Analysis of Trend and Change Point Detection at Kajang Station, Malaysia

Firdaus Mohamad Hamzah[1], Hazrina Tajudin[1], Othman Jaafar[1], Sharifah Mastura Syed Abdullah[2], Ekhwan Toriman[2]

[1]*Smart and Sustainable Township Research Centre (SUTRA), Universiti Kebangsaan Malaysia, UKM Bangi, 43600 Selangor, Malaysia*
[2]*Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia, UKM Bangi, 43600 Selangor, Malaysia*

ABSTRACT
Land use and land cover change gives significant impact on changes in surface runoff, flood frequency, baseflow and also average yearly discharge. This study attempts to analyze trend by using non-parametric tests namely Mann-Kendall test and Sequential Mann-Kendall test. The first is used to determine the presence of monotonic trend whilst the latter is used to detect any possible fluctuations of trend in the data series. Location which is considered for analysis is at Kajang Station, Malaysia with recorded data between 1978 to 2017. Based on the result, there are presence of increasing monotonic trend every month except during September and November, based on the calculated p-value. The sequential Mann-Kendall also shows that there is break point or change point during 2004. It is important to study on present trend and pattern of hydrological parameters since it can be used to assist in water resources planning, design and management.

Keywords: *Monthly streamflow; Kajang Station; Mann-Kendall test; Sequential Mann-Kendall.*

I. INTRODUCTION

Climate change is reported to vary significantly for these past hundreds of years. Increase in global temperature is reported by IPCC to increase around 0.74°C. This is caused by several factors such as change in precipitation pattern, sea level rise and many other. Thus, it is important to monitor change in hydrological parameters regularly. Trend analysis and change point detection had been conducted previously by many researchers. There are several parametric and nonparametric statistical methods which can be applied for analysis including Mann-Kendall test, Mann-Whitney test. These are some examples of non-parametric methods which can be implemented in data series regardless of the probability distribution its fit in. It is important to study on present trend and pattern of hydrological parameters since it can be used to assist in water resources planning, design and management [1]. Trend analysis is important to manage the impact on several hydrological phenomena like drought and floods.

Land use and land cover change gives significant impact on changes in surface runoff, flood frequency, baseflow and also average yearly discharge. Other than that, change in streamflow is also caused by other climatic causes like snowmelt, difference in discharge timing, adjusted springs highest flow and increase in summer droughts. In addition, wet climate also influence change in streamflow trend. In Malaysia, most hydrological parameters are reflected by the amount of downpour. Heavy or continuous rainfall could trigger flooding event, which in the same time alter changes in streamflow trend. Previous study reported increment in annual streamflow by 4% in Johor is caused by land use change and climate change. However, study by [2] shows that there is no significant trend which explains possible changes in climate in Malaysia.

Basically, the evaluation of Mann-Kendall test in terms of hypothesis testing in statistics concept implies that the data series is considered to be independent and identically distributed (iid) when the null hypothesis is accepted [3]. This also shows that there is no trend present in the iid data series. There are two types of error involves in hypothesis testing namely type I error and type II error. The first refers to rejection of null hypothesis when it is actually true while the latter happens when accepting null hypothesis when it is actually false. Type I error tend to occur when there exists serial dependence in the data series. It actually negates the fact that there is no trend present. This is due to variance of Mann-Kendall test is inflated by serial correlation[4]. Hence, it is important to determine the existence of serial correlation in the data series in order to reduce risk of type I error [5].

In general, hydrological data usually exhibit positive serial correlation among the data series [6,10]. Most hydrological data tend to show presence of correlation in the time series data. statistically, in order to determine presence of trend, there are several tests which can be conducted like Durbin-Watson test. The elimination of serial correlation in the hydrological data is crucial to reduce error in estimation of trend including overestimation or underestimation.

The aim for this study is to conduct trend analysis of monthly streamflow at Kajang station between 1978 to 2017 by using Mann-Kendall. Then, Sequential Mann-Kendall test is used to detect for any change point present throughout the study period at Kajang station

II. STUDY AREA

The study area focuses on Langat River at Selangor. It basically has a catchment area of approximately 2350 km2. The latitude is within 20 40'M 152"N to 30 16'M 15" and longitudes of 1010 19'M 20"E to 1020 1'M 10"E. The main river course of Langat River is about 141 km and it is mostly located around 40 km east of Kuala Lumpur. The principal river of Langat River is the Semenyih River, Lui River and Beranang River. There are four streamflow stations along Sungai Langat which are located at Pekan Dengkil, Kajang, Sungai Semenyih and Sungai Lui. The study area focuses around Sungai Langat-Kajang sub basin. The area covers approximately 389.4 km2. The flow measurement at Sungai Langat-Kajang has latitude of 020 59' 34" and longitude of 1010 47' 13".

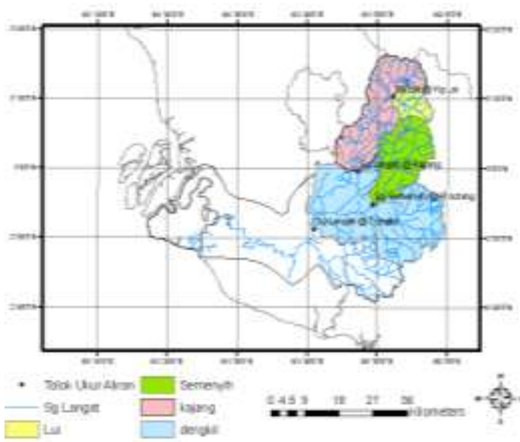

Figure 1: Sungai Langat map

III. METHODOLOGY

This study considers monthly streamflow data at Kajang station between 1978 to 2017. The station ID is 2917401 with coordinate of 02^0 59' 34" and 101^0 47' 13" respectively. The adata used is approximately around 480 data, measured in cubic meter per second (m^3/s). The data is obtained from Department of Irrigation and Drainage (DID) Malaysia. There are no missing data reported for analysis.

Mann Kendall is a non-parametric test which is used to detect trend in time series. The data is not necessarily need to be normal distributed. Mann-Kendal trend test is used to evaluate presence of monotonic trend. Upward movement refers that there is consistent increase through time, but not necessarily linear. MK test is a non-parametric (distribution-free) test. This is usually used in environmental data, climate data and hydrological data. These are test hypothesis of Mann-Kendall test:

H_o: Data comes from independent and identically distributed population

H_A: Data follows monotonic trend

There are few assumptions of using MK test which are listed as follow:

1. When no trend present in the data series, the data is independent and identically distributed. There is no correlation between data series over time.
2. Observations which are obtained over time represents true conditions at sampling stations.
3. Methods of collection, handling and measurement provides unbiased and representative observations of the population.

Independence require large scale data to ensure that there is no correlation between measurements at all time.

It is important to determine presence of correlation among the data series since the existence might affect trend detection even if there is none which actually exist. Autocorrelation function (ACF) and partial correlation function (PACF). Partial correlation between two variables refers to amount of correlation which is not explained by mutual correlation with a specified set of variables. Test statistic for Mann-Kendall is computed using the following equation:

$$S = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} sgn(x_j - x_i)$$

where x_1, x_2, x_3, ..., x_n represents the observation data and x_j is data point and time j, and θ is the difference between two consecutive data value written as $x_j - x_i = \theta$.

$$sgn(\theta) = \begin{cases} 1 \; if \; (\theta) > 0 \\ 0 \; if \; (\theta) = 0 \\ -1 \; if \; (\theta) < 0 \end{cases}$$

The data should be independent and identically distributed (iid). S approximately follows normal distribution when $n \geq 8$ ith mean and statistics are given as follow:

$E(S) = 0$

$$Var(S) = \frac{n(n-1)(2n+5) - \sum_{i=1}^{m} t_i(t_i - 1)(2t_i + 5)}{18}$$

where m is the number of groups with tied ranks and t_i represents the tied observations.

The standardized test statistics Z is calculated using:

$$Z = \begin{cases} \frac{S-1}{\sigma} \; if \; S > 0 \\ 0 \; if \; S = 0 \\ \frac{S+1}{\sigma} \; if \; S < 0 \end{cases}$$

Based on the Z-value, H_0 is rejected when $|Z||Z| \geq Z_{1-\frac{\alpha}{2}}$ at α, significance level. Positive Z value indicate existence of increasing trend whereas negative Z value indicate negative trend in the time series. Basically at $\alpha = 0.05$, $|Z| \geq 1.96$ indicates presence of significant trend.

Sequential Mann-Kendall is conducted to determine trend fluctuation and existence of changing point in the data series (Bari, Rahman, Hoque, & Hussain, 2016) . This test is first proposed by Sneyer in 1990. It determines a beginning of a significant trend, where these two point cross each other (Zarenistanak, Dhorde, & Kripalani, 2014). It involves two series, represented as progressive, $U(t_i)$ and backward series, $U'(t_i)$. The process in conducting sequential Mann-Kendall is first ranked as y_i. Magnitude of y_i is compared with y_j where $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, i-1$. For every comparison, cases where $y_i > y_j$ will be considered and represented as n_i. Test statistics for conducting sequential Mann-Kendall can be mathematically written as follow:

$$t_i = \sum n_i$$

Large sample size is normally distributed and the mean and variance can be written as:

$$E(t_i) = \frac{N(N-1)}{4}$$

$$Var(t_i) = \frac{N(N-1)(2N+5)}{72}$$

Sequential test statistics is denoted as follow:

$$U(t_i) = \frac{t_i - E(t_i)}{\sqrt{Var(t_i)}}$$

Similarly, the computation of $U'(t_i)$ is using the same formula, however, it starts from the end of the series. The intersection between $U(t_i)$ and $U'(t_i)$ represents the time when trend starts.

IV. RESULT AND DISCUSSION

Hydrological data such as streamflow, rainfall is always right skewed, with zero value as lower bound in the data structure. Data recorded for this analysis is approximately 40 years between 1978 to 2017 at Kajang station.

Minimum Flow (m^3/s)	0.9393
Maximum Flow (m^3/s)	65.3871
Mean Flow (m^3/s)	10.1195
Standard Deviation	8.05
Skewness	2.93
Kurtosis	12.63

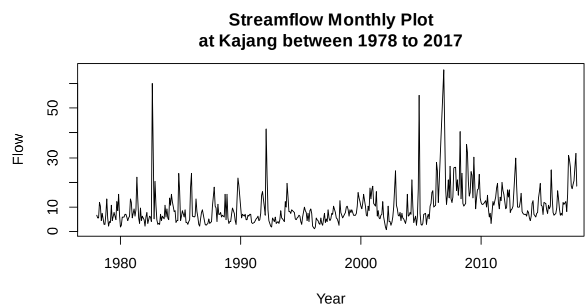

Figure 2: Monthly streamflow plot at Kajang between 1978 to 2017

The following figure represents monthly streamflow plot at Kajang station between 1978 to 2017. There are approximately 480 data involves in the analysis from January to December. Based on the result in the earlier section, the minimum and maximum data which are being considered in this study are 0.9393 and 65.3871 m^3/s respectively. Computation of descriptive data shows skewness and kurtosis of 2.93 and 12.63 reprectively. Skewness measures assymetry degree in the data series. More value greater than the given mean, 10.12 m^3/s indicates more data on the right part of the distribution, leads to tail strecthing to the right.

Basically, normal distribution has skewness value of zero. Skewness value of greater than 2 shows that the data is highly skewed, thus the monthly streamflow data series at Kajang station is highly skewed to the right. On the other hand, kurtosis measures whether the distribution is sharply peaked or not, compared to normal distribution. High kurtosis value indicates that the data series are highly peaked and have longer tails. In statistical terms, this condition is known as leptokurtic. The data series with 12.63 kurtosis value shows that the data have high peaks and longer tail.

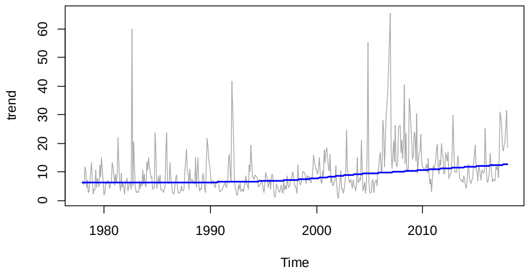

Figure 3: Linear model plot at Kajang station

Roughly, this plot shows that there is upward trend in the monthly streamflow data set. Before conducting Mann-Kendall trend test, the data series should be checked for any presence of serial correlation. If the data is serially correlated, "bootstrapping" should be conducted to eliminate the correlation effect in the time series.

Table 1:

Month	r	p-value	Indication
January	0.395	< 0.05	Increasing monotonic trend
February	0.223	0.044	Increasing monotonic trend
March	0.290	0.009	Increasing monotonic trend
April	0.387	< 0.05	Increasing monotonic trend
May	0.337	0.002	Increasing monotonic trend
June	0.312	0.005	Increasing monotonic trend
July	0.336	0.002	Increasing monotonic trend
August	0.476	0.011	Increasing monotonic trend
Septembe r	0.133	0.230	No trend
October	0.323	0.003	Increasing monotonic trend
November	0.200	0.071	No trend
December	0.321	0.004	Increasing monotonic trend

Based on the overall result calculated using Mann-Kendall test, most of the month in a year excluding September and November shows presence of monotonic trend.

Table 2:

Time	Progressive	Retrograde	Presence of Change Point
July 2004	1.4990	1.6007	True
November 2004	1.7937	1.7683	True
December 2004	1.6485	1.9569	True
January 2007	3.3503	3.2232	True

Based on the analysis of change point using sequential mann-kendall, there is change point detected during 2004 at kajang station throughout 1978 to 2017 period. This is due to change in land use in most area near kajang station.

V. CONCLUSION

It is important to study on present trend and pattern of hydrological parameters since it can be used to assist in water resources planning, design and management. Further trend analysis needs to be updated from time to time. In order to remove presence of serial correlation in the data series, another non-parametric method should be considered for future study

ACKNOWLEDGEMENT

The support provided by the Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia communities in this study is highly appreciated. The authors are very impressed with the government grant FRGS/1/2016/STG06/UKM/02/1 for funding this study.

REFERENCES

[1] Tosunoglu F, Kisi O. "*Trend analysis of maximum hydrologi drought variables using Mann-kendall and Sen's Innovative trend method*". River Res Appl. 2016; 1–14. doi:10.1002/rra.3106

[2] Ramachandra Rao A, Azli M, Pae LJ. "*Identification of trends in Malaysian monthly runoff under the scaling hypothesis*". Hydrol Sci J. 2011;56: 917–929. doi:10.1080/02626667.2011.588172

[3] Blain GC. "*The modified Mann-Kendall test: on the performance of three variance correction approaches*". Bragantia. 2013;72: 416–425. doi:10.1590/brag.2013.045

[4] Önöz B, Bayazit M. "*Block bootstrap for Mann – Kendall trend test of serially dependent data*". 2011. doi:10.1002/hyp

[5] Hamed KH, Rao AR. "*A modified Mann-Kendall trend test for autocorrelated data*". J Hydrol. 1998;204: 182–196. doi:http://dx.doi.org/10.1016/S0022-1694(97)00125-X

[6] Oliva F, Vegas E, Civit S, Garrido T, Fraile J, Munné A. "*Trend Assessment for Groundwater Pollutants : A Brief Review and Some Remarks*". illustrate. In: Munnc

[7] A, Ginebreda A, Prat N, editors. "*Experiences from Ground, Coastal and Transitional Water Quality Monitoring: The EU Water Framework Directive Implementation in the Catalan River Basin District*", Part 2. illustrate. Springer, 2015; 2015. pp. 25–62. doi:10.1007/698

[8] Bari SH, Rahman TU, Hoque MA, Hussain M. "*Analysis of seasonal and annual rainfall trends in the northern region of Bangladesh*." Atmos Res. 2016; 1–28. doi:10.1016/j.atmosres.2016.02.008

[9] Zarenistanak M, Dhorde AG, Kripalani RH. "*Trend analysis and change point detection of annual and seasonal precipitation and temperature series over southwest Iran*". 2014; 281–295

[10] Saeed, S., Nizamani, M. A., Shaikh, A., Memon, M. A., Abbasi, F. A., & Naqvi, S. M. R. (2019). "*Evaluating the Quality of Point of Sale (POS) Software*". *University of Sindh Journal of Information and Communication Technology*, *3*(2), 69-75.