

# Multivariate CUSUM Control Chart Based on the Residuals of Multioutput Least Squares SVR for Monitoring Water Quality

Hidayatul Khusna<sup>1a</sup>, Muhammad Mashuri<sup>1b</sup>, Suhartono<sup>1c</sup>, Dedy Dwi Prastyo<sup>1d\*</sup>, Muhammad Ahsan<sup>1e</sup>

<sup>1</sup>Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, INDONESIA.  
E-mail: khusna16@mhs.statistika.its.ac.id<sup>a</sup>; m\_mashuri@statistika.its.ac.id<sup>b</sup>; suhartono@statistika.its.ac.id<sup>c</sup>;  
dedy-dp@statistika.its.ac.id<sup>d</sup>; ahsan16@mhs.statistika.its.ac.id<sup>e</sup>

\*Corresponding Author: dedy-dp@statistika.its.ac.id<sup>d</sup>

Received: 21<sup>st</sup> April 2019

Revised: 6<sup>th</sup> August 2019

Published: 30<sup>th</sup> September 2019

DOI : <https://doi.org/10.22452/mjs.sp2019no2.7>

**ABSTRACT** Monitoring serially dependent processes using conventional control charts yields a high false alarm rate. Multioutput Least Squares Support Vector Regression (MLS-SVR) has the capability to encompass the cross-relatedness between output variables by learning multivariate output variables simultaneously. This research aims to develop a Multivariate Cumulative Sum (MCUSUM) control chart based on the residual obtained from the MLS-SVR model for monitoring autocorrelated data. The inputs of the MLS-SVR are selected using the significant lag of a partial autocorrelation function. The proposed control chart is applied to monitor water quality data and it can detect the assignable causes in those data caused by a broken pipeline.

**Keywords:** autocorrelated, control chart, multioutput least squares SVR, multivariate CUSUM, water quality.

---

## 1. INTRODUCTION

One of the most useable tools in statistical process control is the control chart (Woodall & Montgomery, 1999). Many assumptions need to be fulfilled in order to propose a control chart. Most of the traditional charts assume that the observations are not dependent and satisfy a multivariate normal distribution. The violation of the independence assumption affecting the control chart performance has been investigated by many researchers. Harris & Ross (1991) proved that autocorrelation affects the Average Run Length (ARL) of an Exponentially Weighted Moving Average (EWMA) and Cumulative Sum (CUSUM) control chart. Moreover, a CUSUM control chart drew an incorrect conclusion when applied to autoregressive AR(1) and moving average MA(1) data (Johnson & Bagshaw, 1974). Serially dependent data might lead to

an incorrect out-of-control signal and break the effectiveness of a control chart (Noorossana & Vaghefi, 2006). Applying the conventional control chart for monitoring the autocorrelated processes will produce a high false alarm rate and reduce its ability to detect a shift of the process (Psarakis & Papaleonida, 2007).

There are two different procedures that deal with monitoring autocorrelated data. First is monitoring serially dependent data using a modified control limit of the conventional control chart (Vanbrackle & Reynolds, 1997). Another approach is developing a residual-based control chart. This second procedure applies a time series model to the serially dependent data then uses a residuals component, i.e., the difference between the actual value and the forecast value, to monitor the process. The obtained residuals are independent, so it is possible to

monitor the process using the residuals component. Yashchin (1993) suggested applying the modified control chart for a low level of autocorrelation, but the residual-based control chart for a high level of autocorrelation.

Many researchers have proposed different procedures for monitoring multivariate time series data. Chan & Li (1994) and Charnes (1995) improved the performance of Hotelling's  $T^2$  control chart for a serially dependent process. Kalgonda & Kulkarni (2004) presented a multivariate control chart based on the residual of a Vector Autoregressive (VAR) process. Theodossiou (1993) introduced a Vector Autoregressive and Moving Average (VARMA)-based MCUSUM control chart. Kramer & Schmid (1997) proposed a Multivariate EWMA (MEWMA) control chart (Lowry et al., 1992) for a multivariate autocorrelated process. Śliwa & Schmid (2005) developed a residual-based MEWMA control chart to monitor a cross-covariance matrix of a multivariate time series process. Furthermore, Wororomi et al. (2014) and Khusna et al. (2018) developed a residual based MEWMA control chart for monitoring the mean vector shift of multivariate autocorrelated data.

When modeling a real application using the traditional time series method, it is usually hard to satisfy the particular assumption. The application of the traditional time series method also needs great expertise due to the complex structure of the autocorrelated data. In order to overcome these limitations, some researchers recommend the utilization of Support Vector Regression (SVR) as an alternative method (Sato et al., 2008; Thissen et al., 2003). The SVR algorithm has two main advantages. First, it is able to tackle both linear and nonlinear patterns data. Second, it is more reproducible since it can yield a globally optimal solution. Khediri et al. (2010) proved that a multivariate control chart based on the residuals of SVR is more effective than a

multivariate control chart based on the residuals of an Artificial Neural Network (ANN). Issam & Mohamed (2008) proved that, in comparison to ANN and VAR, an SVR-based control chart performs better.

Least Squares SVR (LS-SVR) is developed by replacing the quadratic programming problem in the SVR algorithm with a linear programming problem (Vapnik, 1998; Vapnik, 2000). Furthermore, LS-SVR changes the inequality constraints in the SVR formula for equality ones (Suykens & Vandewalle, 1999; Suykens et al., 2002). The linear equation in LS-SVR is simple to solve and useful in computational time-saving. Xu et al. (2013) proposed multioutput LS-SVR (MLS-SVR) by combining the idea of Multioutput SVR (M-SVR) (Tuia et al., 2011) and multiresponse regression (Liu et al., 2009). In order to cover the Hierarchical Bayes intuition, Xu et al. (2013) proposed an MLS-SVR algorithm which has the ability to overcome the differences in slope function for each output variable.

Several researchers have developed a MCUSUM control chart for monitoring multivariate autocorrelated data. Bodnar & Schmid (2007) proposed both a modified MCUSUM control chart (Crosier, 1988; Ngai & Zhang, 2001; Pignatiello & Runger, 1990) and a VARMA-based MCUSUM control chart. Hwang (2016) presented an MLS-SVR-based MCUSUM (Healy, 1987) control chart by considering the covariance of the error term. Hwang (2016) developed a new MLS-SVR which is different from the one proposed by Xu et al. (2013). The residuals of the MLS-SVR model resulting from in-control processes are assumed to satisfy a multivariate normal distribution as well as a white noise assumption. Therefore, the control limit of the MLS-SVR-based MCUSUM chart is equivalent to the control limit of the conventional MCUSUM chart (Healy, 1987). Hwang (2016) pointed out that the ARL of a MCUSUM chart based on the residuals of the MLS-SVR model outperforms

the ARL of a MCUSUM chart based on the residuals of the VARMA and LS-SVR model.

The objective of this research is to develop an MLS-SVR (Xu et al., 2013) based MCUSUM control chart proposed by Crosier (1988) instead of the one proposed by Healy (1987). The proposed control chart is then applied to monitor water quality data. Water turbidity and the chlorine residual are two critical quality characteristics in water manufacturing processes. These quality characteristics are recorded hourly, in which observations show serial dependency. The rest of this paper is organized as follows. Section 2 describes the MLS-SVR algorithm while section 3 presents the proposed control chart. The application of the MLS-SVR-based MCUSUM control chart for monitoring water quality data is shown in section 4. Finally, section 5 summarizes the results found in this work and presents future research.

## 2. MULTIOUTPUT LEAST SQUARES SUPPORT VECTOR REGRESSION

The basic algorithm of LS-SVR only learns the mapping from the input to a single output. If a problem solved using the LS-SVR algorithm consists of a multioutput case, then the cross-relatedness among output is disregarded. This problem inspired (Xu et al.,

2013) to develop the MLS-SVR algorithm, a multitask learning method which is useful to capture the relation between outputs. An observable output variable is defined as  $\mathbf{Y} = [y_{ij}] \in R^{n \times m}$ , for  $i = 1, 2, \dots, n$  observations and  $j = 1, 2, \dots, m$  output variables. Let  $\{(\mathbf{x}_1, \mathbf{y}^1), (\mathbf{x}_2, \mathbf{y}^2), \dots, (\mathbf{x}_n, \mathbf{y}^n)\}$  be a specific independent and identically distributed sample, where  $\mathbf{x}_i \in R^d$ ,  $\mathbf{y}^i \in R^m$ , and  $d$  defines the dimension of the input variables. Let  $\varphi: R^d \rightarrow R^h$  be a mapping function to some higher dimensional Hilbert space with  $h$  dimension. All the MLS-SVR parameters are assumed to be associated with  $\varphi(\mathbf{x})$ , so that vector  $\mathbf{w}_j \in R^h$ , for  $j \in N_m$  can be rewritten as  $\mathbf{w}_j = \mathbf{w}_0 + \mathbf{v}_j$ , where the mean vector  $\mathbf{w}_0 \in R^h$ . The small value of vector  $\mathbf{v}_j \in R^h$  for  $j \in N_m$  indicates that the output variables are similar to each other. The mean vector  $\mathbf{w}_0$  carries commonality information, while vector  $\mathbf{v}_j$  carries specialty information.

Estimating the vector  $\mathbf{w}_0 \in R^h$ , matrix  $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m) \in R^{h \times m}$ , and parameter  $\mathbf{b} = (b_1, b_2, \dots, b_m) \in R^m$  can be simultaneously obtained by minimizing the objective function with constraints as follows (Xu et al., 2013):

$$\begin{aligned} \min J(\mathbf{w}_0, \mathbf{V}, \mathbf{\Xi}) &= \frac{1}{2}(\mathbf{w}_0^T \mathbf{w}_0) + \frac{\gamma''}{2m} \text{trace}(\mathbf{V}^T \mathbf{V}) + \frac{\gamma'}{2} \text{trace}(\mathbf{\Xi}^T \mathbf{\Xi}), \\ \text{s.t } \mathbf{Y} &= \mathbf{Z}^T \mathbf{W} + \text{repmat}(\mathbf{b}^T, n, \mathbf{1}) + \mathbf{\Xi}, \end{aligned} \tag{1}$$

where  $\mathbf{Z} = (\varphi(\mathbf{x}_1), \varphi(\mathbf{x}_2), \dots, \varphi(\mathbf{x}_n)) \in R^{h \times n}$ . The matrix  $\mathbf{\Xi} = (\xi_1, \xi_2, \dots, \xi_m) \in R_+^{n \times m}$  contains the slack variables whereas the matrix  $\mathbf{W} = (\mathbf{w}_0 + \mathbf{v}_1, \mathbf{w}_0 + \mathbf{v}_2, \dots, \mathbf{w}_0 + \mathbf{v}_j, \dots, \mathbf{w}_0 + \mathbf{v}_m) \in R^{h \times m}$  illustrates the MLS-SVR parameter. The

constants  $\gamma', \gamma'' \in R_+$  are regularized parameters. The optimization problem in Equation (1) has the following Lagrange function:

$$L(\mathbf{w}_0, \mathbf{V}, \mathbf{b}, \mathbf{\Xi}, \mathbf{A}) = J(\mathbf{w}_0, \mathbf{V}, \mathbf{\Xi}) - \text{trace}(\mathbf{A}^T (\mathbf{Z}^T \mathbf{W} + \text{repmat}(\mathbf{b}^T, n, \mathbf{1}) + \mathbf{\Xi} - \mathbf{Y})), \tag{2}$$

where matrix  $\mathbf{A} = (\alpha_1, \alpha_2, \dots, \alpha_m)^T \in R^{n \times m}$  consists of the Lagrange multiplier.

Similar to LS-SVR, the MLS-SVR algorithm has a computational advantage by solving linear programming problems. The solutions of MLS-SVR are obtained by

solving two sets of linear equation systems with the same positive definite matrix  $\mathbf{M}$ , as stated in step 9 of Algorithm 1 (Xu et al., 2013).

---

Algorithm 1. Calculation of MLS-SVR solutions

---

1. Save the outputs of MLS-SVR as  $\mathbf{y} = (\mathbf{y}_1^T, \mathbf{y}_2^T, \dots, \mathbf{y}_m^T)^T \in R^{mm}$ .
  2. Specify the kernel function parameter  $\sigma$ , regularized parameters  $\gamma''$  and  $\gamma'$ .
  3. Calculate the matrix containing kernel function  $\mathbf{K} = \mathbf{Z}^T \mathbf{Z} \in R^{n \times n}$  from the inputs of MLS-SVR.
  4. Specify matrix  $\mathbf{N} = \text{blockdiag}(\overbrace{\mathbf{1}_n, \mathbf{1}_n, \dots, \mathbf{1}_n}^m) \in R^{mm \times m}$ .
  5. Specify matrix  $\mathbf{\Omega} = \text{repmat}(\mathbf{K}, m, m) \in R^{mm \times mm}$ .
  6. Specify matrix  $\mathbf{Q} = \text{blockdiag}(\overbrace{\mathbf{K}, \mathbf{K}, \dots, \mathbf{K}}^m) \in R^{mm \times mm}$ .
  7. Calculate the matrix  $\mathbf{M} = \mathbf{\Omega} + (\gamma')^{-1} \mathbf{I}_{mm} + \left(\frac{m}{\gamma''}\right) \mathbf{Q} \in R^{mm \times mm}$ .
  8. Calculate the matrix  $\mathbf{G} = \mathbf{N}^T \mathbf{M}^{-1} \mathbf{N} \in R^{m \times m}$ .
  9. Calculate  $\mathcal{G}$  and  $\nu$  from  $\mathbf{M} \mathcal{G} = \mathbf{N}$  and  $\mathbf{M} \nu = \mathbf{y}$ .
  10. Calculate  $\mathbf{G} = \mathbf{N}^T \mathcal{G}$ .
  11. Find a solution from  $\mathbf{b} = \mathbf{G}^{-1} \mathcal{G}^T \mathbf{y}$  and  $\alpha = \nu - \mathcal{G} \mathbf{b}$ .
- 

Supposing that  $\tilde{\alpha} = ((\tilde{\alpha}'_1)^T, (\tilde{\alpha}'_2)^T, \dots, (\tilde{\alpha}'_l)^T)^T$  and  $\tilde{\mathbf{b}}$  are the solutions for the MLS-SVR model, the decision function of MLS-SVR can be formulated as

$$\begin{aligned} \hat{f}(\mathbf{x}) &= \varphi(\mathbf{x})^T \tilde{\mathbf{W}} + (\tilde{\mathbf{b}})^T = \varphi(\mathbf{x})^T \text{repmat}(\tilde{\mathbf{w}}_0, 1, m) + \varphi(\mathbf{x})^T \tilde{\mathbf{V}} + (\tilde{\mathbf{b}})^T \\ &= \varphi(\mathbf{x})^T \text{repmat}\left(\sum_{j=1}^m \mathbf{Z} \tilde{\alpha}_j, 1, m\right) + \frac{m}{\gamma''} \varphi(\mathbf{x})^T \mathbf{Z} (\tilde{\mathbf{A}}) + \tilde{\mathbf{b}}^T \\ &= \text{repmat}\left(\sum_{j=1}^m \sum_{i=1}^n \tilde{\alpha}_{ij} K(\mathbf{x}, \mathbf{x}_i), 1, m\right) + \frac{m}{\gamma''} \sum_{i=1}^n \tilde{\alpha}_i K(\mathbf{x}, \mathbf{x}_i) + \tilde{\mathbf{b}}^T, \end{aligned} \tag{3}$$

where  $K(\mathbf{x}, \mathbf{x}_i)$  is a kernel function. This research employs the Radial Basis Function (RBF) kernel function. A grid search method (Hsu et al., 2016) is utilized to identify the proper hyper-parameter of the MLS-SVR model. The optimal pair of hyper-parameters  $\sigma$ ,  $\gamma'$ , and  $\gamma''$  is selected based on the criterion of the Minimum Mean Squared Error (MSE) value. The grid search method is carried out using all possible combinations of the kernel function parameter,

$\sigma \in \{2^{-15}, 2^{-13}, \dots, 2^3\}$ , as well as regularized parameters,  $\gamma'' \in \{2^{-10}, 2^{-8}, \dots, 2^{10}\}$  and  $\gamma' \in \{2^{-5}, 2^{-3}, \dots, 2^{15}\}$ .

### 3. MCUSUM CONTROL CHART BASED ON THE RESIDUALS OF MLS-SVR

The observable output variables that

satisfy multivariate autocorrelated data are defined as  $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_j, \dots, \mathbf{y}_m$ , where  $\mathbf{y}_j = (y_{1j}, y_{2j}, \dots, y_{mj})^T$  with  $j = 1, 2, \dots, m$  the number of output variables. Each output

variable is assumed to have a significant partial autocorrelation function (PACF) until lag  $p_1, p_2, \dots, p_m$  so that the input variables of the MLS-SVR model are selected as

$$\mathbf{x} = \left( \mathbf{y}_{1,(i-1)}, \dots, \mathbf{y}_{1,(i-p_1)}, \dots, \mathbf{y}_{j,(i-1)}, \dots, \mathbf{y}_{j,(i-p_j)}, \dots, \mathbf{y}_{m,(i-1)}, \dots, \mathbf{y}_{m,(i-p_m)} \right). \quad (4)$$

Let  $\hat{f}(\mathbf{x}_j)$  be the decision function of the MLS-SVR model utilizing the optimal parameters as in Equation (3). The vector of the residual can be computed as  $\mathbf{e}_j = \mathbf{y}_j - \hat{f}(\mathbf{x}_j)$ . Hence, the  $n \times m$  residual matrix  $\mathbf{e}$  consists of  $e_{ij}$ , where  $i = 1, 2, \dots, n, j = 1, 2, \dots, m$ .

Crosier (1988) presented an MCUSUM control chart in the form of the Hotelling  $T$  statistic, which is usually known as the Cumulative Sum of  $T$  (COT). If the residuals of the MLS-SVR model follow multivariate normal distribution  $N_m(\boldsymbol{\mu}_e, \mathbf{V})$  then it can be transformed into Hotelling  $T$  statistics as follows:

$$T_i = [(\mathbf{e}_i - \boldsymbol{\mu}_e)^T \mathbf{V}^{-1} (\mathbf{e}_i - \boldsymbol{\mu}_e)]^{1/2}. \quad (5)$$

Therefore, MLS-SVR residuals-based MCUSUM statistics can be calculated with the following equation:

$$C_i = \max[0, T_i - k + C_{i-1}], \quad (6)$$

where the initial value  $C_0 = 0$  and the reference value  $k > 0$ . The MLS-SVR residuals-based MCUSUM control chart detects an out-of-control signal if the  $C_i$  statistic is greater than the upper control limit  $H$ .

in the MCUSUM statistics.

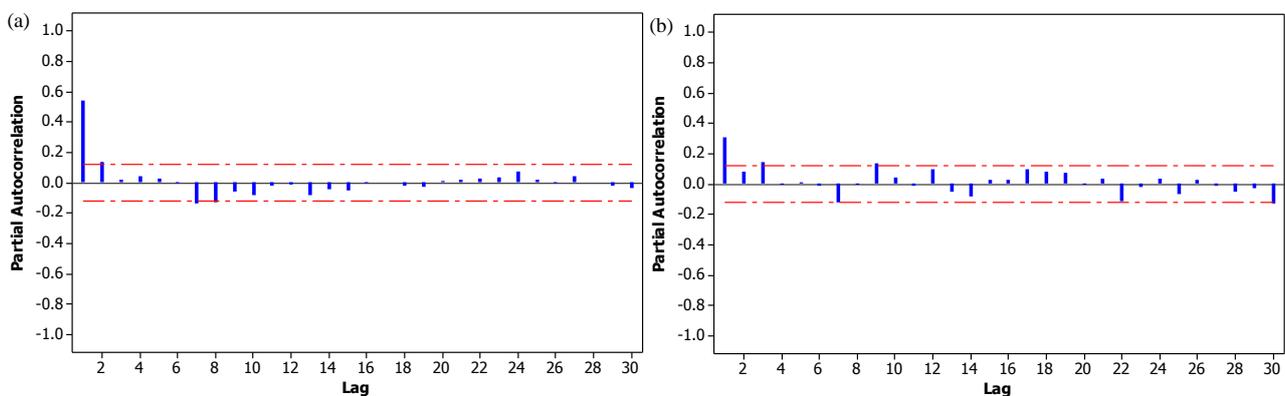
#### 4. RESULTS AND DISCUSSION

The inputs of the MLS-SVR model are selected based on in-control processes (observations from Phase 1) using Equation (4); then the hyper-parameters are optimized using the grid search method so that the residuals satisfy the white noise condition. Once this assumption is satisfied, the value of  $H$  is estimated as in Crosier (1988). The application of MCUSUM on the MLS-SVR residual will ensure the stability and adaptability of the monitoring process. The optimal parameters and hyper-parameters obtained from Phase I can be utilized directly in the Phase II monitoring process. Furthermore, the researcher can adjust the level of tightness by setting reference value  $k$

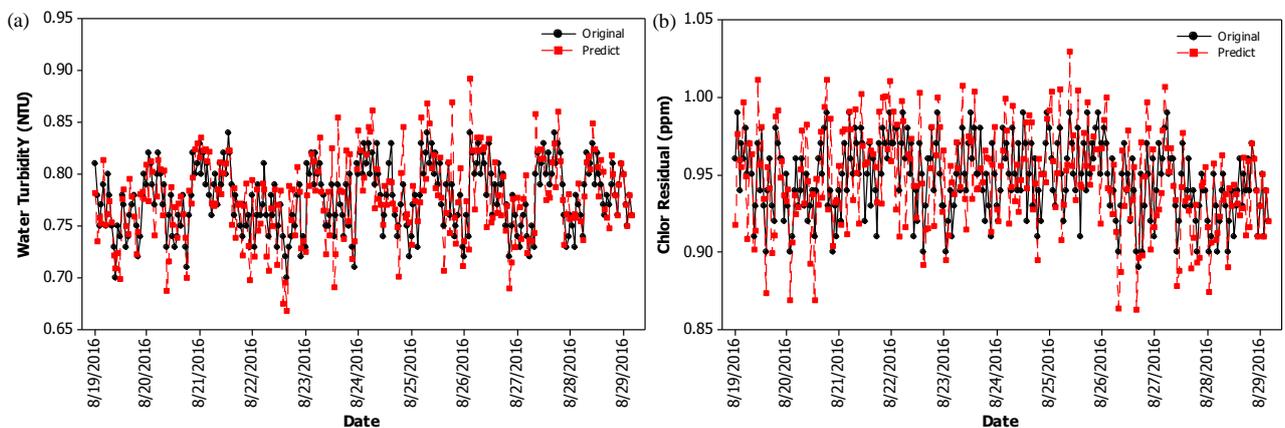
This paper presents the application of the proposed MLS-SVR-based MCUSUM chart to monitor water quality data. Two prominent quality characteristics in the drinking water manufacturing process are the water turbidity and the chlorine residual. The concentration of dissolution and the existence of particles in a liquid are usually referred to as turbidity. Chlorination is affixing chlorine into contaminated water and is principally intended for killing the microbes. Turbidity is measured using Nephelometric Turbidity Units (NTU) whereas the chlorine residual is measured in ppm unit. Drinking water is safe from bacteria if it has a minimum of 0.2 ppm chlorine residual. However, the smell and taste of water are affected by exaggerated chlorine affixation.

This research monitors both water turbidity and chlorine residual data of the water manufacturing process in Surabaya, Indonesia. The water quality data in Phase I were monitored hourly from 19 August to 29 August 2016. Figure 1 exhibits the plots of the PACF for water quality data in Phase I with a 5% significance limit. The PACF of water turbidity ( $y_1$ ) is significant at lag 1, lag 2, lag 7, and lag 8, whereas the PACF of the chlorine residual ( $y_2$ ) is significant at lag 1,

lag 3, and lag 9. Therefore,  $y_{1(i-1)}, y_{1(i-2)}, y_{1(i-7)}, y_{1(i-8)}, y_{2(i-1)}, y_{2(i-3)}, y_{2(i-9)}$  are chosen as the inputs of the MLS-SVR model for Phase I. These selected inputs along with the optimal combination of hyper-parameters  $\gamma' = 2^{-5}$ ,  $\gamma'' = 2^{-8}$ , and  $\sigma = 2^3$  produce the minimum  $MSE = 6.05 \times 10^{-4}$ . It is important to know that this research utilizes the Radial Basis Function (RBF) kernel function with a parameter  $\sigma$ .



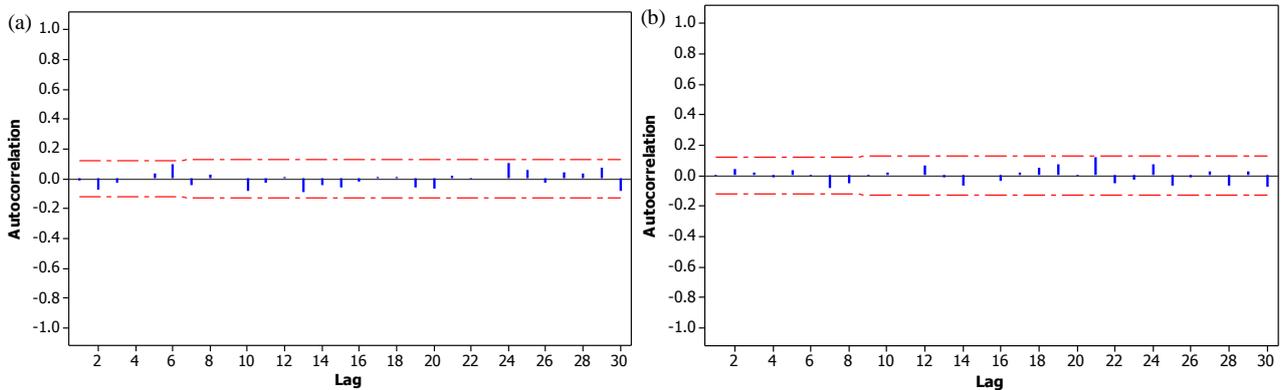
**Figure 1:** Partial autocorrelation function plots of (a) water turbidity and (b) chlorine residual in Phase I.



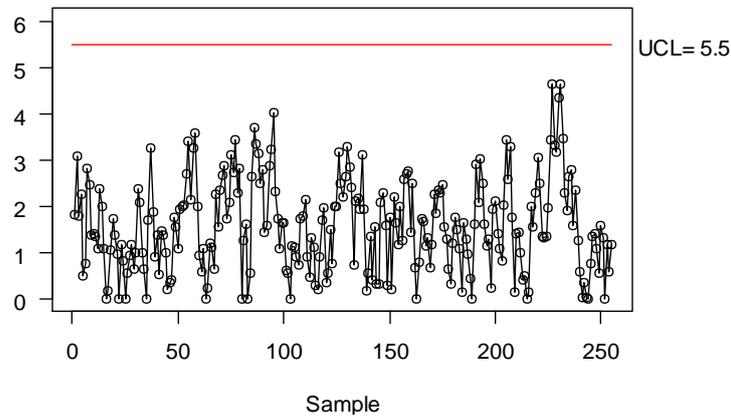
**Figure 2:** Time series plots of (a) water turbidity and (b) chlorine residual in Phase I.

The time series plots of water quality data in Phase I are displayed in Figure 2. The predicted value from the MLS-SVR model shows a similar pattern to the actual value. In addition, the residuals resulting from the MLS-SVR model for water quality data in Phase I satisfy the white noise assumption (confirmed by the ACF plots in Figure 3). The MLS-SVR residuals also fulfill a multivariate

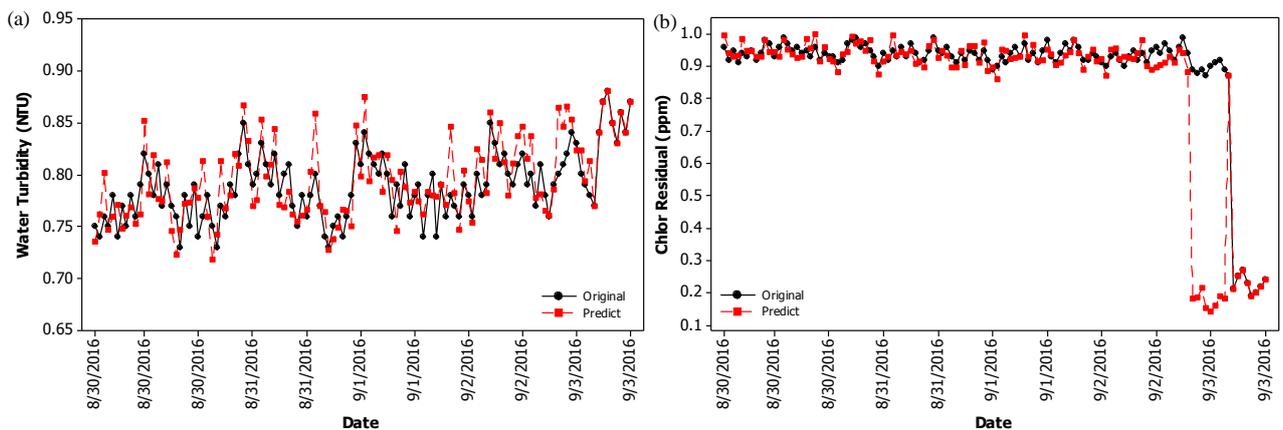
normal distribution. These residuals are then monitored using the MCUSUM control as displayed in Figure 4. It does not detect any out-of-control signal, so the upper control limit  $H$  and the parameters of the MLS-SVR-based MCUSUM control chart can be utilized in Phase II of the water manufacturing process.



**Figure 3:** Autocorrelation function (ACF) plots of MLS-SVR residuals with 5% significance limit for (a) water turbidity and (b) chlorine residual in Phase I.



**Figure 4:** Monitoring water quality data in Phase I using MLS-SVR-based MCUSUM control chart with reference value  $k=0.5$



**Figure 5:** Time series plots of (a) water turbidity and (b) chlorine residual in Phase II.

The water quality data in Phase II were monitored hourly for five days starting from 30 August 2016. The time series plots of the actual water quality data and the predicted

value from the MLS-SVR model in Phase II are displayed in Figure 5. The time series plots of actual water quality data starting from 10.00 AM on 3 September 2016 show an

unusual pattern. The increasing pattern in the water turbidity plot indicates more turbid water. On the contrary, the steeply decreasing shift in the chlorine residual plot points to the higher number of microbes in the water. Both MLS-SVR modeling needs some adaptation before it can follow the extreme changes in

indicators reflect the worsened quality of the water. When the actual values are shifted significantly, the chlorine residual predictions are much less than the actual values (see Figure 5.b). This evidence indicates that actual data.

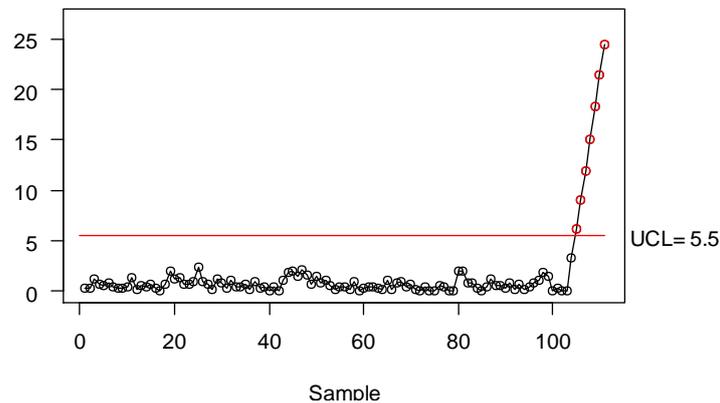


Figure 6: Monitoring water quality data in Phase II using MLS-SVR-based MCUSUM control chart with reference value  $k=0.5$

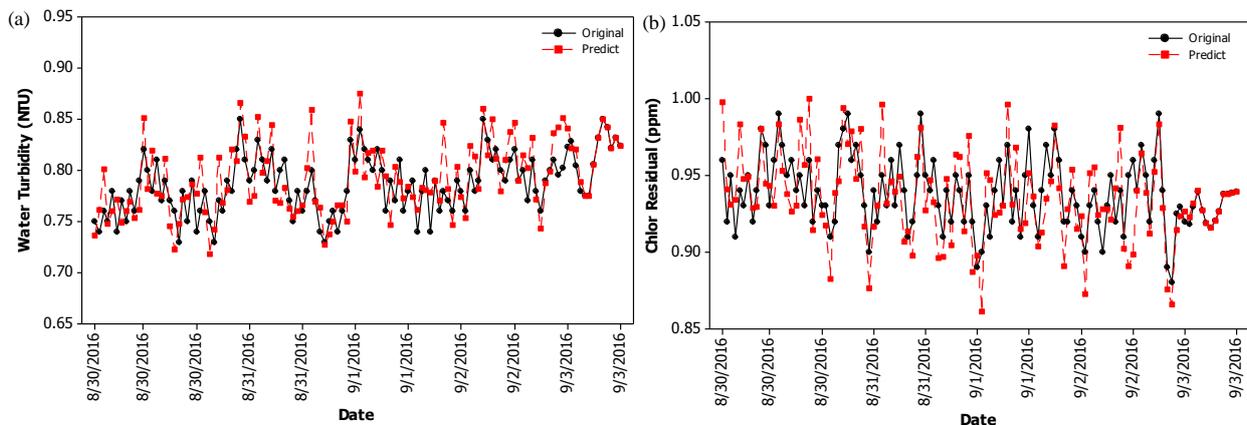


Figure 7: Time series plots of (a) water turbidity and (b) chlorine residual for updated Phase II.

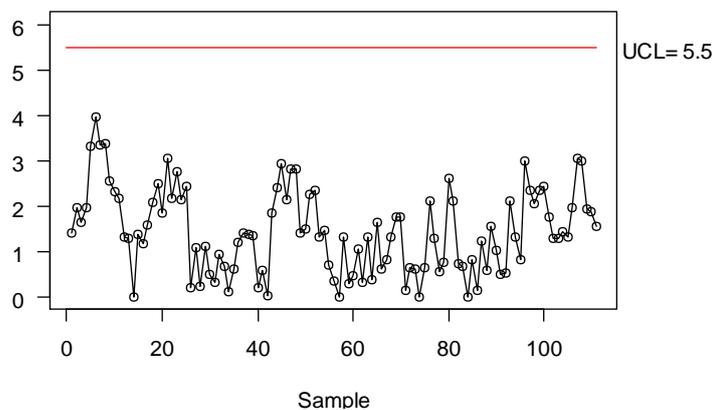


Figure 8: Monitoring updated water quality data in Phase II using MLS-SVR-based MCUSUM control chart with reference value  $k=0.5$

The water quality data in Phase II are trained using the MLS-SVR algorithm by utilizing the optimal parameters and hyper-parameters produced in Phase I. As depicted in Figure 6, the last eight statistics of the MLS-SVR-based MCUSUM control chart in Phase II show significant out-of-control signals. This result indicates a serious problem in both the water turbidity and chlorine residuals data. It is necessary to look for the assignable cause in order to improve the water manufacturing process. At that time point, a broken pipeline was found in one of the water distribution areas. Pipeline maintenance and flow meter installation were then conducted in order to handle the water manufacturing problem.

The MLS-SVR-based MCUSUM control chart needs to be updated in order to fit the new observations in the further monitoring process. That is why the actual unusual pattern displayed in Figure 5 needs to be replaced with the predicted value from the MLS-SVR model. The updated water quality data in Phase II (see Figure 7) consist of actual water quality data in Phase II, except for the data starting from 10.00 AM on 3 September 2016. The actual data in those periods are replaced by the predicted values from the MLS-SVR model. The MLS-SVR-based MCUSUM control chart for the updated Phase II is exhibited in Figure 8. All of the MLS-SVR-based MCUSUM statistics are assigned as being in the in-control condition such that these updated data can be used in the next Phase I monitoring process.

## 5. CONCLUSION

This paper develops a MCUSUM control chart based on the residual of the MLS-SVR model for monitoring the mean vector of time series data. The inputs of the MLS-SVR model are determined based on the significant lag of the PACF. The appropriate inputs and the optimal combination of hyper-parameters yield the residuals of the MLS-

SVR model that fulfill the white noise assumption. The MCUSUM control chart based on the residual of MLS-SVR that is used to monitor the water quality data indicates that corrective actions should be carried out in order to improve the water manufacturing process. Evaluating the performance of the proposed control chart using the Average Run Length (ARL) criterion may be useful in future research. Similarly, optimizing the SVR parameters using an evolutionary algorithm (Härdle et al., 2014) may also be useful in future work.

## 6. ACKNOWLEDGEMENT

This research was supported by Research, Technology, and Higher Education Ministry, the Republic of Indonesia through PMDSU Scheme [Grant number 128/SP2H/PTNBH/DRPM/2018].

## 7. REFERENCES

- Bodnar O., and Schmid W. (2007). Surveillance of the mean behavior of multivariate time series, *Statistica Neerlandica* 61 (4):383–406.
- Chan L K., and Li G-Y. (1994). A multivariate control chart for detecting linear trends, *Communications in Statistics-Simulation and Computation* 23 (4):997–1012.
- Charnes J M. (1995). Tests for special causes with multivariate autocorrelated data, *Computers and Operations Research* 22 (4):443–453.
- Crosier R B. (1988). Multivariate generalizations of cumulative sum quality-control schemes, *Technometrics* 30 (3):291–303.
- Härdle W K., Prastyo D D., and Hafner C M. (2014). Support vector machines with

- evolutionary model selection for default prediction. *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*. Oxford University Press.
- Harris T J., and Ross W H. (1991). Statistical process control procedures for correlated observations, *Canadian Journal of Chemical Engineering* 69 (1):48–57.
- Healy J D. (1987). A note on multivariate CUSUM procedures, *Technometrics* 29 (4):409–412.
- Hsu C W., Chang C C., and Lin C J. (2016). A practical guide to support vector classification. *National Taiwan University*.
- Hwang C. (2016). Multioutput LS-SVR based residual MCUSUM control chart for autocorrelated process, *Journal of the Korean Data and Information Science Society* 27 (2):523–530.
- Issam B K., and Mohamed L. (2008). Support vector regression based residual MCUSUM control chart for autocorrelated process, *Applied Mathematics and Computation* 201 (1–2):565–574.
- Johnson R A., and Bagshaw M. (1974). The effect of serial correlation on the performance of CUSUM tests, *Technometrics* 16 (1):103–112.
- Kalgonda A A., and Kulkarni S R. (2004). Multivariate quality control chart for autocorrelated processes, *Journal of Applied Statistics* 31 (3):317–327.
- Khediri I B., Weihs C., and Limam M. (2010). Support vector regression control charts for multivariate nonlinear autocorrelated processes, *Chemometrics and Intelligent Laboratory Systems* 103 (1):76–81.
- Khusna H., Mashuri M., Prastyo D D., and Ahsan M. (2018). Multioutput least square SVR based multivariate EWMA control chart, *Journal of Physics: Conference Series* 1028:12221. IOP Publishing.
- Kramer H G., and Schmid L V. (1997). EWMA charts for multivariate time series, *Sequential Analysis* 16 (2):131–154.
- Liu G., Lin Z., and Yu Y. (2009). Multi-output regression on the output manifold, *Pattern Recognition* 42 (11):2737–2743.
- Lowry C A., Woodall W H., Champ C W., and Rigdon S E. (1992). A multivariate exponentially weighted moving average control chart, *Technometrics* 34 (1):46–53.
- Ngai H-M., and Zhang J. (2001). Multivariate cumulative sum control charts based on projection pursuit, *Statistica Sinica* 747–766.
- Noorossana R., and Vaghefi S J M. (2006). Effect of autocorrelation on performance of the MCUSUM control chart, *Quality and Reliability Engineering International* 22 (2), 191–197.
- Pignatiello J J., and Runger G C. (1990). Comparisons of multivariate CUSUM charts, *Journal of Quality Technology* 22 (3):173–186.
- Psarakis S., and Papaleonida G. (2007). SPC procedures for monitoring autocorrelated processes, *Quality Technology and Quantitative Management* 4 (4):501–540.

- Sato J R., Costafreda S., Morettin P A., and Brammer M J. (2008). Measuring time series predictability using support vector regression, *Communications in Statistics-Simulation and Computation* 37 (6):1183–1197.
- Śliwa P., and Schmid W. (2005). Monitoring the cross-covariances of a multivariate time series, *Metrika* 61 (1):89–115.
- Suykens J A K., Van-Gestel T., De-Brabanter J., De-Moor B., and Vandewalle J. (2002). Least squares support vector machines. *World Scientific*.
- Suykens J A K., and Vandewalle J. (1999). Multiclass least squares support vector machines, *IEEE* 2:900-903.
- Theodossiou P. (1993). Predicting shifts in the mean of a multivariate time series process: an application in predicting business failures, *Journal of the American Statistical Association* 88:441–449.
- Thissen U., Van-Brakel R., De-Weijer A P., Melssen W J., and Buydens L M C. (2003). Using support vector machines for time series prediction, *Chemometrics and Intelligent Laboratory Systems* 69 (1–2):35–49.
- Tuia D., Verrelst J., Alonso L., Perez-Cruz F., and Camps-Valls G. (2011). Multioutput support vector regression for remote sensing biophysical parameter estimation, *IEEE Geoscience and Remote Sensing Letters* 8 (4):804–808.
- Vanbrackle L N., and Reynolds M R. (1997). EWMA and CUSUM control charts in the presence of correlation, *Communications in Statistics-Simulation and Computation* 26 (3):979–1008.
- Vapnik V N. (1998). *Statistical Learning Theory*. John Wiley & Sons.
- Vapnik V N. (2000). *The Nature of Statistical Learning Theory* 8.
- Woodall W H., and Montgomery D C. (1999). Research issues and ideas in statistical process control, *Journal of Quality Technology* 31 (4):11.
- Wororomi J K., Mashuri M., Irhamah, and Arifin A Z. (2014). On monitoring shift in the mean processes with vector autoregressive residual control charts of individual observation, *Applied Mathematical Sciences* 8:3491–3499.
- Xu S., An X., Qiao X., Zhu L., and Li, L. (2013). Multi-output least-squares support vector regression machines, *Pattern Recognition Letters* 34 (9):1078–1084.
- Yashchin E. (1993). Performance of CUSUM control schemes for serially correlated observations. *Technometrics* 35 (1):37–52.