Flood Identification and Mitigation:
Flood risk trend by using PCA and SPC Analysis

Jurina Jaafar[1], Noor Sa'adah Abdul Hamid[1], Zulhafizal Othman[1], Nur Qudus[2]

[1] School of Civil Engineering, College of Engineering, Universiti Teknologi MARA, Malaysia
[2] Civil Engineering Department, Universitas Negeri Semarang, Semarang, Indonesia

jurina@uitm.edu.my, saadahhamid.pg@gmail.com, zulhafizal445@uitm.edu.my, nurqudus@mail.unnes.ac.id
Tel: +6019-380 0877

Abstract

PCA and SPC were used to identify data variables that trigger flooding and assess flood risk in the Muda River basin. Three (3) hydrologic variables, RL, WL, and SF, are analyzed using correlation tests, PCA and SPC. The Pearson correlation test shows SF and WL have significant correlations. The PCA indicates that all hydrologic variables are significant. The SPC shows ideal flood control values for the Muda River basin. The runoff value exceeding the UCL increases flood risk. The rapid expansion of development and anthropogenic activities have caused heavy rainfall and hydrologic variables to increase above normal levels.

Keywords: Correlation test; Principle component analysis; Statistical process control keyword

DOI: https://doi.org/10.21834/aje-bs.v8i24.422

1.0 Introduction

A flood study is an ongoing concern in the Sustainable Development Goals (SDGs) studies in Yan Kedah, Malaysia. The Sustainable Development Goals (SDGs) are a set of goals that aim to create a more sustainable future. One of the SDGs is to make cities and human settlements inclusive, safe, resilient, and sustainable. However, one of the challenges is the risk of flooding. Flooding is a common environmental hazard that causes loss of life, property, infrastructure, and financial resources. The number of people at risk from devastating floods will continue to increase due to massive urbanization and population growth in flood-prone areas. The risk of flash floods in the Muda River basin in northern Malaysia usually occurs during the northeast and southwest monsoons. The rainy season from November to February is known as North-East Monsoon. The relatively drier weather with minimal monthly rainfall in Peninsular Malaysia experiences the Indian Ocean's Southwest Monsoon from May to August. During the monsoon transition seasons from September through November, most of the state receives average monthly rainfall during the changeover between monsoons (Julien et al., 2010). More frequent rainfall events due to climate change occurred on August 18, 2021, resulting in six human life losses. Follow by. a second deluge on September 28, 2021, was no less terrifying, though no fatality rate was recorded. The third event on October 20, 2021, the flash floods in Yan considerably caused damage to infrastructure and property (Berita Harian, 2021).

The irregular distribution of severe rainfall in the area caused the river to overflow and crash with the tide, causing Yan's August 18, September 28, and October 20 flash floods, according to KASA (Bernama, 2021). The August 18 rainstorm accounted for 60% of the month's average rainfall (Berita Harian, 2021). Conditions in the Yan River basin and rough topography lead air condensation to climb faster and cool off more quickly, forming orographic clouds that form rainfall. Climate change and human activity may cause more frequent heavy rains. The Ministry of Energy and Natural Resources (KeTSA) stated that there had been no logging in the Gunung Jerai Forest Reserve for 30 years after the region was gazetted as a Protection Forest (Geopark Jerai) in 2018. Chan Ngai Weng, a water expert at Universiti Sains Malaysia, claimed that floods occur due to climate change and human activities, including forest clearing, construction, agriculture, and converting virgin forests into developed areas (The Vibes, 2021).

Flood disaster resilience and sustainable development are the main focus in the environment that aims to the societal lives healthy and safe. For this reason, three hydrological parameters at two hydrological stations near Muda River are studied to determine the cause of floods in Kedah. The Chemometric Analysis is used in this study as it is a straightforward approach and gives significant findings. This technique classifies which elements that affect the other variables in the study. The Principal Component Analysis (PCA) correlation test determines the significant hydrological variable contributing to the flood occurrence. Statistical Process Control (SPC) is used to analyze the movement of flood risk patterns for each hydrological parameter as an early assessment for flood warning systems. The PCA and SPC are used to analyze and monitor the control limit for each PCA-derived parameter. Due to a prominent component causing flooding in Kedah, the rainfall data and primary forest loss areas are studied to determine the main reason for the event.

2.0 Literature Review

2.1 Factors Contributing to Increasing Flood Risk

Most flash floods are caused by short-lived, severe thunderstorms (usually 6 hours or less). Rainfall can cause flash floods. A rainfall distribution is needed to study flooding. Floods are caused by heavy rainfall, river overflow, and the South China Sea surges (DID, 2010). When it rains, rivers that store and release water affect their watersheds. Rain causes flooding. When rain falls upstream, it is more challenging to reach the flood stage (Kamarudin et al., 2015). Malaysia is humid, where land-sea interactions are very harsh.

Rapid urbanization, conflicting land use, and population growth cause calamities. Flooding in Batman, Turkey, has reached catastrophic levels due to human effects on the environment, causing loss of life and property (SUKAR & TOBUL, 2011). Unzoned public institutions, infrastructure, and residential, commercial, and agricultural operations caused floods (Zaidee et al., 2018a). According to Berita Harian (2021), a nearby river's rapid expansion and flow ruined many resorts and waterfalls. Insufficient flow to the estuary and ocean hinders drainage and rivers. Sediment and debris block rivers and floodplains, reducing drainage capacity. This means rainwater will not be absorbed and will overflow into rivers and drains. Climate change could increase rainfall. Global warming will increase flood risk because water covers 70% of the planet. Human activity influences river basin ecosystems and drainage, whether directly or indirectly.

2.2 Effects of Flooding

Flooding damages lives, property, and infrastructure. According to Talbot et al. (2018), flooding disrupts public services too. Direct and indirect flood damages exist (Bubeck et al., 2017). Direct impacts are more accessible to predict than indirect ones. Direct losses result from floodwater's physical contact with people and damageable property. Indirect losses may be as or more significant than direct ones. Effects on urban human settlements vs rural agricultural fields differ. Flooding on farmland can ruin crops (Talbot et al., 2018). Floods can damage the ecosystem, especially if pollution sources are nearby. Adding or releasing contaminants into the atmosphere causes most pollution. Human-caused pollution can harm water bodies (Zeleáková et al., 2016). Farmland fertilizers can leach into groundwater as an unintended side effect. There are other adverse effects, such as reduced fish production due to pollution and habitat loss.

Floods also preserve ecosystem processes and biodiversity in many natural systems. They also recharge groundwater, fill wetlands, connect aquatic habitats, and convey sediment and nutrients to the ocean. Floods help species breed, migrate, and disperse. Natural systems may resist even devastating floods. Floods boost fish production, recharge groundwater, and preserve recreational areas (Svetlana et al., 2015).

2.3 Hydrological Parameters Approach

Some aspects of flood control, irrigation, drainage, and dams rely on rainfall, particularly in the early planning and building stages. At various time scales, rainfall is the primary source of flood discharge and plays a part in the flood-receding process. If there is much rain, the daily streamflow (and water level) may drop quickly (Cheng et al., 2021). DID has set up a nationwide network of rain gauges to track the amount and distribution of rain. Rainfall is a significant factor in the development of irrigation water resources. Rainfall is an important factor in urban drainage and flood control planning. According to Reddy & Reddy (2013), increased rainfall from a typhoon or front may result in catastrophic floods. Urban drainage data on short-term rainfall intensity and area distribution are crucial for large river basin flood simulations.

The water level is one of the most often monitored quantities because it is vital in many applications. A real-time river flood warning system is needed in the Philippines, where flooding is regular (Collenn et al., 2020). The water level is a simple water factor. If a river's water level rises abruptly, panic and tragedy may ensue. According to Monjardin et al. (2021), a system that monitors river levels and predicts how they will rise after rainfall reduces evacuation and flood damage. River or stream gauge height shows the water level. Due to human activity like illegal logging, soil silt and suspended particles dissolve and enter waterways during rainy days. Because of shallower rivers and faster stream flow, rising water levels cause destructive floods (Abd Halim et al., 2018).

It is vital to determine the stream's stage and discharge. Discharge is water flowing down a river or stream. Temperature, precipitation, and elevation affect the river and stream flow (Whitfield et al., 2003). According to Araujo (2017), streamflow fluctuations due to climate change have been the focus of many studies. If daily rainfall increases, too much surface runoff will be created since daily channel flow will likewise increase (Araujo, 2017). Streamflow monitoring can help manage climate-friendly water (Dobriyal et al., 2017). Changing streamflow due to rainfall and temperature must be communicated for computing semantics (Mughal et al., 2021).

2.4 Flood Monitoring Warning System (FMWS) as A Mitigation Measure

Disaster management officials employ several systems to monitor floodwater levels. Flooding is not the primary goal; it is to warn the public. It improves survival and reduces electronic damage. An early warning or real-time alerting system like this FMWS can help individuals know about the flood level that threatens their lives (Zakaria et al., 2021). Using moving range charts for each hydrological parameter, Statistical Process Control (SPC) can detect the likelihood of flooding.

Volunteers can be trained as first responders in a crisis. Public address systems, such as sirens, must be set up in places of worship and shopping malls as a precaution because only some people are tech-savvy enough to get flood alerts online or on their cell phones. Security forces must be stationed in flooded areas to protect citizens' homes if they must evacuate. Flood victims want temporary housing for evacuees. Flood warning systems, such as local government monitoring control charts in SPC analysis, can be made more effective, cost-efficient, and easy. SPC charts helped identify a flood risk pattern (Abd Halim et al., 2018). Visible peak surge Upper Control Limit (UCL), Control Limit (CL), Average Value (AVG), and Lower Control Limit (LCL) ensured no river parameter exceeded maximum capacity. When the maximum limit control is about to be reached, this system can warn local communities to prepare for floods, save important documents, and evacuate.

3.0 Methodology

This study is primarily quantitative, relying on the three years of recent hydrological data collected (2019-2021) from the Department of Irrigation and Drainage (DID). There are three (3) hydrological parameters, which are Rainfall (RL), Water Level (WL), and Stream Flow (SF), to be analyzed through the application of Principle Component Analysis (PCA) and Statistical Process Control (SPC). PCA and SPC tests were conducted through XLSTAT Base 2021 add-in software and Excel. In pre-processed data of PCA, Pearson's Correlation test is conducted to identify variables with a strong correlation between variables and then to determine the variables contributing to flooding based on load factors. Two types of hypothesis testing were employed to determine whether or not variables suitable for PCA factoring are suitable: Bartlett's Sphericity Test and Kaiser-Meyer-Olkin (KMO). Through SPC analysis, the control charts for each hydrological parameter represent the movement of flood risk pattern. The factors contributing to the flooding issue in Kedah could be justified through the comparison of statistics on primary forest loss areas in Yan, Kedah, obtained from the Global Forest Watch (GFW) Organization website (2019-2021) and the rainfall data obtained from DID (2019-2021).

3.1 Study Area

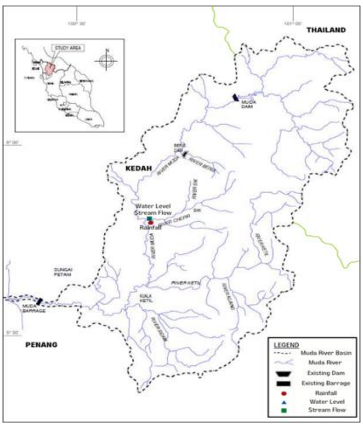

Figure 1: The Topography of the Study Area at Muda River Basin
(Source: National Water Balance System (NAWABS))

The geology of the study area is at the coordinates of 5˚06'N and 100˚17'E, with a catchment area of 4,210 km2 and a length of 180 km. The three significant flash flood events occurred in 2021 at Yan, Gunung Jerai and the southern highland areas encountered as the areas that receive heavy yearly rainfall above 5,000 millimetres. Figure 1 and Table 1 show the topography of the study area at the Muda River basin and the locations of monitoring stations at the Muda River basin. Secondary hydrological data, including Rainfall (RF), Stream Flow (SF) and Water Level (WL), were provided by the Department of Drainage and Irrigation (DID) from 2019 to 2021.

Table 1: The Locations of Monitoring Stations at Muda River Basin

Station No.	Latitude	Longitude	Name of Station	Parameters
Site 5806066	5˚48’50N	100˚37’55E	Jeniang Klinik	Rainfall
Site 5806414	5˚49’10N	100˚37’55E	Sg. Muda di Jeniang	Stream Flow
Site 5806414	5˚49’10N	100˚37’55E	Sg. Muda di Jeniang	Water Level

(Source: Department of Irrigation and Drainage (DID))

3.2 Statistical Analysis

This study used statistical Analysis to determine the relationship between hydrologic parameters, RL, SF, and WL. Correlation tests were used to determine the fit of the variables. Principal component analysis (PCA), statistical process control (SPC) was used to determine the strongest loading factor significantly influencing flooding. In contrast, SPC determines the movement of the flood risk pattern per the control boundary.

3.2.1 Correlation Test

This study employed the Correlation test to find variables with a vital link for further Analysis. The test measures two variables with a relationship of -1 to 1. Pearson and Spearman coefficients can be utilized in this investigation, but Pearson was used consistently (Nor et al., 2018). This study was used to determine the strongest Correlation between hydrological data parameters and the strongest Correlation. A positive Correlation means both variables are increasing linearly, while a negative Correlation shows one variable increasing and the other declining. Pearson rank coefficient requires actual data and ratio-scaled variables. The correlation test was run using XLSTAT Base 2021 add-in software, and the following equation (1) was used:

$$r_p = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \qquad (1)$$

Where,
r_p = Pearson Correlation Coefficient,
X_i = Values of the x-variable in a Sample,
$\bar{X}$ = Mean of the Values of the x-variable,
Y_i = Values of the y-variable in a Sample,
$\bar{Y}$ = Mean of the Values of the y-variable

3.2.1.1 Bartlett's Sphericity Test

Bartlett's Test of Sphericity compares a correlation matrix to the identity matrix. This test is typically carried out before we apply data reduction approaches such as Principal Component Analysis (PCA) or Factor Analysis to check that a data reduction technique can compress the data meaningfully (Academy, 2009). According to (Ayuni & Sari, 2018), the statistic test can be computed using this formula (2):

$$X^2 = -\left(s - 1 - \frac{2p+5}{6}\right) \ln|R| \tag{2}$$

Where,
s = Number of Samples
p = Number of Variables
R = Correlation Matrix of Variables

3.2.1.2 Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy

The Kaiser-Meyer-Olkin (KMO) statistic, or called the Measure of Sampling Adequacy (MSA), is a statistic that examines whether other factors may explain correlations between variables in the data set. KMO and MSA threshold values are reported in Table 2 by statistician Kaiser (Al-Hashmi et al., 2014), who developed the metric. As a result, the KMO statistic should be used as the primary criterion for determining whether or not the data are suitable for Principal Components Analysis (PCA). If these measurements suggest that the variables are sufficiently associated, we can proceed with further investigation. KMO must have a value of at least 0.5 to be considered a viable factor in factor analysis. Otherwise, it should be omitted from consideration. The formula in (3) for calculating KMO is as follows:

$$KMO = \frac{\sum\sum r_{IJ}^2}{\sum\sum r_{IJ}^2 + \sum\sum a_{IJ}^2} \tag{3}$$

Where,
r_{IJ} = simple Correlation between i-th and j-th variable
a_{IJ} = partial Correlation between i-th and j-th variable

Table 2: Threshold Values for KMO and MSA

KMO/MSA value	Adequacy of the correlations
Below 0.50	Unacceptable
0.50 – 0.59	Miserable
0.60 – 0.69	Mediocre
0.70 – 0.79	Middling
0.90 and higher	Marvellous

(Source: An index of factorial simplicity." Psychometrika 39, no. 1 (1974): 31-36. (Kaiser, 1974))

3.2.2 Principle Component Analysis (PCA)

According to Nor et al. (2018), Principal Component Analysis can find significant variables in a small number of data sets (PCA). This method reduces the number of variables, examines the structure or relationship between variables in hydrological data, and finds non-dimensionality in the theoretical construct. This study has multi-co-linearity, which means two or more variables are re-correlated. The PCA will find the most vital link between variables. This method visualizes the most and least essential metrics without losing much data (Zaidee et al., 2018b). Based on the data, it may be possible to identify which variables have the most significant impact on hydrological models in the Muda River basin. The following is the equation that was utilized in (4):

$$Z_{ij} = ai^1xj^1 + ai^2xj^2 + ai^3xj^3 + aimx \tag{4}$$

Where,
Z = Component Score,
a = Component Loading,
i = Component Number,
m = Total Variables,
x = Measured of Variables

3.2.3 Statistical Process Control (SPC)

This method calculated the limits of all flood-causing variables in the Muda River basin. UCL, CL, or AVG, and LCL were shown on the control charts. A control chart's Sigma value falls within a particular data range. In the event of a variable, the UCL value is undesirable and high risk for floods (Saudi et al., 2018). This Analysis helped anticipate future hydrological modelling. Control charts can identify trends and patterns by showing how accurately data deviates from past baselines, indicating the best baselining and dynamic threshold, and capturing abnormal resource utilization. The following equations (5) and (6) were employed in this Analysis:

$$\text{Moving Range} = \text{Plot: MRt for } t = 2, 3, \ldots, n \tag{5}$$

Where,
MR = Average Moving Range,
t = time,
n = Individual ValuesZ = Component Score,

$$\text{Average Value: } \widetilde{X} = \frac{\sum_{i=1}^{m} x_i}{n} \tag{6}$$

Where,
$\tilde{X}$ = Moving Range,
n = Individual Values,
x_i = Difference Between Data Points

4.0 Results and Discussions

4.1 Analysis of Correlation Test

P-values were used to describe the rates of Correlation between the examined variables. In these results, the p-values for the Correlation between RF and SF, between RF and WL, and between SF and WL are different from 0 with a significance level alpha equal to 0.05, indicating a significant correlation between these variables. This finding is based on the correlation coefficient size and interpretation in Table 4. Table 3, reveals an insignificant correlation between RF and WL at 0.162, followed by SF at 0.091. WL and SF have a significant positive correlation of 0.899. Overall correlation test shows that WL and SF are highly correlated. According to prior research, a correlation coefficient of 0.7 or above indicates a substantial correlation and should be chosen for further Analysis (Nor et al., 2018).

Table 3: Threshold Values for KMO and MSA

Variables	Rainfall (RL)	Stream Flow (SF)	Water Level (WL)
Rainfall (RL)	**1**	**0.091**	**0.162**
Stream Flow (SF)	**0.091**	**1**	**0.899**
Water Level (WL)	**0.162**	**0.899**	**1**

Values in bold are different from 0 with a significance level of alpha=0.05

Table 4: Threshold Values for KMO and MSA

Size of Correlation	Interpretation
.90 to 1.00 (-.90 to -1.00)	Very high positive (negative) Correlation
.70 to .90 (-.70 to -.90)	High positive (negative) Correlation
.50 to .70 (-.50 to -.70)	Moderate positive (negative) Correlation
.30 to .50 (-.30 to -.50)	Low positive (negative) Correlation
.00 to .30 (.00 to -.30)	Negligible correction

Source: Statistic Corner: A guide to the appropriate use of Correlation coefficient in medical research (Campi et al., 1997)

4.1.1 Bartlett's Sphericity Test

According to the Table 5 below, the correlation matrix is not an identity matrix (the null hypothesis has been rejected), as shown by a significant statistical test where the p-value (<0.0001) is less than the set significance level, alpha (0.050). The fact that the observed Chi-square value for Bartlett's test of sphericity was 1671.178 would suggest that the factor model is appropriate (P-value less than 0.0001).

Table 5: Bartlett's Sphericity Test

Chi-square (Observed value)	1671.178
Chi-square (Critical value)	7.815
DF	3
p-value (Two-tailed)	<0.0001
alpha	0.050

4.1.2 Measure of Sample Adequacy of Kaiser-Meyer-Olkin

From Table 6, the value of KMO was discovered to be 0.495 or rounded to the nearest value of 0.5, which is an acceptable number and indicates that the correlation matrix can be used for factoring in PCA. According to Table 4,"Threshold Values for KMO and MSA", a KMO value of 0.5, is considered miserable for the adequacy of the correlations. This is because there are relatively significant partial correlations compared to the sum of correlations. In other words, broad correlations could cause a problem for Principle Component Analysis (PCA).

Table 6: Kaiser-Meyer-Olkin Measure of Sampling Adequacy

Rainfall (RL)	0.412
Stream Flow (SF)	0.497
Water Level (WL)	0.497
KMO	0.495 ≈ 0.5

4.2 Results from Principle Component Analysis (PCA)

This study applied PCA to RF, WL, and SF to establish the most discriminating parameter and the likely primary source of flooding in the Muda River Basin. Only the most vital analytical factor was used to create PCs (Zaidee et al., 2018b). Table 7 states that a factor's eigenvalue must be larger than 1.0 to be considered significant. Its high factor makes it a crucial analytical factor. Table 7 shows that F1, F2, and F3 each represent a Principle Component (PC). It "regroups" data into smaller components. This regrouping is based on RF, WL, and SF data. The initial variables' information is squeezed into the new variables (major components). This ordering of components maintains maximal diversity with a small number of newly created components. Instead of variables, investigate data with components. The most variable factor will be examined using each variable's data to determine the essential variable contributing to floods.

Table 7 presents the eigenvalue after the initial PCA. The findings show that F1 (1.933) were gained as the eigenvalue, more than (>1.0), and that the variability percentage of this factor was 64.437%. Both F2 and F3 resulted in eigenvalue of less than (<1.0) with 0.968 and 0.099, respectively. The proportion of both variances F2 and F3 are cumulatively less than the F1 at 32.275% and 3.288% consecutively. By regrouping the variables, the results show that the F1 component has the most variation. The F2 component has the second most variation, while the F3 component should have the least fluctuation.

Table 7: Eigenvalues for Factors from PCA

	F1	F2	F3
Eigenvalue	1.933	0.968	0.099
Variability (%)	64.437	32.275	3.288

Cumulative %	64.437	96.712	100.000

Eigenvalues > 1.0 consider a selected factor
(Source: Flood Risk Pattern Recognition In Rajang River Basin, Sarawak (Syafiqah et al., 2020a))

Table 8, shown factor loading result of WL with the highest value of 0.970 followed by SF with a value of 0.960. RF shown the largest impact of factor loading with a value of 0.965.

Table 8: Correlations between Factors and Variables in PCA from 2019-2021 in Muda River Basin

	F1	F2	F3
Rainfall (RL)	0.263	0.965	-0.018
Stream Flow (SF)	0.960	-0.171	-0.220
Water Level (WL)	0.970	-0.092	0.223
Eigenvalue	1.933	0.968	0.099
Variability (%)	64.437	32.275	3.288
Cumulative %	64.437	96.712	100.000

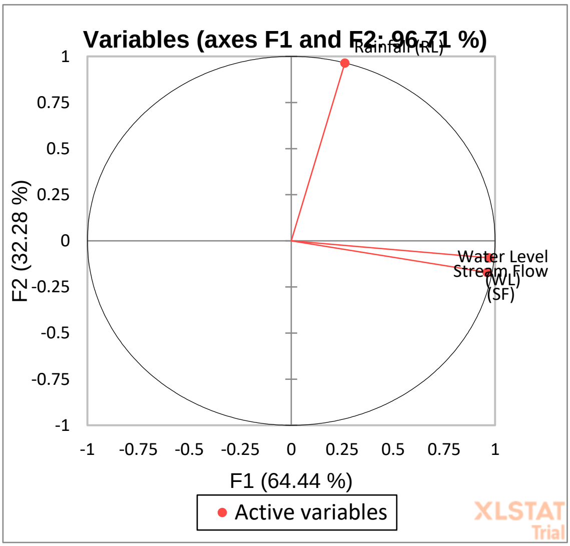

Figure 2: Correlations Circle between Factors and Variables in PCA

Figure 2 shown the original 3 variables of Rainfall (RF), Water Level (WL) and Stream Flow (SF) are shown in red on this two-dimensional factor space within this two-dimensional circle. A strong correlation exists when two red lines are pointing in the same direction (r close to 1), an orthogonal relationship exists when the lines are 90 degrees apart where no correlation occur (r close to 0), and a negative relationship exists when the lines are pointing in the opposite way or far from the center (r close to -1). In this correlation circle, it can be seen that the WL and SF have strong correlation between these two variables while RF have no correlation with any of the variables involved.

Table 9: Factor Loadings after Varimax Rotation

	VF1	VF2
Rainfall (RL)	0.063	0.998
Stream Flow (SF)	0.975	0.026
Water Level (WL)	0.969	0.105
Eigenvalue	1.933	0.968
Variability (%)	63.130	33.582
Cumulative %	63.130	96.712

Values in bold correspond for each variable to the factor for which the squared cosine is the largest
A cutoff value > 0.95 consider as a threshold for the above loading factor

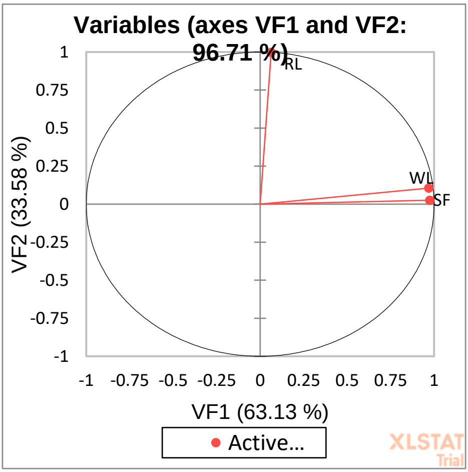

Figure 3: Factor Loadings Plot after Varimax Rotation

According to Zaidee et al. (2018), only significant factor loadings were used to select the principal component (PC). Varimax factor (VF) was generated using the essential component in rotation (Syafiqah et al., 2020a). Factor loading findings must have an eigenvalue larger than 1.0 to be considered significant. Cross-loadings make early extraction factors challenging to interpret. Orthogonal factor rotation simplifies the study's results. Varimax is a column-based method for maximizing pattern structure coefficient squared differences, which can rotate orthogonally. Shrestha (2021) states rotating PCs yield varimax factors (VFs). Rotation enhances extraction-induced high and low loads.

Table 9 and Figure 3 show a significant factor loading for SF and WL with values of 0.975 and 0.969, respectively. RF (0.714) exhibited a favorable moderate factor loading in VF2 with an eigenvalue of 0.968 and a variance of 33.582 %. Syafiqah et al. (2020b) excluded PCA with eigenvalues less than (1.0). Final PCA results after Varimax rotation demonstrate that SF contributes to flooding.

4.3 Results from Statistical Process Control (SPC)

The flood control warning system uses SF as its primary parameter. The correlation coefficient between the SF and other variables, such as WL and RF, shows that SF is the highest and most feasible for use in the flood control warning system. The variable is the most dynamic and should be used as the primary parameter in the flood warning system. Table 9 shows the results of Statistical Process Control (SPC) for all variables at Muda River Basin.

Table 9: Results of Statistical Process Control (SPC) for Muda River Basin

Area	Points Plotted	Lower Control Limit (LCL)	Average Limit (AVG)	Upper Control Limit (UCL)	Limit Sigma	Sample Size
Stream Flow	Individual	-58.698 m^3/s	19.567 m^3/s	97.831 m^3/s	Moving Range	1
Water Level	Individual	19.202 m	20.998 m	22.793 m	Moving Range	1
Rainfall	Individual	-33.031 mm	6.416 mm	45.863 mm	Moving Range	1

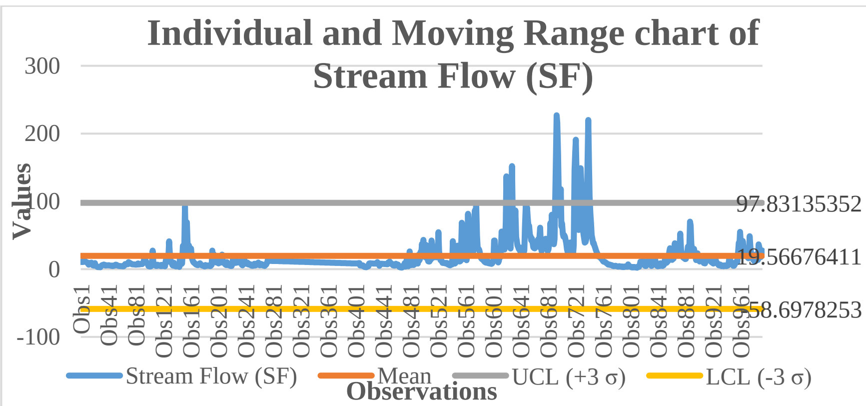

Figure 4: SPC Result for Muda River Basin's Stream Flow

In Figure 4, SF is an essential component of floods, and data beyond the UCL are at significant risk of flooding that might devastate the stream bank. All values below the LCL indicate drought or minimal flood risk. These findings are essential for enhancing Muda River Basin's flood early warning system. SF peaks at 227 m3/s in November 2020. Heavy rain in this month led the WL in low-lying areas or near the river to rise one to two metres (NST, 2020). Kedah's rainy season lasts from April to November. The bulk of the spikes that occurred over the years showed a high value of SF in the study region, which was caused by changes in the stream's characteristics and runoff from precipitation along the stream (Zaidee et al., 2018b).

Figure 5 shows that the WL peaked in November 2020 at 24.13 metres. The higher peak value showed that human activities had impacted water storage since the stream had been widened, raising the WL. Both WL level and SF peak simultaneously due to their significant correlation. Increasing WL increases Muda River SF. SF is often calculated from WL data using a rating curve (Fentaw et al., 2019). SFs and WLs are utilized to produce the rating curve, which involves a huge number of measurements over a lengthy period (concurrent SF and WL data sample is termed gauging).

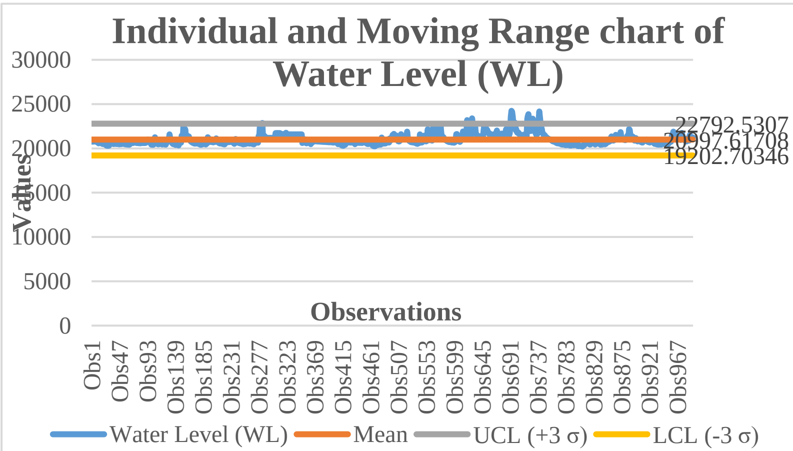

Figure 5: SPC Result for Muda River Basin's Water Level

Figure 6 shown similar rainfall trends, with a peak of 103.500 mm in September 2020. It was considered an anomaly, not a usual occurrence. The Muda River basin's monthly, yearly, and monsoon rainfall patterns varied greatly. March, April, May, June, July, August, September, October, November, and December are the rainiest months. The Muda River's catchment basin floods frequently around April to May and September to November, causing near-annual flooding (Ghani et al., 2010). In 30 years, Kedah's central region, comprising Kota Setar, Pendang, Yan, Kuala Muda, and Sik, is expected to receive more rain (2010 to 2039). The average amount of rain will likely stay the same

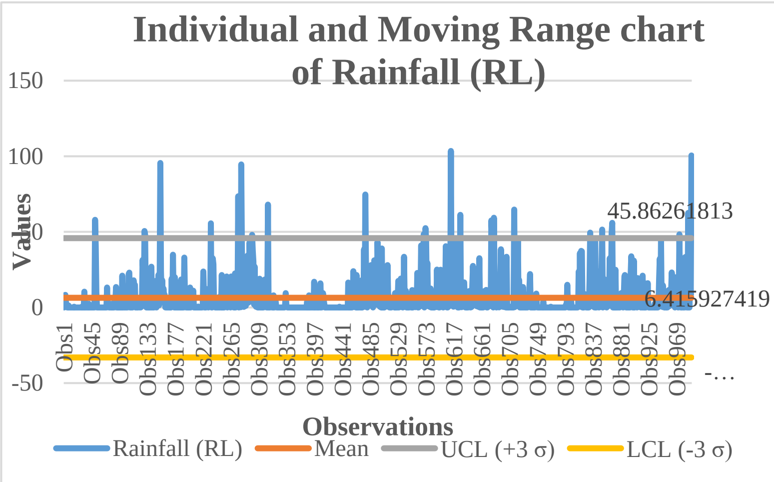

Figure 6: SPC Result for Muda River Basin's Rainfall

4.4 Comparison of Rainfall and Primary Forest Loss Data

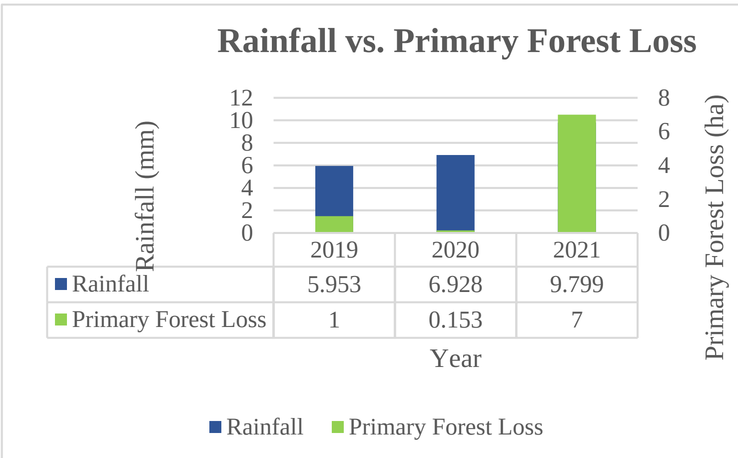

Figure 7: Correlation Between Year (2019-2021) with Rainfall and Primary Forest Loss Data

According to Grosso (2021), all forest regions preserve and contribute the most humidity to the atmosphere. This mechanism increases rainfall as air flows over the forest. Rainfall levels reduce due to deforestation. From Figure 7, the cause of floods can be justified by comparing rainfall data and primary forest loss areas data. Flooding may be natural when rainfall data is higher than in primary forest loss regions. Otherwise, forest loss indicates human-caused flooding. Deforestation via agricultural expansion, wood exploitation (e.g., logging or wood extraction for house fuel or charcoal), and infrastructural expansion (road building and urbanization) are human activities.

In 2019, primary forest loss was 1 hectare, and yearly rainfall was 5.953 mm. In 2020, fewer human activities caused 0.153 hectares of primary forest loss, while rainfall readings increased from 6.928 mm to 9.799 mm. Primary Forest loss areas were 7 hectares in 2021. This shows active deforestation this year (2021), causing rainfall levels to be lower than the recorded primary forest loss regions.

5.0 Conclusion

Three hydrological parameters, RF, WL, and SF are used in flood risk models. To ensure the collected variables can be used for Principle Component Analysis (PCA), a Correlation test is conducted using Bartlett's Sphericity test and Kaiser-Meyer-Olkin (KMO) test. In both hypotheses tests, the p-value for Bartlett's test is below 0.05 (0.0001), deemed significant, while the KMO value is roughly 0.5, an acceptable result for PCA factoring. Therefore, this study has proved the capacity to employ RF, WL, and SF data in flood risk models. In PCA, the first procedure is to analyze the factors' eigenvalues, in which the factor had to be more than 1.0 (>1.0). The result of F1 is chosen since it shows the highest variance (64.43%) compared to F2. In the second step, the Correlation between factors and variables is analyzed, where in F1, SF (0.960) and WL (0.970) are regarded as substantial loading due to coefficients being more than 0.7. Varimax rotation has simplified the study's results. The result of Varimax showed that SF contributed mainly to the occurrence of floods in the Muda River basin beside RF and WL. For the Basin River in Malaysia, the statistically proven variables can be utilized to highlight flood patterns and the suitable rates for maximum flood management (Saudi et al., 2018). This study has shown that the most critical variables identified are valid to determine the flood alert warning system for risk models.

The result from the SPC analysis can improve the Muda River Basin's flood warning system. The SPC analysis showed that most variables exceeded the Upper Control Limit (UCL), which determines the risk factor pattern of the chosen variable. The control limit developed as a result of this study allows the early response actions to be implemented based on the river basin SF levels. Early warning systems can be made more robust and precise based on the present pattern formation in floods. The economic sector can benefit from a control limit system in preparation for flood and drought events. Rice farming is an important source of revenue and food security in the Muda River Basin; thus, the water supply must be sufficient to meet the demand for a sustainable environment. Study through integrating Chemometric Technique through Factor Analysis, Factorial Method and Time

Series Analysis, the primary cause of a flood is determined. The limitation system can measure the effectiveness of installing a warning system. SPC charts helped detect the flood risk trend. The UCL, LCL, and AVG ensured no parameters exceeded the river's maximum tolerance.

The comparison of Rainfall and Primary Forest Loss Areas from 2019 to 2021 indicates that in 2021 there is active deforestation in Yan, clearing 7 hectares. This action causes the lower result of RF levels than in primary forest loss areas. The flooding crisis occurred thrice in 2021, causing substantial property damage. News reported abnormally heavy rainfall causes an increase in the runoff, thus causing rivers to overflow and floods to occur. The main ultimate reason is also due to deforestation in this area. Deforestation in tropical rainforests could impact the planet's climate and biodiversity (Rhett A. Butler, 2019). The event occurs during the rainy season, and the loss of forest cover increases runoff into streams, rising river levels and triggering flooding downstream in the study area. Loss of trees with anchoring roots worsens erosion in the tropics. Heavy tropical rains carry forest runoff into streams and rivers. For the well-being of societies to be achieved, it is crucial to protect the environment from the natural disaster. Conserving forests is a solution to environmental degradation.

Acknowledgement
This study was financially funded by a grant from the Universiti Teknologi Mara (UiTM), Malaysia (Grant No: 600-RMC/LESTARI SDG-T 5/3 (120/2019)).

Article Contribution to Related Field of Study
State how this article has contributed to the related field of study

Authors Declaration
This article is an extended version of the original conference paper published in the E-BPJ, Vol (8), Mar, 2023, 39-49.

References

Abd Halim, M., Shakir Mohd Saudi, A., Khairul Amri Kamarudin, M., Mahmud, M., Bala Krishnan, A., & Nizam Mohd Isa, K. (2018). Assessment on Regional Flood Risk Trend in Northern Region of Malaysia: Case Study in Muda River Basin, Kedah. International Journal of Engineering & Technology, 7(4.34), 75. https://doi.org/10.14419/ijet.v7i4.34.23584

Araujo, 2010. (2017). Инновационные подходы к обеспечению качества в здравоохраненииNo Title. Вестник Росздравнадзора, 6(May), 5–9.

Bernama. (2021, September 20). Tiada Pembalakan di Hutan Simpan gunung jerai [METROTV]. Harian Metro. Retrieved January 20, 2022, from https://www.hmetro.com.my/mutakhir/2021/09/757004/tiada-pembalakan-di-hutan-simpan-gunung-jerai-metrotv

Bubeck, P., Otto, A., & Weichselgartner, J. (2017). Societal Impacts of Flood Hazards. Oxford Research Encyclopedia of Natural Hazard Science, March 2018. https://doi.org/10.1093/acrefore/9780199389407.013.281

C E F Monjardin, K M Transfiguracion, J P J Mangunay, K M Paguia, F. A. A. U. and F. J. T. (2021). Determination of River Water Level Triggering Flood in Manghinao River in Bauan , Batangas , Philippines. 18(3), 181–192.

Cheng, Y., Sang, Y., Wang, Z., Guo, Y., & Tang, Y. (2021). Effects of Rainfall and Underlying Surface on Flood Recession—The Upper Huaihe River Basin Case. International Journal of Disaster Risk Science, 12(1), 111–120. https://doi.org/10.1007/s13753-020-00310-w

Collenn, C., Carl, A. A., Kathleen, H., Gwen, S. T., Paul, V., & Clare, M. V. G. (2020). River water lever sensor as river flood warning system. International Journal of Physical Sciences, 15(4), 138–150. https://doi.org/10.5897/ijps2020.4866

Dobriyal, P., Badola, R., Tuboi, C., & Hussain, S. A. (2017). A review of methods for monitoring streamflow for sustainable water resource management. Applied Water Science, 7(6), 2617–2628. https://doi.org/10.1007/s13201-016-0488-y

Fentaw, F., Melesse, A. M., Hailu, D., & Nigussie, A. (2019). Precipitation and streamflow variability in Tekeze River basin, Ethiopia. Extreme Hydrology and Climate Variability: Monitoring, Modelling, Adaptation and Mitigation, Exodus 7, 103–121. https://doi.org/10.1016/B978-0-12-815998-9.00010-5

Julien, P. Y., Ghani, A. A., Zakaria, N. A., Abdullah, R., & Chang, C. K. (2010). Case Study: Flood Mitigation of the Muda River, Malaysia. Journal of Hydraulic Engineering, 136(4), 251–261. https://doi.org/10.1061/(asce)hy.1943-7900.0000163

Ghani, A. A., Ali, R., Zakaria, N. A., Hasan, Z. A., Chang, C. K., & Ahamad, M. S. S. (2010). A temporal change study of the Muda River system over 22 years. International Journal of River Basin Management, 8(1), 25–37. https://doi.org/10.1080/15715121003715040

Grosso, M. (2021). MAPPING THE EFFECT OF DEFORESTATION ON RAINFALL A CASE STUDY FROM THE STATE OF. October.

'Gunung jerai floods not an act of god but man-made'. The Vibes. (2021, August 20). Retrieved January 20, 2022, from https://www.thevibes.com/articles/news/38872/gunung-jerai-floods-not-an-act-of-god-but-man-made

Mughal, M. H., Shaikh, Z. A., Khand, Z. H., Rajput, A., & Akhtar, F. (2021). A Systematic Review of Ontology-Based River Streamflow and Flood Data Management Challenges. Quaid-e-Awam University Research Journal of Engineering Science & Technology, 19(1), 40–50. https://doi.org/10.52584/qrj.1901.06

Nor, P., Kandari, A., Shakir, A., Saudi, M., Chyang, J., Khairul, M., Kamarudin, A., Hafiz, M., Saad, M., Azid, A., Surayya, N., & Mahmud, M. (2018). Flood Risk Pattern Recognition Analysis in Klang River Basin. International Journal of Engineering & Technology, 7, 86–90.

Reddy, P. R., & Reddy, D. V. (2013). Hydrology and water resources engineering. International Journal of Earth Sciences and Engineering, 6(5).

Saudi, A. S. M., Kamarudin, M. K. A., Ridzuan, I. S. D., Ishak, R., Azid, A., & Rizman, Z. I. (2018). Flood risk index pattern assessment: case study in Langat River Basin. Journal of Fundamental and Applied Sciences, 9(2S), 12. https://doi.org/10.4314/jfas.v9i2s.2

SUKAR, M., & TOBUL, S. (2011). Effect of Human Factors on Occurrence of Flood and Torrent Events In Batman, Southeastern Turkey. Batman'Da (GüneydoğuTürki̇ye) Sel Ve Taşkınların OluşumundaBeşeri̇Faktörleri̇Etki̇si̇., 10(37), 267–292.

Svetlana, D., Radovan, D., & Ján, D. (2015). The Economic Impact of Floods and their Importance in Different Regions of the World with Emphasis on Europe. Procedia Economics and Finance, 34(15), 649–655. https://doi.org/10.1016/s2212-5671(15)01681-0

Talbot, C. J., Bennett, E. M., Cassell, K., Hanes, D. M., Minor, E. C., Paerl, H., Raymond, P. A., Vargas, R., Vidon, P. G., Wollheim, W., & Xenopoulos, M. A. (2018). The impact of flooding on aquatic ecosystem services. Biogeochemistry, 141(3), 439–461. https://doi.org/10.1007/s10533-018-0449-7

Whitfield, P. H., Wang, J. Y., & Cannon, A. J. (2003). Modelling Future Streamflow Extremes - Floods and Low Flows in Georgia Basin, British Columbia. Canadian Water Resources Journal, 28(4), 633–656. https://doi.org/10.4296/cwrj2804633

Zaidee, W. N. A. W. F., Saudi, A. S. M., Kamarudin, M. K. A., Toriman, M. E., Juahir, H., Abu, I. F., Shafii, M. M. N. Z., Nizam, K., & Elfithri, R. (2018a). Flood risk pattern recognition using chemometric techniques approach in Golok River, Kelantan. International Journal of Engineering and Technology(UAE), 7(3.14 Special Issue 14), 75–79. https://doi.org/10.14419/ijet.v7i3.14.16865

Zakaria, M. I., Jabbar, W. A., Technology, E. E., & Computing, I. (2021). F LOOD M ONITORING AND W ARNING S YSTEMS : A B RIEF R EVIEW 洪水監測和預警系統 ： 簡要回顧. June.

Zeleňáková, M., Blišťan, P., Alkhalaf, I., Gaňová, L., & Zvijáková, L. (2016). Assessment of environmental damages in case of flood in Bodva River basin, Slovakia. International Journal of Safety and Security Engineering, 6(3), 498–507. https://doi.org/10.2495/SAFE-V6-N3-498-507