process mean and variability of individual observations Quality and Reliability Engineering International
[13] Xie H 1999 Contributions to qualimetry
[14] Cheng S W and Thaga K 2005 Multivariate Max-CUSUM Chart Quality Technology & Quantitative Management 2 221–35
[15] Khusna H, Mashuri M, Ahsan M, Suhartono S and Prastyo D D 2018 Bootstrap-based maximum multivariate CUSUM control chart Quality Technology & Quantitative Management 1–23
[16] Wei W W S 2006 Time series analysis The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2
[17] Tsay R S 2013 Multivariate time series analysis: with R and financial applications (John Wiley & Sons)
[18] Pan X and Jarrett J 2007 Using vector autoregressive residuals to monitor multivariate processes in the presence of serial correlation International Journal of Production Economics 106 204–16
[19] Liao J-C and Tsay W-J 2016 Multivariate least squares forecasting averaging by vector autoregressive models Available at SSRN 2827416