# Monitoring the Quality of Water Production Process in Surabaya Using Max-Mcusum Control Chart Based on Residual Deep Learning LSTM Model

Veneza Rafa Aliyah and Muhammad Ahsan



August 26, 2024

# Monitoring The Quality of Water Production Process in Surabaya Using Max-MCUSUM Control Chart based on Residual Deep Learning LSTM Model

Veneza Rafa Aliyah[1], Muhammad Ahsan[2]

[1] Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Indonesia
venezarafaaliyah24@gmail.com
[2]Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
muh.ahsan@its.ac.id

**Abstract.** This study examines the water quality at the Surya Sembada water treatment plant in Surabaya, Indonesia, by analyzing turbidity, pH, permanganate index, and chlorine residual. Recognizing the inherent autocorrelation within these parameters, a Long Short-Term Memory (LSTM) neural network was implemented to model their temporal dependencies. Optimal LSTM hyperparameters were determined through rigorous experimentation using MSE, RMSE, and MAE as evaluation metrics. Residuals from the LSTM model was subsequently analyzed using a Maximum Multivariate Cumulative Sum (MCUSUM) control chart. Phase I analysis indicated a statistically non-conforming process, suggesting a significant process shift. Subsequent Phase II monitoring confirmed ongoing process instability. The application of LSTM modeling and Max-MCUSUM control charting in this study provides a robust framework for early detection of anomalies and process deviations in water treatment operations, facilitating timely corrective actions and improvements in water quality management.
**Keywords:** Control Chart, LSTM, Max-MCUSUM, Statistical Quality

## 1 Introduction

Sustainable water resource management is critical for mitigating the challenges posed by rapid urbanization in cities like Surabaya. Balancing escalating water demands with resource conservation is essential to preserving public health and ecological integrity. Implementing robust strategies to safeguard both water quantity and quality is paramount for the city's long-term sustainability. Surya Sembada, a state-owned water treatment plant established in 1976, is mandated to provide potable water to the Surabaya metropolitan area. As a vital public utility, it plays a substantial role in the local economy, contributing significantly to Regional Original Revenue. To safeguard public health, Surya Sembada adheres to stringent water quality standards as outlined in Indonesia's Ministry of Health Regulation No.

492/MENKES/PER/IV/2010.

The Surabaya River is characterized by a deteriorated water quality profile, primarily attributed to elevated levels of organic pollutants and an aging wastewater treatment infrastructure. Comprehensive assessments utilizing the Pollution Index (PI) and STORET methodologies have consistently categorized the river's water quality as ranging from moderately to severely polluted, thereby emphasizing the urgent need for remediation interventions.

To address this, Statistical Quality Control (SQC) methods, such as control charts, monitor water treatment quality[1]. Analysis of water quality data reveals the presence of autocorrelation, rendering traditional monitoring methods less effective. A variety of statistical and machine-learning techniques can be employed to address this issue. Traditional time series methods such as Vector Autoregression (VAR) models [2], as well as advanced machine learning algorithms including Artificial Neural Networks (ANN) [3, 4], Multioutput Least Squares Support Vector Regression (MLS-SVR) [5, 6], XGBoost [7], and Long Short-Term Memory (LSTM) [8, 9], offer potential solutions. Consequently, this study proposes a multivariate control chart methodology that leverages residuals from time series-based machine learning models to address the complexity of multivariate data [10, 11]. This research introduces a multivariate control chart approach that utilizes residuals derived from time series deep learning techniques to monitor and manage multivariate data effectively [12, 13].

Comprehensive water quality assessment necessitates the application of multivariate control charts to simultaneously monitor multiple quality parameters. To address the inherent autocorrelation present in water quality time series data, this study proposes a novel approach combining the Max-MCUSUM chart [14–16] with a Long Short-Term Memory (LSTM) model. The efficacy of this methodology was evaluated at the Ngagel II water treatment plant with the objective of guaranteeing the delivery of high-quality water, thereby contributing to public health and environmental sustainability in Surabaya.

This investigation employs a Max-MCUSUM control chart integrated with an LSTM residual model for the surveillance of water quality. The study aims to evaluate the efficacy of this methodology in monitoring water quality parameters. The analysis is circumscribed to the calendar year 2022, utilizing hourly time-step inputs within the LSTM model for the examination of quality control data. Specific objectives encompass the development and implementation of an LSTM model for the water production process, the optimization of the Max-MCUSUM control chart through LSTM residual integration to minimize false alarms, and the statistical assessment of process capability to ensure adherence to established water quality standards.

# 2 Literature Review

## 2.1 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks are a specialized variant of Recurrent Neural Networks (RNNs) designed to mitigate the vanishing gradient problem inherent in standard RNN architectures when processing sequential data. By incorporating a sophisticated cell state regulated by input, forget, and output gates, LSTMs effectively capture and maintain long-term dependencies. This architectural innovation empowers LSTMs to excel in tasks demanding the preservation of information over extended temporal intervals, such as time series forecasting and handwriting recognition [8].

## 2.2 Maximum Multivariate Cumulative Sum (Max-MCUSUM) Control Chart based on the Long Short-Term Memory (LSTM) residual

A control chart is a fundamental tool in Statistical Process Control (SPC) employed to graphically represent process variability over time. By plotting sample means of a quality characteristic, it establishes a centerline indicative of the expected process average, flanked by upper and lower control limits. Excursions of sample means beyond these control limits signal the presence of anomalous variability sources, prompting investigative and corrective actions. Consistent application of control charts facilitates a systematic reduction in process variability [17].

The Cumulative Sum (CUSUM) control chart, introduced by Page in 1954 [18], is employed to detect subtle shifts in process mean or variance attributed to assignable causes. Thaga subsequently extended this to the Max-CUSUM control chart. For processes exhibiting correlated quality characteristics, multivariate control charts such as Max-MCUSUM have been proposed to monitor both central tendency and dispersion. While Max-MCUSUM assumes independent, normally distributed data, it demonstrates relative robustness to deviations from normality. However, autocorrelation in data can lead to an elevated false alarm rate, diminishing the chart's efficacy. To mitigate this, Long Short-Term Memory (LSTM) models can be leveraged to extract residuals from the autocorrelated data, which are then incorporated into the Max-MCUSUM control chart to enhance monitoring performance.

$$e_i = y_i - f(y_{f_{i-1}}, y_{f_{i-2}}, y_{f_{i-m}}) \tag{1}$$

Where $e_i$ is residual at time $-i$, $y_i$ is actual value at time$-i$ , $y_{f_{i-1}}, y_{f_{i-2}}, y_{f_{i-m}}$ is the value at the previous time with a distance of m time periods backward, $f$ is The function used to predict actual values. We need to create a random variable $Z_i$that can represent changes in the data calculated based on the residual vector, namely:

$$Z_i = \frac{(\boldsymbol{\mu}_{e(b)} - \boldsymbol{\mu}_{e(g)})^T \boldsymbol{\Sigma}_{e(g)}^{-1} (\boldsymbol{e}_i - \boldsymbol{\mu}_{e(g)})}{[\left(\boldsymbol{\mu}_{e(b)} - \boldsymbol{\mu}_{e(g)}\right)^T \boldsymbol{\Sigma}_{e(g)}^{-1} \left(\boldsymbol{\mu}_{e(b)} - \boldsymbol{\mu}_{e(g)}\right)]^{\frac{1}{2}}} \tag{2}$$

Where, $\boldsymbol{\mu}_{e(b)}$ is Mean vector when the residual comes from the out-of-control limit (uncontrolled), $\boldsymbol{\mu}_{e(g)}$ is Mean vector when the residual comes from the in-control limit (controlled), $\boldsymbol{\Sigma}_{e(g)}$ is Covariance matrix included (in-control)

If $Z_i$ do not follow a standard normal distribution, then $Z_i$follows a normal distribution with a mean of λ. The non-centrality parameter λ can be calculated as in Equation 3:

$$\lambda = [(\boldsymbol{\mu}_{e(b)} - \boldsymbol{\mu}_{e(g)})^T \boldsymbol{\Sigma}_{e(g)}^{-1} (\boldsymbol{\mu}_{e(b)} - \boldsymbol{\mu}_{e(g)})]^{\frac{1}{2}} \tag{3}$$

the vector experiences a shift in the covariance matrix $\Sigma_{e(b)}$, expressed by:

$$W_i = \Phi^{-1}\left\{H\left[(\boldsymbol{e}_i - \boldsymbol{\mu}_{e(g)})^T \boldsymbol{\Sigma}_{e(b)}^{-1}\left(\boldsymbol{e}_i - \boldsymbol{\mu}_{e(g)}\right); m\right]\right\} \tag{4}$$

where $\Phi(z) = P(Z \leq z)$ with $Z \sim N(0,1)$. $H(x;\, m) = P(X \leq x|m)$ with $X \sim X^2{}_m$ and the function $\Phi^{-1}$ is the inverse of the cumulative distribution function (CDF) of the standard normal distribution can be expressed by.

$$\Phi(x) = P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \tag{5}$$

If $Zi$ and $Wi$ are independent and follow a normal distribution.

$$M_{r(i)} = max\ [C_{r(i)}, S_{r(i)}] \tag{6}$$

Where, $C_{r(i)}$ is transformation from random variable $Zi$, $S_{r(i)}$ is transformation from random variable $Wi$. Thus, $C_{r(i)}$ and $S_{r(i)}$can be expressed:

$$\begin{gathered} C_{r(i)} = max\left[C_{r(i)}{}^{+}, C_{r(i)}{}^{-}\right] \\ C_{r(i)}{}^{+} = max\left[0, Z_i - k + C^{+}{}_{(r)i-1}\right] \\ C_{r(i)}{}^{-} = max[0, -k - Z_i + C^{-}{}_{(r)i-1}] \end{gathered} \tag{7}$$

$$\begin{gathered} S_{r(i)} = max\ [S_{r(i)}{}^{+}, S_{r(i)}{}^{-}] \\ S_{r(i)}{}^{+} = max\left[0, W_i - k + S^{+}{}_{(r)i-1}\right] \\ S_{r(i)}{}^{-} = max[0, -k - W_i + S^{-}{}_{(r)i-1}] \end{gathered} \tag{7}$$

Max-MCUSUM in LSTM models compare the statistic to the upper control limit (UCL) when $M_{r(i)} \geq 0$, detecting out-of-control signals when M(i) exceeds the UCL. The UCL is estimated using the bootstrap method.

$$UCL = \frac{1}{N'} \sum_{l=1}^{N'} M^l{}_{(r)(100(1-\alpha))} \tag{2.377}$$

Where $M^l{}_{(100(1-\alpha))} = M^l{}_{((1-\alpha)B)}$ and $l = 1, 2, \ldots, N'$ replication.

# 3 Methodology

## 3.1 Data Source

This study employs secondary water quality data obtained from Surya Sembada Surabaya spanning the period from January 1, 2022, to December 31, 2022. The dataset is divided into two phases: Phase I (January 1 to July 31, 2022) and Phase II (August 1 to December 31, 2022). Water samples were collected post-filtration and subsequently analyzed in a laboratory setting. Four water quality parameters were investigated: turbidity, pH, organic matter (measured as KMnO4 consumption), and residual chlorine.

### 3.2 Research Variable

The research variables employed in this study encompass four distinct quality characteristics, as detailed in Table 1.

**Table 1.** Research Variables

| Variables | Explanation | Unit | Specification |
|---|---|---|---|
| $y_1$ | Turbidity | NTU | $0 - 1$ |
| $y_2$ | pH | - | $6.5 - 8.5$ |
| $y_3$ | Organic Matter (MnO4) | Mg/l | $0 - 10$ |
| $y_4$ | Chlorine Residual | ppm | $0.5 - 1$ |

### 3.3 Analysis Steps

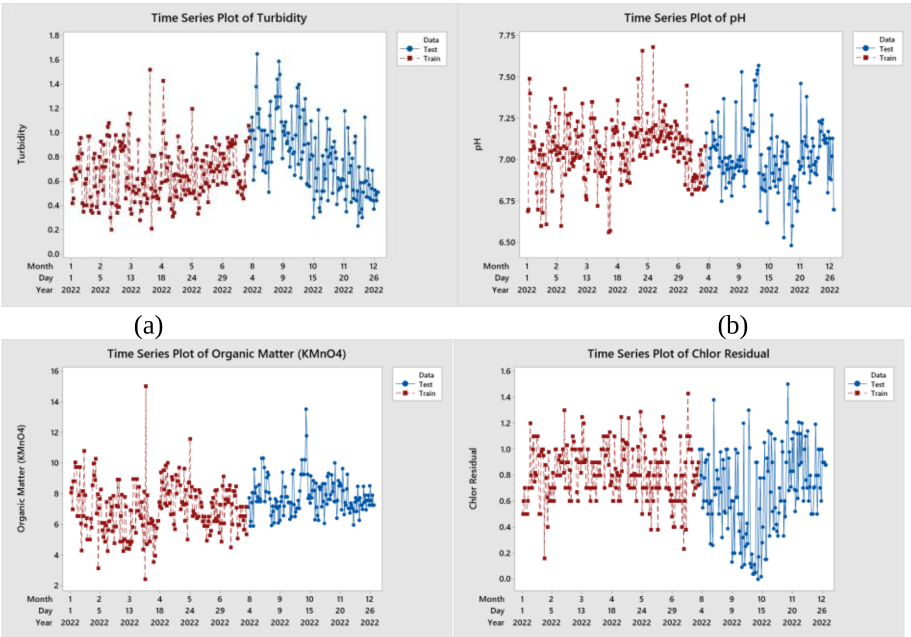
This study employs an LSTM residual-based Max-MCUSUM control chart to monitor water quality. The methodology comprises the following steps:

1. **Data Acquisition**: Secondary water quality data encompassing turbidity, pH, organic matter (KMnO4), and chlorine residual were collected from Surya Sembada Surabaya for the period of January 1 to December 31, 2022.
2. **Time Series Modeling**: A Long Short-Term Memory (LSTM) neural network was constructed to model the time series behavior of the water quality parameters. Data preprocessing, including scaling and splitting into training and testing sets, was performed prior to model development. Hyperparameter tuning was conducted to optimize model performance.
3. **Model Validation and Control Chart Implementation**: The adequacy of the LSTM model was assessed through the Shapiro-Wilk test for normality and the Portmanteau test for autocorrelation. Subsequently, a Max-MCUSUM control chart was applied to the model residuals to monitor for process shifts.
4. **Conclusion and Recommendations**: Based on the control chart analysis, conclusions regarding water quality stability were drawn and appropriate recommendations for process improvement were formulated.

## 4 Analysis and Discussion

### 4.1 LSTM Modelling

Figure 1 illustrates the time series plot of the studied parameters during Phase I (red) and Phase II (blue) from January 1 to December 31, 2022. Turbidity exhibited a declining trend from September to December, while Organic Matter (KMnO4) demonstrated sporadic increases throughout the year. pH measurements indicated an elevation in May followed by a decline in October. Residual Chlorine levels remained relatively consistent with a minor reduction observed between August and October.




(c) (d)

**Fig. 1.** Time Series Plot of the Data of Quality Characteristics: (a) Turbidity, (b)pH, (c)Organic Matter (KMnO4), (d) Chlorine Residual

Prior to autocorrelation analysis, standardization of the three quality characteristics was performed to account for disparate measurement units. Autocorrelation was evaluated through the application of Multivariate Cross-Correlation Function (MCCF) plots, as visualized in Figure 2 for each water quality parameter. The MCCF plots in Figure 2 unequivocally demonstrate the presence of significant autocorrelation in all four water quality parameters at the Ngagel II water treatment plant, indicating complex interrelationships among these variables.

Schematic Representation of Cross Correlations

| Variable/Lag | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Turbidity | +..- | +..- | +..- | +..- | +... | +..- | +.+- | +.+- | +.+- | +..- | +.+- | +-+- | +.+- | +.+- | +.+- | +.+- |
| pH | .+.. | .+.. | .+.. | .+.. | .+.. | .+.- | .+.. | .+.. | .... | .... | .... | .... | .... | .... | .... | .... |
| OrganicMatter | ..+- | +.+- | ..+- | ..+. | .-+- | .-+- | .-+- | ..+. | ..+. | .... | .... | .-.. | .-.. | .... | .... | .... |
| ChlorResidual | -.-+ | -.-+ | -.-+ | -.-+ | -.-+ | -.-+ | -+-+ | -.-+ | .+-+ | -.-+ | ..-+ | .... | ...+ | ..-+ | -+-+ | -+.+ |

+ is > 2*std error, - is < -2*std error, . is between

**Fig. 2.** Multivariate Cross-Correlation Function of the Data of Quality Characteristics: (a)Turbidity, (b)pH, (c)Organic Matter (KMnO4), (d) Chlorine Residual

Long Short-Term Memory (LSTM) modeling was employed to address autocorrelation within Surya Sembada's production process at the Ngagel II water treatment plant. Model architecture and hyperparameters were meticulously optimized to minimize residual error. Phase-1 model configurations were subsequently applied to phase-2 data. Prior to LSTM modeling, input data underwent stationarity and scaling transformations. A comprehensive experimental design was implemented to identify optimal LSTM architecture and hyperparameters based on minimizing Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE), as outlined in Table 2. To achieve optimal model fit, the training process was iterated for 250 epochs. The architecture and hyperparameters of the selected LSTM model are presented in Appendix 1.

**Table 2** Evaluation Metrics

| Evaluation Metric | Value |
|---|---|
| The Average of Mean Squared Error (MSE) | 0.049 |
| The Average of Root Mean Squared Error (RMSE) | 0.222 |
| The Average of Mean Absolute Error (MAE) | 0.170 |

### 4.1.1 LSTM Modelling for Phase I

After modeling using LSTM in phase 1, the predicted values were compared with the actual data values, which can be seen in Figure 4.3. Figure 4.3 depicts the predicted values of Chlorine Residual, Organic Matter (KMnO4), pH, and Turbidity from LSTM modeling, mirroring the pattern of actual data in phase I. This suggests the reliability of LSTM modeling as it captures similar data patterns.

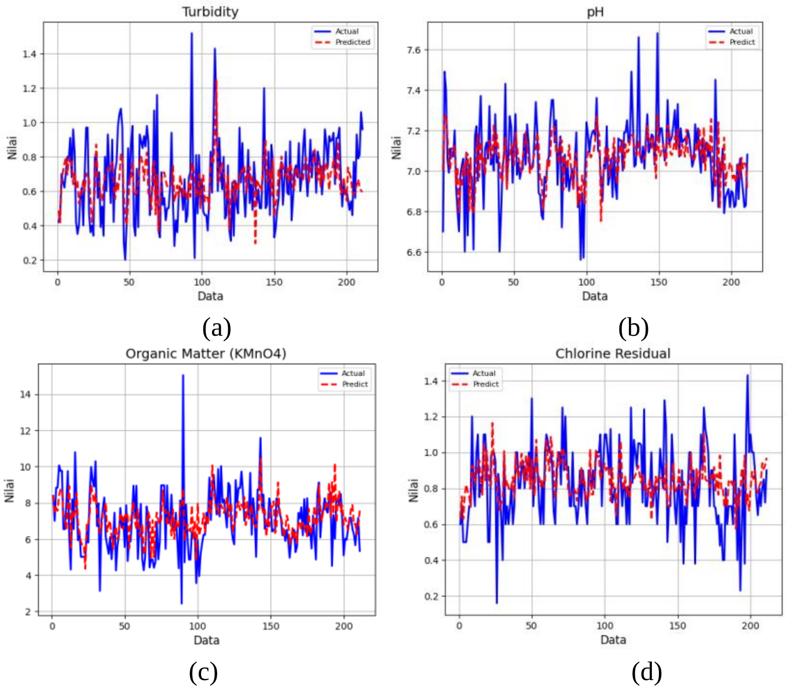



**Fig. 2** Line Plot of Actual Data vs Predicted Data of the Quality Characteristics: (a) Turbidity, (b)pH, (c)Organic Matter (KMnO4), (d) Chlorine Residual in Phase I

### 4.1.2 LSTM Modelling for Phase II

LSTM modeling in Phase II employs identical architecture and hyperparameters as in Phase I, covering water production data from August 1 to December 31, 2022.

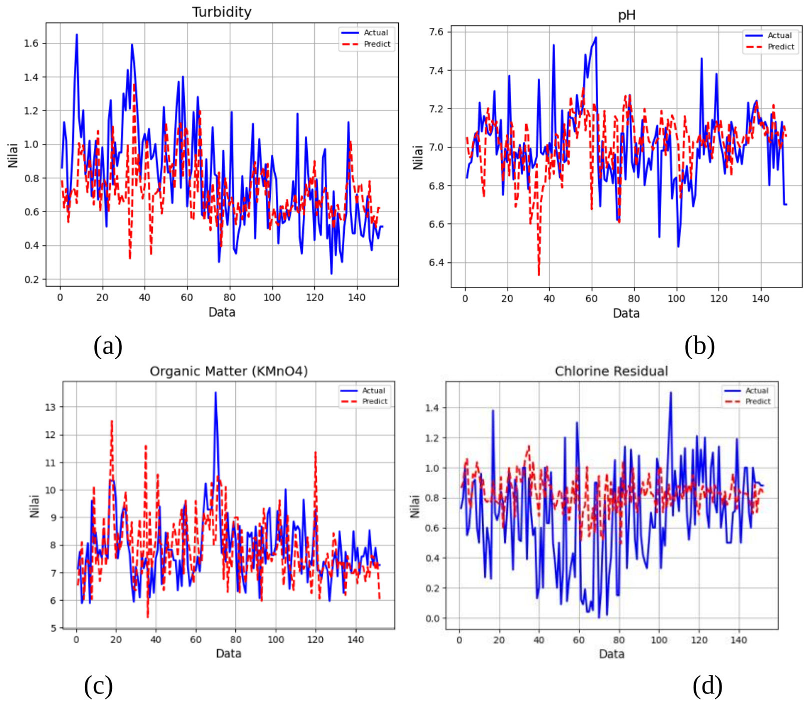
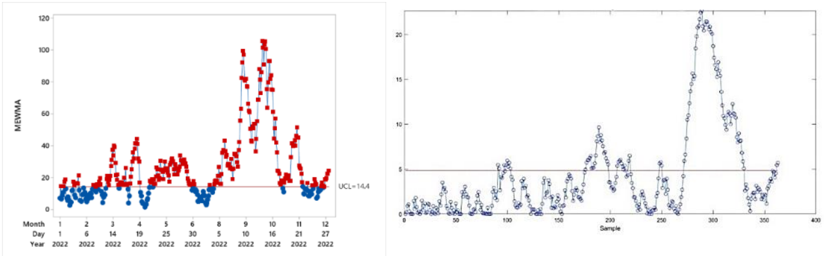



**Fig. 3.** Line Plot of Actual Data vs Predicted Data of the Quality Characteristics: (a) Turbidity, (b)pH, (c)Organic Matter (KMnO4), (d) Chlorine Residual in Phase II

Figure 3 depicts a comparative analysis of predicted and actual values for Chlorine Residual, KMnO4, and turbidity. While Chlorine Residual and KMnO4 exhibit comparable trends, a systematic overestimation of Chlorine Residual and underestimation of turbidity is evident. The observed discrepancies in the time series patterns of all four quality characteristics during the phase II period necessitate a more rigorous investigation using a control chart methodology. To this end, a Max-MCUSUM control chart based on LSTM model residuals is employed.

## 4.2 Max-MCUSUM Control Chart Based on LSTM Model Residual

During the initial phase of water production, traditional quality control methods aimed to maintain process stability within predefined control limits. However, as illustrated in Figure 4, all data points exhibited exceedance of these limits due to the presence of unaccounted autocorrelation, resulting in an elevated false alarm rate. To mitigate this issue, residual data derived from a phase I LSTM model were incorporated into a Max-MCUSUM control chart. This multivariate control chart, capable of simultaneously monitoring both process mean and variance, demonstrated superior sensitivity in detecting subtle process shifts, thereby improving overall process control.




(a) (b)

**Fig. 4** Control Chart using the Actual Data: (a) MEWMA Control Chart (b) Max M-CUSUM Control Chart

The Max-MCUSUM control chart integrates CUSUM and MCUSUM methodologies to provide simultaneous monitoring of process mean and variance. Employing exclusively upper control limits, the chart relies on bootstrap resampling to determine control limits, evaluated using Average Run Length (ARL) criteria. An ARL0 of 370 was achieved with a significance level (α) of 0.00273. In the initial phase (Phase I), the Max-MCUSUM chart is applied to LSTM model residuals to establish process control and determine initial control limits. These limits serve as reference points for subsequent monitoring (Phase II), where the bootstrap method is used for estimation.

#### 4.2.1 Choosing the optimal reference value *k*.

The reference value, k, of $0 < k \leq 1.5$, chosen for its sensitivity in detecting out-of-control points. Actual observation data surpassing specification limits are utilized to test various $k$ values and determine the optimal one.

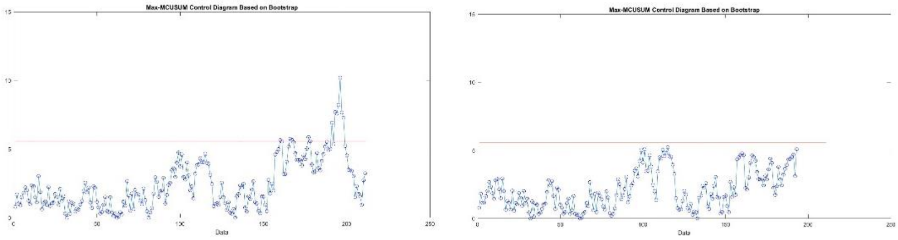
**Table 3**. Number of Out-of-Control Points for Residual Data of Max-MCUSUM Control Chart

| $k$ | UCL | Number of Out-of-Control Points |
|---|---|---|
| 0.25 | 91.2310 | 12 |
| 0.50 | 63.2456 | 15 |
| 0.75 | 40.6152 | 12 |
| 1.00 | 22.2152 | 12 |
| 1.25 | 11.2036 | 14 |
| 1.50 | 6.5401 | 11 |

Table 3 demonstrates that a k-value of 0.5 optimizes the control chart's performance, accurately identifying 15 data points as out-of-control, which closely align with actual instances of specification exceedance.

### 4.2.2 Max MCUSUM chart for Phase I

Based on the obtained reference value k, the optimal hyper-parameter, the control limit utilized in this study is 5. 5655. The following are the results obtained after processing using the Max-MCUSUM control chart in phase I.




(a) (b)

**Fig. 5** Max-MCUSUM Control Chart for Phase I Data: a) Initial chart b) Revised chart

Figure 5 (a) illustrates multiple data points exceeding the upper control limit, suggesting a state of process instability. The Max-MCUSUM control chart, implemented in Phase 1 with a reference value k of 0.5, provides statistical confirmation of an out-of-control process for both methodologies under evaluation. These findings indicate the presence of assignable causes of variation, necessitating in-depth diagnostic analysis to identify and eliminate root causes. Subsequent replacement of anomalous observations with the target value yielded consistent results from the Max-MCUSUM chart, maintaining a state of statistical non-control as evidenced by multiple out-of-control points in Figure 5 (b).

### 4.2.3 Max MCUSUM chart for Phase II

A Phase II quality control analysis was conducted on Surya Sembada's water quality data spanning August to December 2022. A Max-MCUSUM control chart, informed by LSTM-derived Phase I parameters, was employed for the evaluation.

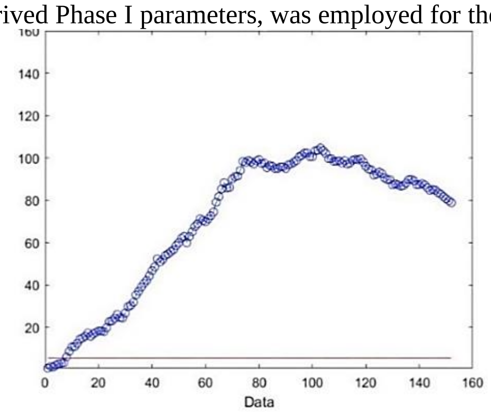



**Fig. 6** Max-MCUSUM Control Chart Based on Bootstrap for Phase II Data

Figure 6 presents a Phase II Max-MCUSUM control chart (*k* = 0.5) for the water production process at the water treatment plant. A notable departure from statistical control is evident from August 9, 2022, as numerous data points transgress the established control limits. This anomalous pattern strongly indicates a shift in the mean of the water production process, signifying a state of instability. The occurrence of this out-of-control signal coincides with the onset of the rainy season in Surabaya during that period, suggesting a potential correlation between meteorological conditions and process variability. Further investigation is warranted to elucidate the precise mechanisms underlying this relationship and to develop strategies for mitigating the impact of rainfall on water production

## 5 Conclusion and Suggestions

This study introduces a novel hybrid monitoring framework that integrates a Long Short-Term Memory (LSTM) model with a Max-MCUSUM control chart for the rigorous surveillance of water quality at the Surya Sembada water treatment plant in Surabaya. The LSTM model, optimized through minimization of Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE), effectively mitigated autocorrelation in the Phase I data, enabling the establishment of statistical control using the Max-MCUSUM chart with a k value of 0.5. Nevertheless, the Phase II analysis revealed 144 data points exceeding the control limits, signifying a process deviation from the established state of control. Comparative assessments demonstrated the superior sensitivity of the Max-MCUSUM chart to process shifts relative to conventional control charting techniques. To further enhance the proposed methodology, future research should explore the potential of advanced time series models, such as Transformers, Generative Adversarial Networks [19], or Temporal Convolutional Networks [20], to address autocorrelation challenges more comprehensively. Also, several robust methods can also be applied to improve the accuracy of monitoring process [21, 22].

## References

1. Johnson, R.: Applied Multivariate Statistical Analysis, (2002).
2. Veerendra, G.T.N., Kumaravel, B., Rao, P.K.R., Dey, S., Manoj, A.V.P.: Forecasting models for surface water quality using predictive analytics. Environ Dev Sustain. 26, 15931–15951 (2024).
3. Palani, S., Liong, S.-Y., Tkalich, P.: An ANN application for water quality forecasting. Mar Pollut Bull. 56, 1586–1597 (2008).
4. won Seo, I., Yun, S.H., Choi, S.Y.: Forecasting water quality parameters by ANN model using pre-processing technique at the downstream of Cheongpyeong Dam. Procedia Eng. 154, 1110–1115 (2016).

5. Khusna, H., Mashuri, M., Suhartono, S., Prastyo, D.D., Ahsan, M.: Multioutput least square SVR-based multivariate EWMA control chart: The performance evaluation and application. Cogent Eng. 5, 1531456 (2018).
6. Ahsan, M., Aulia, T.R.: Comparing the Performance of Several Multivariate Control Charts Based on Residual of Multioutput Least Square SVR (MLS-SVR) Model in Monitoring Water Production Process. In: Journal of Physics: Conference Series. p. 012018. IOP Publishing (2021).
7. Yusri, H.I.H., Hassan, S.L.M., Halim, I.S.A., Abdullah, N.E.: Water quality classification using SVM and XGBoost method. In: 2022 IEEE 13th Control and System Graduate Research Colloquium (ICSGRC). pp. 231–236. IEEE (2022).
8. Wang, Y., Zhou, J., Chen, K., Wang, Y., Liu, L.: Water quality prediction method based on LSTM neural network. In: 2017 12th international conference on intelligent systems and knowledge engineering (ISKE). pp. 1–5. IEEE (2017).
9. Tejoyadav, M., Nayak, R., Pati, U.C.: Multivariate water quality forecasting of river ganga using var-lstm based hybrid model. In: 2022 IEEE 19th India Council International Conference (INDICON). pp. 1–6. IEEE (2022).
10. Khusna, H., Mashuri, M., Suhartono, Prastyo, D.D., Lee, M.H., Ahsan, M.: Residual-based maximum MCUSUM control chart for joint monitoring the mean and variability of multivariate autocorrelated processes. Prod Manuf Res. 7, 364–394 (2019).
11. Khusna, H., Mashuri, M., Suhartono, S., Prastyo, D.D., Ahsan, M.: Multioutput least square SVR based multivariate EWMA control chart: The performance evaluation and application. Cogent Eng. (2018). https://doi.org/10.1080/23311916.2018.1531456.
12. Prastyo, D.D., Khusna, H., Mashuri, M., Suhartono, S., Ahsan, M.: Multivariate CUSUM control chart based on the residuals of multioutput least squares SVR for monitoring water quality. MJS. 38, 73–83 (2019).
13. Sulistiawanti, N., Ahsan, M., Khusna, H.: Multivariate Exponentially Weighted Moving Average (MEWMA) and Multivariate Exponentially Weighted Moving Variance (MEWMV) Chart Based on Residual XGBoost Regression for Monitoring Water Quality. Engineering Letters. 31, (2023).
14. Khusna, H., Mashuri, M., Suhartono, Prastyo, D.D., Lee, M.H., Ahsan, M.: Residual-based maximum MCUSUM control chart for joint monitoring the mean and variability of multivariate autocorrelated processes. Prod Manuf Res. 7, 364–394 (2019).
15. Cheng, S.W., Thaga, K.: Multivariate max-CUSUM chart. Qual Technol Quant Manag. 2, 221–235 (2005).
16. Khusna, H., Mashuri, M., Ahsan, M., Suhartono, S., Prastyo, D.D.: Bootstrap-based maximum multivariate CUSUM control chart. Qual Technol Quant Manag. 17, 52–74 (2020).
17. Montgomery, D.C.: Introduction to statistical quality control. John wiley & sons (2019).
18. Page, E.S.: Cumulative sum charts. Technometrics. 3, 1–9 (1961).

19. Qi, S., Chen, J., Chen, P., Wen, P., Niu, X., Xu, L.: An efficient GAN-based predictive framework for multivariate time series anomaly prediction in cloud data centers. J Supercomput. 80, 1268–1293 (2024).
20. Hu, Y., Lyu, L., Wang, N., Zhou, X., Fang, M.: Application of hybrid improved temporal convolution network model in time series prediction of river water quality. Sci Rep. 13, 11260 (2023).
21. Mashuri, M., Ahsan, M., Lee, M.H., Prastyo, D.D.: PCA-based Hotelling's T2 chart with fast minimum covariance determinant (FMCD) estimator and kernel density estimation (KDE) for network intrusion detection. Comput Ind Eng. 158, 107447 (2021).
22. Ahsan, M., Mashuri, M., Lee, M.H., Kuswanto, H., Prastyo, D.D.: Robust adaptive multivariate Hotelling's T2 control chart based on kernel density estimation for intrusion detection system. Expert Syst Appl. 145, 113105 (2020).

**Appendix 1**. Architecture and Hyperparameters of LSTM Model

| Layer | Hyperparameter |
|---|---|
| LSTM Layer 1 | LSTM Units: 512 |
| | Recurrent Dropout: 0.2 |
| | Kernel Regularizer: l2(0.001) |
| Batch Normalization | - |
| Dropout | Rate: 0.2 |
| | LSTM Units:256 |
| Skip Connection 1 | Recurrent Dropout: 0.2 |
| | Kernel Regularizer: l2(0.001) |
| | LSTM Units:256 |
| Skip Connection 1 Residual | Recurrent Dropout: 0.2 |
| | Kernel Regularizer: l2(0.001) |
| Skip Connection 2 | LSTM Units: 128 |
| | Recurrent Dropout: 0.2 |
| | Kernel Regularizer: l2(0.001) |
| | LSTM Units: 128 |
| Skip Connection 2 Residual | Recurrent Dropout: 0.2 |
| | Kernel Regularizer: l2(0.001) |
| | LSTM Units: 64 |
| Skip Connection 3 | Recurrent Dropout: 0.2 |
| | Kernel Regularizer: l2(0.001) |
| | LSTM Units: 64 |
| Skip Connection 3 Residual | Recurrent Dropout: 0.2 |
| | Kernel Regularizer: l2(0.001) |
| | LSTM Units: 32 |
| Skip Connection 4 | Recurrent Dropout: 0.2 |
| | Kernel Regularizer: l2(0.001) |
| | LSTM Units: 32 |
| Skip Connection 4 Residual | Recurrent Dropout: 0.2 |
| | Kernel Regularizer: l2(0.001) |
| | LSTM Units: 32 |
| Final LSTM Layer | Recurrent Dropout: 0.2 |
| | Kernel Regularizer: l2 (0.001) |
| Batch Normalization | - |
| Dropout | Rate: 0.2 |
| | LSTM Units: 256 |
| Dense Layer 1 | Activation: leaky_relu |
| Batch Normalization | - |
| | LSTM Units: 128 |
| Dense Layer 2 | Activation: leaky_relu |
| Batch Normalization | - |
| | LSTM Units: 64 |
| Dense Layer 3 | Activation: leaky_relu |
| Batch Normalization | - |

| Layer | Hyperparameter |
|---|---|
| | LSTM Units: 32 |
| Dense Layer 4 | Activation: swish |
| Batch Normalization | - |
| Final Dense Layer | Activation: linear |