Journal of Quality Measurement and Analysis *JQMA* 21(1) 2025, 253-265
e-ISSN: 2600-8602 *http://www.ukm.my/jqma*
https://doi.org/10.17576/jqma.2101.2025.16

A COMPARISON OF VARIOUS RESIDUAL-BASED CONTROL CHARTS FOR DETECTING AND MONITORING ABNORMAL RIVER WATER LEVELS

(Perbandingan Pelbagai Carta Kawalan Berasaskan Residual untuk Mengesan dan Memantau Paras Air Sungai yang Tidak Normal)

SITI NUR ATIQAH MOHD SHAFIE, NORSHAHIDA SHAADAN*,
FIRDAUS MOHAMAD HAMZAH & MAHAYAUDIN M. MANSOR

ABSTRACT

Malaysia experiences recurring floods with numerous issues that have surfaced annually. The rapid overflow of river water necessitates monitoring abnormal levels with a visualization tool as an alarm system. This study utilized a control chart approach to monitor the Klang River's water level in Taman Sri Muda for a signal of potential flood occurrence. Aiming to identify a suitable method, this study compares several residual-based control charts to achieve the purpose. Due to the autocorrelation issue that might cause false alarms, this study emphasizes the implementation of control charts based on residuals. The control chats were developed based on the residual observations obtained from the best time series model, ARIMA (1,1,1), involving the Mean Shewhart (I-Shewhart), Moving Average (MA), and Exponential Weighted Moving Average (EWMA). The results show that EWMA Residual with λ=0.05 is found to have the best performance compared to other control charts; followed by EWMA Residual with λ=0.1 and the I-Shewhart Residual with a detection score of 7.1, 3.84, and 3.56 percent respectively. The results revealed that the detected water level anomalies occurred dominantly in April, October, and November. The study emphasizes the efficiency of early signals produced by the control charts in detecting abnormal water levels by visualizing their occurrence and behaviour in the Klang River at the study location.

Keywords: anomaly; autocorrelation; control chart; floods; river water level

ABSTRAK

Malaysia mengalami banjir berulang dengan pelbagai isu yang timbul pada setiap tahun. Limpahan air sungai yang cepat memerlukan pemantauan tahap yang tidak normal dengan alat visualisasi sebagai sistem amaran. Kajian ini menggunakan pendekatan carta kawalan untuk memantau aras air Sungai Klang di Taman Sri Muda terhadap kemungkinan berlakunya banjir. Dengan matlamat untuk menghasilkan kaedah pengesanan yang sesuai, kajian ini membandingkan beberapa carta kawalan berasaskan data reja untuk mencapai tujuan ini. Disebabkan isu autokorelasi yang mungkin menyebabkan amaran palsu, kajian ini menekankan aplikasi carta kawalan yang berasaskan data reja sebagai teknik untuk membina carta kawalan kualiti. Carta kawalan dibangunkan berdasarkan data reja yang diperoleh daripada model siri masa terbaik iaitu ARIMA (1,1,1), meliputi Mean Shewhart (I-Shewhart), Moving Average (MA), dan Exponential Weighted Moving Average (EWMA). Hasil kajian menunjukkan carta kawalan reja EWMA dengan λ=0.05 didapati memberikan prestasi terbaik berbanding carta kawalan yang lain; diikuti oleh carta kawalan reja EWMA dengan λ=0.05 dan carta kawalan I-Shewhart dengan peratus skor pengenalpastian 7.1, 3.84 dan 3.56 masing-masing. Keputusan analisis mendapati anomali paras air sungai berlaku secara dominan dalam bulan April, Oktober, dan November. Kajian ini menekankan kecekapan isyarat awal yang dihasilkan oleh carta kawalan melalui visualisasi dan tingkahlaku paras air yang tidak normal di Sungai Klang di lokasi kajian.

Kata kunci: anomali; autokorelasi; carta kawalan; banjir; aras air Sungai

1. Introduction

Malaysia, a Southeast Asian country with a tropical climate and diverse geography, faces significant challenges from recurring floods, typically occurring from November to March (Northeast Monsoon) and May to September (Southwest Monsoon). However, Asmat *et al.* (2021) disclose that climate changes extend flood durations beyond these seasons. The combination of heavy rainfall, unique topographical features, and human activities contributes to the vulnerability of various regions in Malaysia to flooding. The relationship between river water levels and floods is critical for understanding and predicting flood events (Ahmed *et al.* 2022). River water levels that exceed the Upper Control Chart Limit (UCL) or below the Lower Control Limit (LCL) are relatively known as anomalies or abnormal levels (Kadri *et al.* 2016). The river water levels are typically collected every minute, with hourly and daily data treated as univariate time series variables. A study by Bjelonja (2021) proposed control charts to monitor these levels as an early flood warning. However, data collected at intervals often exhibit positive autocorrelation (Lowry & Montgomery 1995), negatively impacting the control chart's evaluation of the statistical process (Sharifuddin *et al.* 2019). Autocorrelation increases the occurrence of false alarms, where abnormal points are detected without any real irregularities (Horng Shiau & Ya-Chen 2005; Shaadan *et al.* 2015). This provides the need for an approach to mitigate autocorrelation effects for monitoring the river water level over time.

Several researchers adopted quality control charts (Mehmood *et al.* 2020; Yusuf *et al.* 2020) to monitor flood risk indexes. These charts are widely used in environmental studies to help track data over time, identify trends, and detect anomalies for environmental management purposes. Walter Shewhart's control chart, developed in the 1920s, is valuable for monitoring flood variability and issuing early warnings. The application compares historical and current data to assess whether variations fall within control limits. Generally, the control charts are categorized into two main types: the memory-less control chart and the memory type (da Cunha Alves *et al.* 2019). The Shewhart control chart is an example of a memory-less chart, which relies solely on current data and exhibits less sensitivity to small-to-moderate shifts but displays enhanced capability in detecting large shifts in independent and normally distributed processes (Gove *et al.* 2013). Conversely, memory-type control charts utilize current and past information, demonstrate greater sensitivity in detecting small to moderate shifts (Riaz *et al.* 2011), and exhibit robustness against control chart assumptions (Horng Shiau & Ya-Chen 2005). However, the control charts were less effective in identifying large shifts (Mehmood *et al.* 2020).

Numerous authors have proposed various methods to address autocorrelated data within the statistical process control environment (Grimshaw 2023; Silpakob *et al*. 2023; Sabahno & Celano 2023; Sheu *et al*. 2023). Some approaches fit the time series model to the dataset and obtain the residual to establish the control chart such as those discussed by Zaidi *et al*. (2023) and Nguyen *et al*. (2023). A residual control chart is a tool used to monitor the residuals, which are the differences between observed and predicted values by a statistical model. Concerning an autocorrelation issue, residual control charts are commonly employed after applying a time series model to account for autocorrelated errors. The residuals control chart offers a clearer view of the underlying process behavior by accounting for autocorrelated errors, thereby ensuring more accurate and reliable statistical process control (Jafarian-Namin *et al*. 2021). The application of the Individual Mean Shewhart (I-Shewhart) control chart, Moving Average (MA) control chart, Exponentially Weighted Moving Average with λ=0.05, 0.10, and 0.20 (EWMA) control chart using the time series model (Residual) are briefly explained to improve the performance of control charts in the methodology section.

Generally, this study employs autocorrelation removal with the time series model (Residual) on the three different control charts that is Individual Mean Shewhart (I-SHEWHART) control chart, the Moving Average (MA) control chart, and the Exponentially Weighted Moving Average (EWMA) control chart. Besides handling the autocorrelation problems, this study helps to determine the most appropriate control chart for monitoring the river water level to serve as an early warning system in assisting responsible bodies in reducing the impact of floods on communities and the environment when dealing with autocorrelated data.

2. Research Methodology

2.1. *Materials and Methods*

This research utilized a secondary dataset collected over five years, consisting of hourly recorded data from 2017 to 2022, sourced from the Department of Irrigation and Drainage (DID). The variable under investigation was the hourly water levels, measured in meters per hour (m/h), as outlined in Table 1.

Table 1: Data description

Variable Name	Types of Data	Unit of Measurement	Description
River Water Level	Quantitative: Continuous data	Meter per hour (m/h)	The hourly data recorded each day

The research area chosen for this study focuses on data from the telemetry station located at Taman Sri Muda along the Klang River, as depicted in Figure 1, created using ArcGIS software. Situated within the Klang district, this telemetry station monitors the main and sub-river basin of the Klang River. The details of the station utilized in this study are presented in Table 2.

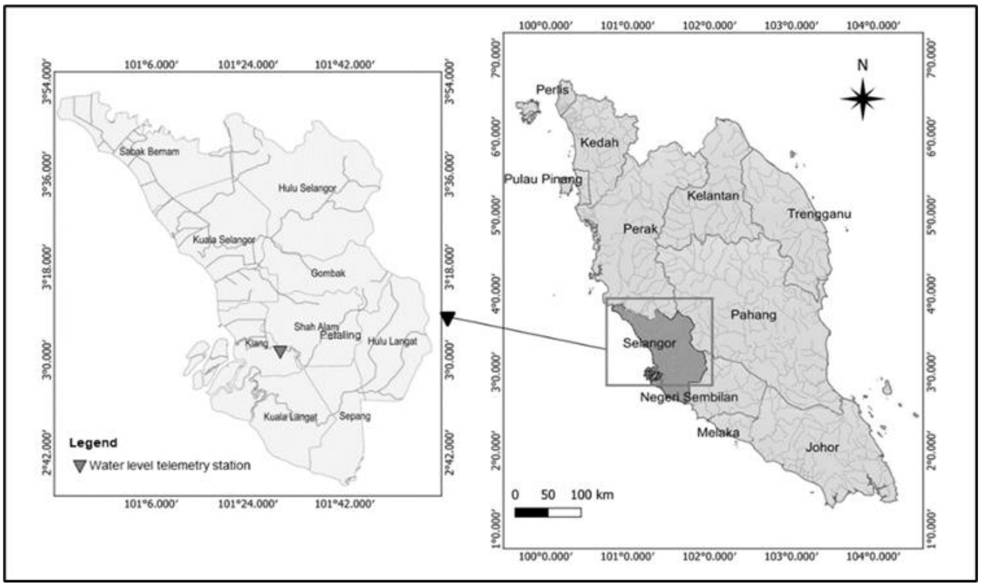

Figure 1: The telemetry stations of river water level in Selangor

Table 2: Summary of telemetry stations of river water level

Station Identification	Station Name	District	Main Basin
3015432	Klang River at Taman Sri Muda	Klang	Klang River

2.2. *Materials and methods*

The hourly river water level dataset is transformed into a new series of univariate daily mean datasets from 2017 to 2019 and divided into two parts where the dataset from the year 2017 to 2018 was used for Phase I analysis to obtain an in-control process and serve as a baseline or reference for understanding the unusual behaviour of river water levels in Phase II. The remaining data in the year 2019 is used for Phase II analysis for control charting, comparative analysis, and monitoring of the abnormal records of river water level behaviour. Table 3 describes a two-phase approach in the data analysis to achieve the research objectives. Before conducting the analysis, the missing values are treated with the mean values.

Table 3: Process flow of control chart establishment

Process Steps	Phase I: Obtaining Normal Daily Mean River Water Level Dataset
1	The mean river water level from the year 2017-2018 dataset with the estimation of the normal process means value μ and standard deviation σ were obtained by traditional outlier removal using the boxplot method. The outliers are removed from the initial daily mean river water level dataset and the normal Quantile-Quantile Plot (Q-Q plot) is performed to confirm the data normality.
2	An approach of autocorrelation treatment is considered for the river water level dataset using residuals to fulfill the *i.i.d* process of control charting. The reference time series model with the best-fit model of ARIMA (1,1,1) is built based on the normal mean river water level dataset.
3	In-control residual mean river water level dataset of ARIMA (1,1,1), n=696 are obtained in Phase 1 analysis after achieving normality assumption and *i.i.d* process.
	Phase II: Control Charting, Performance Comparison Analysis and Application
4	The testing dataset of river water level from the year 2019 was utilized in the Phase II control charts for monitoring whereby the control limits obtained from a normal process in Phase I were used in Phase II. This testing dataset typically includes anomalous data due to its association with irregular processes.
5	The new dataset using testing data (2019, n=365) based on the reference time series model based on the best-fitted model of ARIMA (1,1,1) in Phase I is obtained for autocorrelation removal treatment.
6	The control chart limits (UCL, LCL) are calculated based on the in-control process mean (μ) and standard deviation (σ) residuals of the reference time series model in Phase I.
7	The Individual Mean Shewhart residual (I-Shewhart residual) control chart, Moving Average residual (MA residual) control chart, and Exponentially Weighted Moving Average residual (EWMA residual) control chart are constructed to address autocorrelation effects using the best fit time series models of ARIMA (1,1,1). The different parameters of λ are applied in the EWMA control chart since the weighting factor, λ directly affects the results obtained (Ozilgen *et al.* 2013). A smaller value of λ leads to a smoother plot by giving older data more weight and recent data with less weight. Therefore, the parameters of λ=0.05, 0.10, and 0.20 proposed by (Montgomery 2019) are constructed in the EWMA residual control chart to achieve better results compatible with the study of river water levels.
8	The control chart performance is compared based on the results of the detection score (percent detection) reflecting the sensitivity power of the chart.

Table 3 (Continued)	
9	The capability of a control chart is identified based on the control chart's performance
10	The confirmation of abnormalities is validated in the application analysis through monitoring with the optimal control chart, which is determined after identifying the most reliable control chart

2.3. *Autocorrelation treatment using residuals*

In this study, an initiative to reduce the autocorrelation effect in control charts is considered in establishing the control charts to overcome false alarms. The false alarms referred to incorrect detected anomalies which can lead to misleading alerts. In monitoring systems or control charts applications, false alarms can lead to unnecessary actions or interventions (Magaji *et al.* 2015). The serial correlation in the data was found to be the cause of false alarms to occur (Lu & Reynolds 1999). Thus, to overcome the effect of data autocorrelation, this study employed the control chart development based on residuals.

The time series ARIMA model is obtained and the control charts will be applied to the residuals produced by the models. Since a time series model takes into account the correlated process data over time, this approach enables the creation of a robust control chart that can effectively distinguish between normal process variation and signals indicating that the process is moving out of control without false alarms (Chen & Yu 2019). The best fit ARIMA (*p, d, q*) model where *p* is denoted as an autoregressive component, *d* as a differencing component, and *q* as a moving average component were computed using the R software with the ARIMA forecast package using auto.arima function that fits the best ARIMA model to the univariate time series. Next, the Individual Mean Shewhart residual (I-Shewhart residual) control chart, Moving Average residual (MA residual) control chart, and Exponentially Weighted Moving Average residual with λ=0.05, 0.10, and 0.20 (EWMA residual) control chart were established in this study.

2.3.1. *I-Shewhart residual control chart*

The residual data is extracted from the autocorrelated data of daily mean river water level data based on the best-fit model of ARIMA (1, 1, 1). The central line (CL) that represents the mean of the residuals and control limits (UCL and LCL) are obtained to plot the residual control charts as follows:

$$CL = \bar{x}_{residual} \tag{1}$$

where $\bar{x}_{residual}$ is the mean of residuals of m samples. Nonetheless, the structure of UCL and LCL is involved with the average of moving range statistics. Therefore, the moving range values and the average of the moving range are present as $MR_i = |\, x_{i\,residual} - x_{i\,residual-1}|$ and $\overline{MR} = \frac{\sum_{i=1}^{m} MR_i}{m}$. Thus, the *UCL* and *LCL* are given as follows:

$$UCL = \bar{x}_{residual} + 3\frac{\overline{MR}}{d_2} \tag{2}$$

$$LCL = \bar{x}_{residual} - 3\frac{\overline{MR}}{d_2} \tag{3}$$

d_2 refer to the factors for the centre line.

2.3.2. *MA residual control chart*

This chart is a statistical tool used in quality control to monitor and control a process based on the residuals of a moving average model. This type of control chart is designed to detect unusual patterns or trends in the residuals, which represent the differences between observed values and the corresponding values predicted by a moving average model. The moving average (M_i) values are computed after obtaining the residual data based on the best-fitted time series model of ARIMA (1,1,1). The M_i statistics are computed based on the residual data with the span (w) at the time i where the span (w) indicates how much the moving average is used. The M_i statistics are defined by:

$$M_i = \frac{x_{i\,residual} + x_{i\,residual-1} + \cdots + x_{1\,residual-w+1}}{w} \tag{4}$$

The same span as the AR1 control chart is applied to the residual control chart which is $w = 2$. The mean and standard deviation of residual data from the residual in-control dataset of Phase I are used in control charting. The central limit of the MA residual control chart was denoted by the mean of the residual data as Eq. (1). While the *LCL* and *UCL* are mentioned as:

$$UCL = \mu_0 + \frac{3\sigma}{\sqrt{w}} \tag{5}$$

$$LCL = \mu_0 - \frac{3\sigma}{\sqrt{w}} \tag{6}$$

where the μ_0 was the mean of the in-control residual dataset or the *CL* obtained based on Eq. (1). Besides, *UCL* and *LCL* are often set based on statistical principles, such as ±3 standard deviations from the mean.

2.3.3. *EWMA residual control chart*

An Exponentially Weighted Moving Average (EWMA) Residual Control Chart is effective in detecting abnormalities with small or gradual shifts and large or abrupt shifts (Zhou & Tang 2016). This chart is useful when there is a need to give more weight to recent observations in detecting subtle changes in the process since this chart used all the information of previous data to establish the control chart.

$$Z_i = \lambda\, x_{i\,residual} + (1-\lambda) z_{i-1} \tag{7}$$

where i was the sample number and λ was a smoothing constant weighting factor. Equivalent to the EWMA AR1 control chart, the λ suggested by Montgomery (2019) is applied in the EWMA Residual control chart that is as λ=0.05, 0.10 and 0.20. The initial Z_i is denoted as Z_0 since there was z_{i-1} when $i = 1$. Therefore the Z_0 (starting value) was taken equal to the mean of the target value μ_0 or the average of residual data. Where the μ_0 is corresponding to the *CL*, *LCL* and *UCL* as follows:

$$CL = \mu_0 = \bar{x}_{residual} = \frac{\sum_{i=1}^{m} x_{residual}}{m} \tag{8}$$

$$UCL = \mu_0 + L_0 \sigma \sqrt{\frac{\lambda}{(2-\lambda)} [1-(1-\lambda)^{2i}]} \tag{9}$$

$$LCL = \mu_0 - L_0\sigma\sqrt{\frac{\lambda}{(2-\lambda)}[1-(1-\lambda)^{2i}]} \qquad (10)$$

The size of the sigma limit (L=3) is used to determine the width of the EWMA limits and σ was the process standard deviation (Montgomery 2019).

3. Research Methodology

The presence of autocorrelation is shown in each year based on the autocorrelation function (ACF) plot depicted in Figure 2. In 2017 and 2018, not all lags exceeded the significance level of 0.05, unlike in 2019 where all lags surpassed the level. However, across all ACF plots, there was a seasonal cyclical autocorrelation pattern in the river water level data. In 2017, there was a prominent peak correlation coefficient at lag 0, followed by a rapid decline in correlation until lag 4, indicating a weaker correlation below 0.4. Subsequently, it became negative at lag 6 before gradually rising again at lag 11. This suggests that there was the presence of both positive and negative autocorrelation between the lags in 2017. Additionally, the river water level data in 2019 exhibited the highest positive correlation among the lags compared to 2017 and 2018.

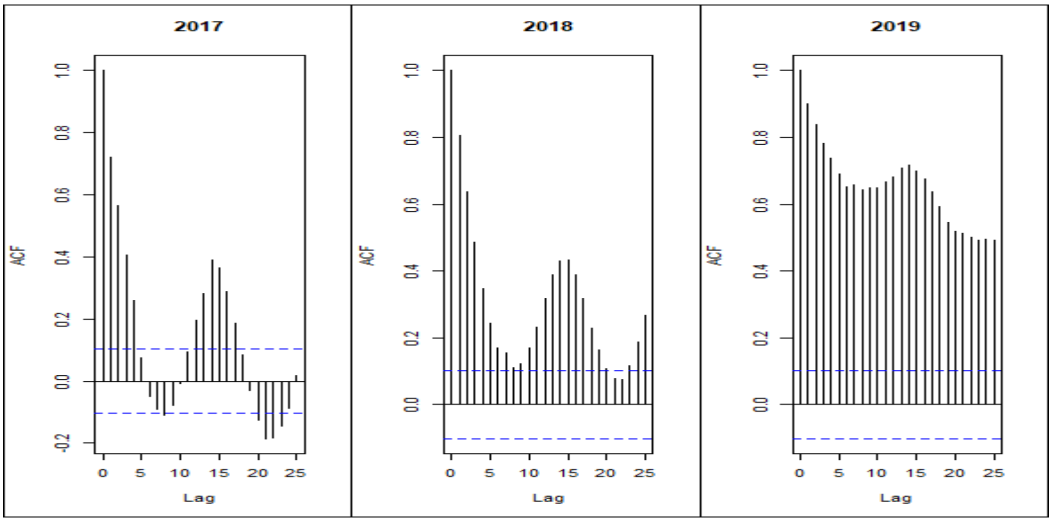

Figure 2: Autocorrelation function (ACF) plot of daily mean river water level in Klang River at Taman Sri Muda annually (period 2017-2019)

The above insights are confirmed based on the results of the Durbin-Watson test depicted in Table 4. The test was conducted based on the residual data of a regression function by treating water levels and time as the dependent and the independent variables respectively. At a 5% significance level, the small p-values (0.0001), which is less than 0.05, provide evidence that the serial correlation is significant for the series of river water levels of each year. This is shown by the p-value obtained less than the 5% significance level.

Recall that, the first two years of data (2017-2018) were utilized for Phase I control chart analysis to obtain normal process data so that the in-control control limits LCL, CL, and UCL can be established. The data from 2019 was used for control charting in Phase II, employing the control limits established in Phase I, and then determining the most suitable control chart for monitoring river water levels.

Table 4: Summary of Durbin Watson Test of mean daily river water level in Klang River at Taman Sri Muda annually (period 2017-2019)

Year	Durbin-Watson Statistics	*p*-value
2017	0.5578	0.0001
2018	0.3861	0.0001
2019	0.1924	0.0001

The boxplot method was employed in Phase I analysis to identify outliers in the dataset. The upper part of Figure 3 illustrates several outliers detected above the upper limits of the boxplot, suggesting a slight positive skewness in the data. Additionally, the Normal Q-Q plot indicates deviations from the normality assumption due to outliers, as some sample quantile points deviate from the reference quantiles of the Normal distribution line. This suggests that the daily mean river water level data from 2017 to 2018 were not normally distributed. Subsequently, outliers were identified primarily in mid-October, specifically on the 10^{th}, 11^{th}, and 12^{th}, as well as in November, with river water levels ranging from 2.25 to 2.70 m/h. These outliers were removed from the dataset giving a symmetric distribution exhibited in the lower part of Figure 3. This outcome confirms the achievement of a normal dataset in Phase I. After removing the outliers, the total number of observations decreased from n=730 to n=722. The remaining data represent the observations under stable or normal conditions so that the control limits computed based on these observations can be used for monitoring in Phase II.

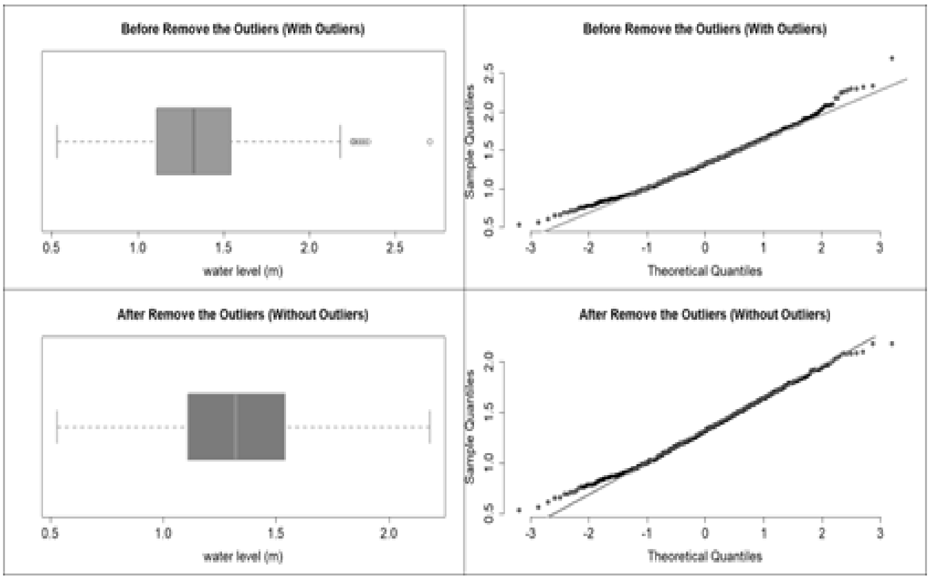

Figure 3: Boxplot plot and normal Q-Q plot of daily mean river water level in Klang River at Taman Sri Muda before and after removing outliers (period 2017-2018)

The Durbin Watson test in Table 4 confirmed that the river water level data at the Klang River telemetry stations in Taman Sri Muda exhibited autocorrelation over time. Therefore, when applying control charts to this dataset, it is crucial to address autocorrelation, as it can result in more false alarms or incorrect signals during monitoring. Next, the subsequent analysis in Phase II aims to explore how data autocorrelation affects the performance of various control charts considered in this study, specifically focusing on their sensitivity.

There were three types of control charts, namely Individual Mean Shewhart (I-Shewhart) control chart, Moving Average (MA) control chart, and Exponentially Weighted Moving Average (EWMA) control chart was deployed based on the autocorrelation removal using residual via the best-fitted ARIMA model to achieve i.i.d assumptions. Note that the EWMA control charts were implemented using three distinct values of the weighted factor, λ (0.05, 0.10, 0.20). The dataset from 2019 was utilized in all the control charts constructed in this study for the monitoring stage Phase II. The performance of the control charts was validated using a detection score, represented by the percentage of signals falling outside the control limits. The signals also represent the sensitivity power of the control chart. Under, the *i.i.d* condition, the higher the sensitivity of a control chart, the higher its performance (Lee & McGreevey 2002). The percentage of signals for each control chart is presented in Table 5.

Table 5: Summary of control limits establishment and percentage of detection (%)

Type of approaches	**Types of control charts**	**Central line (CL)**	**Upper control limit (UCL)**	**Lower control limit (LCL)**	**No. of signal**	**% Signals**
Residuals by using ARIMA time series model	I-Shewhart residual Control Chart	-0.0074	0.4951	-0.5100	13	3.56
	MA residual Control Chart	-0.0074	0.3479	-0.3627	11	3.01
	EWMA residual Control Chart (λ=0.05)	-0.0074	0.0730	-0.0879	26	7.12
	EWMA residual Control Chart (λ=0.1)	-0.0074	0.1079	-0.1227	14	3.84
	EWMA residual Control Chart (λ=0.2)	-0.0074	0.1601	-0.1750	11	3.01

The residual-based control charts in Table 5 were established based on the in-control process mean, μ (-0.0074), and standard deviation, σ (0.16752) residual of reference best fit ARIMA (1,1,1) time series model. Notably, the I-Shewhart residual chart demonstrated higher sensitivity detection percentages (3.56 %) compared to the MA residual chart (3.01 %) and MA control chart which also resulted in 3.01 % detection. However, the EWMA residual control chart proves to be the most sensitive, with the lowest frequency of false alarms that were detected at 7.12 percent of anomalies signal and the detection sensitivity reduces when λ=0.2. The results indicate that the EWMA residual control charts, λ=0.05 are deemed more appropriate for monitoring and detecting anomalies in daily mean river water levels at the study location, as they impose stricter control limits. Previous studies suggest that smaller values of λ produce reliable results by placing more weight on older data. Therefore, λ = 0.05 is recommended due to the importance of older data in determining daily water level measurements. Supported by the previously mentioned literature, the uncorrelated residuals control chart obtained from the ARIMA (1,1,1) model effectively removes the effects of autocorrelation in the monitoring process and thus, reduces false alarms, as it satisfies Normal distribution and independent and identically distributed (*i.i.d*) assumptions. The points that appeared beyond the control limits are considered anomalies. The control chart developed under this data quality consideration, with higher signals indicating more sensitive the control chart, thus reflecting how efficient it is. In addition, for the practical manufacturing field of usage for example, Western Electric rules for identifying true signals of instability rather than false alarms may help to further distinguish between random variation and genuine process shifts (Shmueli & Cohen 2003).

Figure 4 illustrates the monitoring results obtained from the EWMA residual control chart with a smoothing parameter of $\lambda = 0.05$, revealing several anomalous levels detected above the upper control limits. Anomalies exhibit a seasonal fluctuation pattern in the daily mean river water level detection. The earliest detection occurred in Aril with a significant shift, while a higher number of alerts were observed from mid-October to the end of November. This period coincides with the monsoon wind changes in Peninsular Malaysia, occurring twice a year in April to May and September to October. Consequently, flooding occurs not only during the monsoon season but also from mid-October to mid-January annually. This aligns with the dates and values of anomalies.

The control chart's limits indicate that the river water level fluctuated within certain bounds, with a maximum of 3.30 m/h and a minimum of 1.91 m/h recorded outside the normal range. The earliest anomaly, on April 7th, showed a slight difference from the minimum anomaly recorded, suggesting no flood event occurred in April. However, anomalies were detected on October 12th (2.14 m/h) and October 17th (2.12 m/h), followed by consecutive anomalies from October 18th to 20th, coinciding with a flash flood in Taman Sri Muda on October 17th to 18th. River cleaning and water pumping actions by authorities reduced anomaly levels on October 19th and 20th. November saw frequent anomalies, possibly indicating early signs of the flood event that occurred on January 29th, 2020. Notably, no anomalies were detected below the lower control limits EWMA residual control chart with $\lambda = 0.05$.

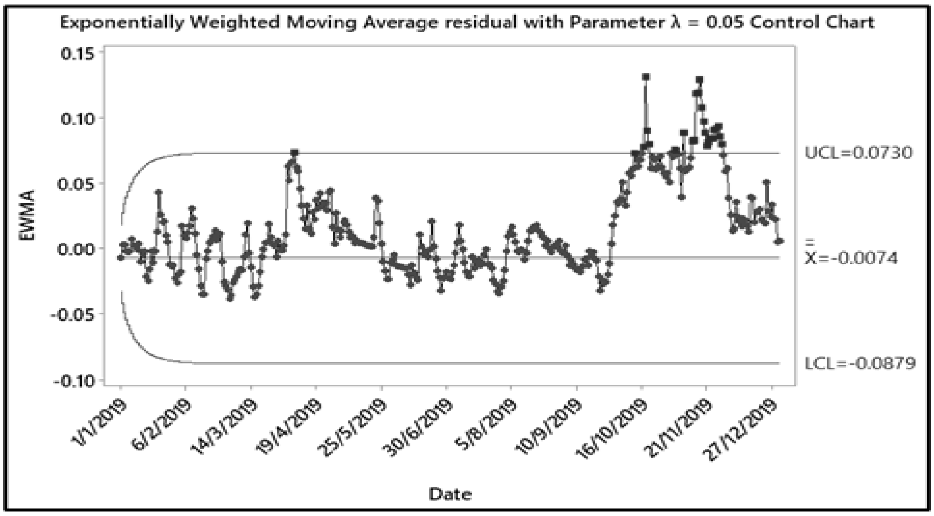

Figure 4: Exponentially Weighted Moving Average residual with parameter λ = 0.05 control chart of daily mean river water level

On the other hand, anomaly detection using DID thresholds in Figure 5 depicted no anomalies were recorded, with the process remaining within normal limits. The threshold values defined by DID indicate normal at 2.80 m/h, with alerts at 4.40 m/h, warnings at 4.70 m/h, and danger at 5.00 m/h. Despite several points lying outside the norm or usual level of

2.80 m/h on specific dates, such as October 18th, early and late November, and late December, the process remained in control as it did not exceed the alert, warning, or danger levels.

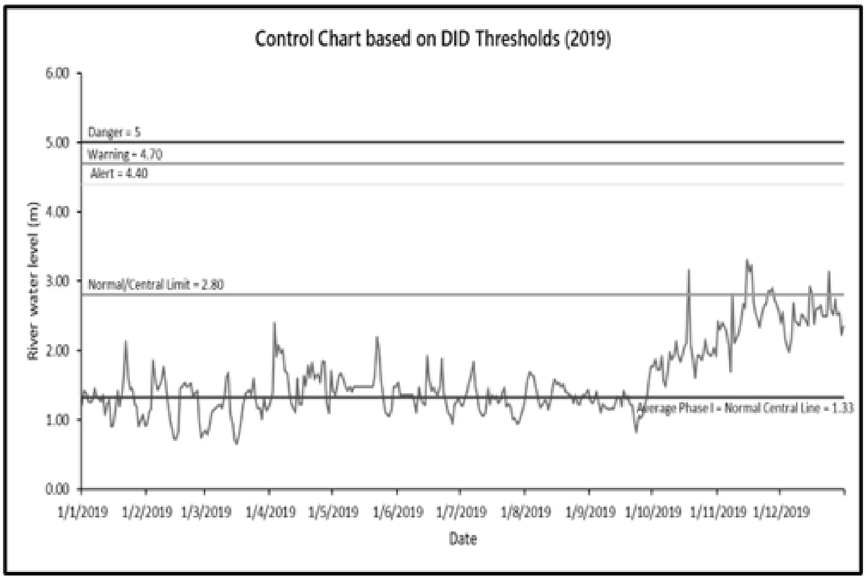

Figure 5: DID Threshold Control Chart of daily mean river water level based on Department of Irrigation and Drainage (DID) thresholds

This study revealed that the normal level, in statistical control, was 1.33 m/h, narrower than the DID's normal threshold. The number of points exceeding the statistical control standard is higher compared with the DID's normal threshold of 2.80 m/h. Anomalies detected outside the statistical control norm level, 1.33 m/h, occurred earlier on September 28, 2019, than those detected above the DID's normal level on October 18, 2019, indicating a three-week delay in detection based on the DID's threshold. Thus, utilizing the new normal central limit based on statistical control is advisable for monitoring river water levels. Nevertheless, the DID's thresholds were less stringent than control chart limits, potentially contributing to a lack of early flood warnings in 2019. Noticeably, the application of the proposed residual control chart revealed the water level anomalies earlier, indicating the importance of using a proper statistical control chart for river water level monitoring, especially the EWMA residual control chart with $\lambda = 0.05$ since the control chart removes the autocorrelation effect in the dataset.

4. Conclusion

The study proposed implementing the residual-based control chart to consider effective anomaly detection and assessment when monitoring highly autocorrelated processes in Sungai Klang River at Taman Sri Muda. Additionally, the research emphasized the importance of ensuring adherence to the assumption of normal distribution before constructing control charts for monitoring purposes. In building control charts based on the residuals approach, efforts were made to achieve the remedial assumption of independence and identically distributed (i.i.d) data. These methods were compared in terms of their ability to effectively detect anomalies, with the EWMA residual control chart with λ=0.05 emerging as the best approach in detecting anomalies and offering a smoother monitoring process due to its consideration of older data. Hence, the appropriate selection of λ is crucial in enhancing sensitivity. The detection of anomalies using this approach revealed significant occurrences typically towards the end of the year, mostly occurring between October and November. This study emphasizes the importance of employing rigorous control charting methods for early and accurate anomaly detection, particularly in monitoring river water levels subject to seasonal and extreme events like floods. In summary, the residual control chart method effectively addressed the autocorrelation issue. This was demonstrated by the improvement in control chart performance, based on the performance of the sensitivity detection of control charts. Supported by previous studies, the employment of residual data in control chart development helps to mitigate river water floods via an alarm system.

References

Ahmed A.N., Yafouz A., Birima A.H., Kisi O., Huang Y.F., Sherif M., Sefelnasr A. & El-Shafie A. 2022. Water level prediction using various machine learning algorithms: a case study of Durian Tunggal river, Malaysia. *Engineering Applications of Computational Fluid Mechanics* **16**(1): 422-440.

Asmat A., Wahid S.N.S. & Deni S.M. 2021. Identifying rainfall patterns using Fourier series: A case of daily rainfall data in Sarawak, Malaysia. *Journal of Physics: Conference Series* **1988**(1): 012086.

Bjelonja A. 2021. Comparative analysis of process control charts in order to define most efficient sampling frequency. Master Thesis. Lappeenranta–Lahti University of Technology.

Chen S. & Yu J. 2019. Deep recurrent neural network-based residual control chart for autocorrelated processes. *Quality and Reliability Engineering International* **35**(8): 2687-2708.

da Cunha Alves C., Konrath A.C., Henning E., Walter O.M.F.C., Paladini E.P., Oliveira T.A. & Oliveira A. 2019. The Mixed CUSUM-EWMA (MCE) control chart as a new alternative in the monitoring of a manufacturing process. *Brazilian Journal of Operations & Production Management* **16**(1): 1-13.

Gove A.D., Sadler R., Matsuki M., Archibald R., Pearse S. & Garkaklis M. 2013. Control charts for improved decisions in environmental management: a case study of catchment water supply in south-west Western Australia. *Ecological Management & Restoration* **14**(2): 127-134.

Grimshaw S.D. 2023. Constructing control charts for autocorrelated data using an exhaustive systematic samples pooled variance estimator. *Journal of Quality Technology* **55**(3): 349-365.

Horng Shiau J.J. & Ya-Chen H. 2005. Robustness of the EWMA control chart to non-normality for autocorrelated processes. *Quality Technology & Quantitative Management* **2**(2): 125-146.

Jafarian-Namin S., Fallahnezhad M.S., Tavakkoli-Moghaddam R., Salmasnia A. & Fatemi Ghomi S.M.T. 2021. An integrated quality, maintenance and production model based on the delayed monitoring under the ARMA control chart. *Journal of Statistical Computation and Simulation* **91**(13): 2645-2669.

Kadri F., Harrou F., Chaabane S., Sun Y. & Tahon C. 2016. Seasonal ARMA-based SPC charts for anomaly detection: Application to emergency department systems. *Neurocomputing* **173**: 2102-2114.

Lee K. & McGreevey C. 2002. Using control charts to assess performance measurement data. *The Joint Commission Journal on Quality Improvement* **28**(2): 90-101.

Lowry C.A., & Montgomery D.C. 1995. A review of multivariate control charts. *IIE Transactions* **27**(6): 800-810.

Lu C.W. & Reynolds Jr M.R. 1999. EWMA control charts for monitoring the mean of autocorrelated processes. *Journal of Quality Technology* **31**(2): 166-188.

Magaji A.A., Yahaya A. & Asiribo O.E. 2015. Assessing the effects of autocorrelation on the performance of statistical process control charts. *International Journal of mathematics and statistics invention* **3**(6): 15-23.

Mehmood R., Lee M.H., Baqar M. & Riaz M. 2020. Application of quality control charts for early detection of flood hazards. *Polish Journal of Environmental Studies* **29**(6): 4241-4248.

Montgomery D.C. 2019. *Introduction to Statistical Quality Control*. 8th Ed. Hoboken, NJ: John Wiley & Sons, Inc.

Nguyen H.D., Nadi A.A., Tran K.D., Castagliola P., Celano G. & Tran K.P. 2023. The Shewhart-type RZ control chart for monitoring the ratio of autocorrelated variables. *International Journal of Production Research* **61**(20): 6746-6771.

Ozilgen S., Bucak S. & Ozilgen M. 2013. Improvement of the safety of the red pepper spice with FMEA and post processing EWMA quality control charts. *Journal of Food Science and Technology* **50**(3): 466-476.

Riaz M., Abbas N. & Does R.J. 2011. Improving the performance of CUSUM charts. *Quality and Reliability Engineering International* **27**(4): 415-424.

Sabahno H. & Celano G. 2023. Monitoring the multivariate coefficient of variation in presence of autocorrelation with variable parameters control charts. *Quality Technology & Quantitative Management* **20**(2): 184-210.

Shaadan N., Jemain A.A., Latif M.T. & Deni S.M. 2015. Anomaly detection and assessment of PM10 functional data at several locations in the Klang Valley, Malaysia. *Atmospheric Pollution Research* **6**(2): 365-375.

Sharifuddin A.N., Ali H. & Rahman A.A. 2019. Exponential Weighted Moving Average (EWMA) control chart for air pollution index at SK Batu Muda, WP Kuala Lumpur. *Test Engineering and Management* **81**: 3944-3953.

Sheu W.T., Lu S.H. & Hsu Y.L. 2023. The generally weighted moving average control chart for monitoring the

process mean of autocorrelated observations. *Annals of Operations Research* 1-29.
Shmueli, G. & Cohen, A. 2003. Run-length distribution for control charts with runs and scan rules. *Communication in Statistics-Theory and Methods* **32**(2): 475-495.
Silpakob K., Areepong Y., Sukparungsee S. & Sunthornwat R. 2023. A new modified EWMA control chart for monitoring processes involving autocorrelated data. *Intelligent Automation and Soft Computing* **36**(1): 281-298.
Yusuf N.M., Saudi A.S., Shafii N. & Roespinoedji D. 2020. Flood risk pattern recognition in Kinabatangan River, Sabah. *PalArch's Journal of Archaeology of Egypt/Egyptology* **17**(10): 4018-4025.
Zaidi F.S., Dai H.L., Imran M. & Tran K.P. 2023. Monitoring autocorrelated compositional data vectors using an enhanced residuals hotelling T2 control chart. *Computers & Industrial Engineering* **181**: 109280.
Zhou Z.G. & Tang P. 2016. Improving time series anomaly detection based on exponentially weighted moving average (EWMA) of season-trend model residuals. *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 3414-3417.

College of Computing, Informatics and Mathematics
Universiti Teknologi MARA, Kampus Mukah
96400 UiTM Mukah,
Sarawak MALAYSIA
E-mail: sitinuratiqahms@uitm.edu.my

College of Computing, Informatics and Mathematics
Universiti Teknologi MARA
40450 UiTM Shah Alam
Selangor, MALAYSIA
E-mail: shahida@tmsk.uitm.edu.my[*]

Centre for Defence Foundation Studies
Universiti Pertahanan Nasional Malaysia
57000 UPNM Kem Sungai Besi,
Wilayah Persekutuan Kuala Lumpur,MALAYSIA
E-mail: firdaus.hamzah@upnm.edu.my

College of Computing, Informatics and Mathematics
Universiti Teknologi MARA
40450 UiTM Shah Alam
Selangor, MALAYSIA
E-mail: maha@uitm.edu.my

Received: 8 May 2024
Accepted: 9 January 2025

[*]Corresponding author