Flood Frequency Analysis of Tidal Data at Pelabuhan Klang, Malaysia

Firdaus Mohamad Hamzah[1], Hazrina Tajudin[1], Sharifah Mastura Syed Abdullah[2], Ekhwan Toriman[2], Hafizan Juahir[3]

[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*

[3]*East Coast Environmental Research Institute (ESERI), Universiti Sultan Zainal Abidin, Gong Badak, 21300 Kuala Terengganu, Terengganu, Malaysia*

ABSTRACT

Flooding will continue being an issue as long as there are group of peoples who stays nearby the shore line area with high risk factor. Thus, flood frequency analysis at this location should always be updated from time to time. This study focuses on analysis at Pelabuhan Klang area. This port is one of the most important port located in Peninsular Malaysia. It is greatly functional and there are many active activities going on there such as transferring goods between one ship to another, port throughput and the port infrastructure. This study attempts to determine magnitude of high tide at selected return period at Pelabuhan Klang. There are several distributions namely Weibull, Gumbel, Generalized Extreme Value, Generalized Pareto and Lognormal with 3 parameters. L-moment method is selected to determine estimation of parameter due to its robustness to outliers and is most commonly applied in hydrology. Based on the analysis, all five distribution is significant, however, Weibull distribution is most fitted to the annual maximum data of high tide at Pelabuhan Klang. The analysis shows that magnitude at return period of 5-, 10-, 20-, 50- and 100-years are 5.54, 5.60, 5.65, 5.71 and 5.75 m respectively. It is important to conduct the flood frequency analysis and keep it updated from time to time to prevent the occurrence of tidal flooding and reduce the impacts especially for those who lives nearby the shoreline.

Keywords: *Flood Frequency Analysis, Annual Maximum, Pelabuhan Klang, Weibull.*

I. INTRODUCTION

Tide is a fluctuation of sea level, where the rise and fall of the level of the sea is influenced by astronomical and non-astronomical factors. The former involves the gravitational force produced by the moon and sun. Several non-astronomical factors contribute to the occurrence of tides. These include earth rotation, topography, wind and weather pattern, and human factors [1]. Tides start in the ocean and move towards the coastline. They are essentially the rise and fall of sea water. High tide occurs when the highest wave hits certain location during full moon and new moon phases. When the earth, moon and sun lie in a line, both the gravitational forces of the moon and sun exert a greater magnetic force on the water, thereby producing bigger waves.

Floods may have a serious impact on human lives and the environment. In Malaysia, the government agency responsible for the acquisition, processing, archiving and dissemination of sea level data is the Department of Survey and Mapping Malaysia (DSMM). Malaysia has 21 tidal stations, of which 12 are located along the Peninsular Malaysia and nine are located along the coast of Sabah and Sarawak. Most of the tidal stations in Malaysia are established between 1981 to 1986 with the technical aid provided by the Hydrographic Department, Maritime Safety Agency, Japan and is funded by the Japan International Cooperation Agency (JICA) under the Colombo Plan.

Floods in Malaysia are associated with the northeast and southwest monsoon and the inter-monsoon between the two monsoons. Malaysia has an equatorial climate with constant high temperature and high relative humidity. The monsoon season between the months of November and February may bring 600 mm of rainfall within 24 hours in the east coast of Peninsular Malaysia, Sabah and Sarawak. There are three types of flood in Malaysia, namely flash flood, monsoon flood and coastal flood [2]. Flash floods occur as a result of a heavy downpour within a short time period. Monsoon floods are the result of continuous downpour during the monsoon season, while coastal floods are typically caused by the high tides in the coastal area.

According to Department of Statistics Malaysia, total population had increased by 0.35 million from year 2015 to 2019. With only small increase in area, the development in Selangor especially in construction of building and infrastructure had surely reduce the surface runoff and therefore, increasing the risk of flooding especially in urban area. The same issue goes to sub-urban, with addition of improper irrigation and drainage which should be properly monitored by the authorities. Increase in the amount of housing and illegal shop-lots in the suburban area had blocked or reduce the size of the drainage system.

Pelabuhan Klang is known as one of the most important port in Malaysia. One of the activities happening at Pelabuhan Klang is a place for trading and transferring goods from one ship to another. This port is located at a strategic place and ranked top 100 port in the world due to its functionality which includes logistic activity, port throughput and infrastructure [3]. Due to its various functionality, it triggers lot of other issues including land subsidence, coastal erosion which indirectly influences the physical of the port. Other than that, other factor such as sea level rise also cause erosion of shoreline, floods and inundation of land [4].

Thus, the objective of this study is to determine the magnitude at selected return period by considering annual maximum data of high tide at Pelabuhan Klang

II. STUDY AREA

Tidal data is collected at Pelabuhan Klang station. It is located at latitude of 3.05 and longitude of 101.358333. The data is obtained from tide gauge station collected by Department of Survey and Mapping Malaysia. The data consist of daily data recorded for the period from 2004 to 2017. The location of the tide gauge is shown in Figure 1. The tide data collected shows that the pattern of tidal is of semidiurnal type where both low and high tides have the same height.

Figure 1: Location of Pelabuhan Klang Station

III. METHODOLOGY

Analysis of extreme involving two aspects which are study on extreme maximum and extreme minimum. The first refers to analysis on risk and strategies involving flooding, whilst the latter refers to the analysis of issues related to drought. In Malaysia, high rainfall amount throughout the year had caused several flooding which include both, flash and monsoonal floods. This is also caused by improper drainage system, high intensity of rainfall and decrease in surface runoff.

Flood frequency analysis (FFA) is conducted using annual maximum (AM) data [5]. There are several advantages of using AM. One of it is that it can be easily extracted from the daily data. AM data consist of the highest tidal reading in each particular year. For example, period between 2004 to 2017 have 14 AM data. Although AM does not consider moderate and small flood, it definitely considers the most significant flood during that particular year.

Among statistical distribution which are commonly used to describe extreme events are Lognormal (3 Parameters), Generalized Extreme Value (GEV), Generalized Pareto (GPA), Weibull, Fréchet, Gumbel, Pearson (3 Parameters), Generalized Logistics and Kappa distribution [6][7][8]. L-moment is a linear combination of probability weighted moments which gives a simple interpretation of the location, shape and dispersion of the sample data. This study employs L-moment method since it is able to describe the shape of distribution and thus provides a greater degree of accuracy. Additionally, this method is particularly robust when dealing with outliers. This method is not affected by sample variability in contrast to other product moment methods [9], [10]. It is an unbiased method compared to other methods of parameter estimation [11], [12]. L-moment method is suitable for a large sample size due to its asymptotic properties [13].

Hydrological parameters often involve an investigation of extreme events. Researchers investigating flood magnitudes with high return period are concerned with the upper tail of the distribution while those investigating droughts focus on the lower tail of the distribution [14]. Goodness of fit testing is carried out to determine whether the selected distribution explains the data set in the earlier section. The hypotheses for the goodness of fit testing are as follow:

H_0: The data follows a specific distribution
H_A: The data does not follow a specific distribution

AD test statistics is used to compute the probability that a sample may have a normal distribution. It has the ability to overcome the limitation of the KS test even though it can only be used for certain distributions. In contrast to the KS method, AD is more sensitive towards the tail of the distribution. Mathematically, the test statistics for Anderson Darling can be written as follows:

$$A = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left(lnF(X_i) + \ln\left(1 - F(X_{n-i+1})\right)\right) \quad (1)$$

where n is sample size, $F(x)$ is the cumulative distribution function for the tested distribution, and the i-th sample is calculated after sorting the data in ascending order. The p-value is dependent on the statistical value obtained from the above equation. $p-value$ is the probability that the data in this sample is from a random data. The formula for calculating the $p-value$ of a goodness of fit technique proposed by Agostino and Stephen is presented in Table 1.

Table 1: *p*-value formula for AD test statistics

A-D test statistics	$p-value$
$A \geq 0.60$	$p = \exp(1.2937 - 5.709(A) + 0.0186(A)^2)$
$0.34 < A < 0.60$	$p = \exp(0.9177 - 4.279(A) - 1.38(A)^2)$
$0.20 < A < 0.34$	$p = 1 - \exp(-8.318 + 42.796(A) -59.938(A)^2)$
$A \leq 0.20$	$p = 1 - \exp(-13.436 + 101.14(A) -223.73(A)^2)$

The analysis of return period using selected distribution applied the usage of quantile function. Table 2 shows list of distribution which is applied in the FFA.

Table 2: Quantile function for each tested distribution

Distribution	Quantile function
Gumbel	$x(F) = \sigma(\ln(-\ln(F))) + \mu$
Weibull	$x(F) = \beta\left(-(\ln(1-F))^{\frac{1}{\alpha}}\right) + \gamma$
Generalized Extreme Value	$x(F) = \xi + \frac{\alpha}{k}\{1 - (-\log(F))^k\}$
Lognormal (3 Parameter)	$x(F) = e^{\sigma\phi^{-1}(F)+\mu} + \gamma$
Generalized Pareto	$x(F) = \xi + \frac{\alpha}{k}\{1 - (1-F)^k\}$

where $\phi^{-1}(F)$ is the inverse of probability distribution function for normal distribution.

IV. RESULTS AND DISCUSSION

Descriptive analysis of tidal data at Pelabuhan Klang is shown in Table 3 and daily plot of tidal data is in Figure 2.

Table 3: Descriptive analysis of daily high tide data

Station	Minimum (m)	Maximum (m)	Mean (m)
Pelabuhan Klang	2.3	5.7	4.46

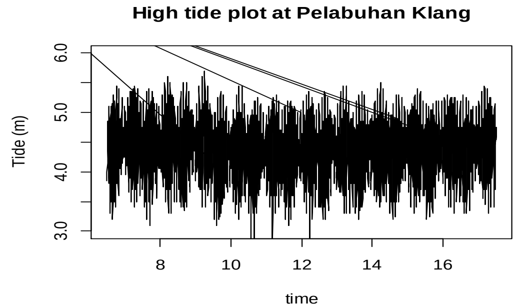

Figure 2: Daily plot of high tide at Pelabuhan Klang

Maximum data which is 5.7 results in significant floods which especially affect those near the Pelabuhan Klang. The descriptive analysis of AM data is tabulated in Table 4. The AM plot is shown in Figure 3.

Table 4: Descriptive analysis of AM data

Station	Pelabuhan Klang
Minimum (m)	5.3
Maximum (m)	5.7
Mean (m)	5.45
Skewness	0.55
Kurtosis	-0.19

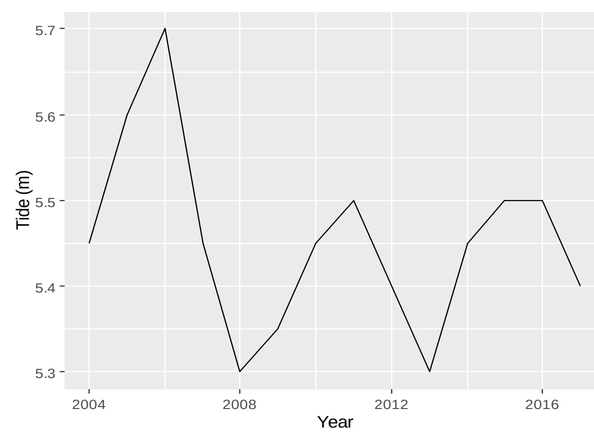

Figure 3: Annual maximum plot 2004-2017 at Pelabuhan Klang

Based on the above Table, minimum and maximum value of tidal data does not show significant difference, however, even a slight increase in the water depth could trigger flood at the study area. The average of high tide data shows 5.45 m near the maximum value. Skewness measures how far the data deviates from horizontal symmetry. The skewness shows positive value which indicate that the data are skewed to the right. Kurtosis measures how tall and sharp the central peak when compared to normal distribution. Based on the AM data, it does not heavy tails compared to normal distribution. It can be seen by histogram plot of AM in Figure 4.

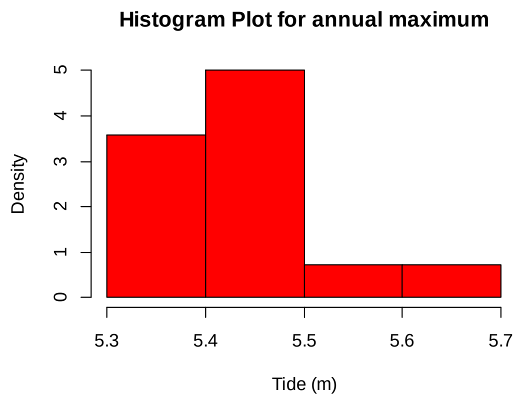

Figure 4: Histogram plot for AM

The result for Anderson-Darling test is shown in Table 5.

Table 5: Goodness of fit testing

Distribution	p-value
Lognormal (3 Parameters)	0.4454
Weibull	**0.461**
Generalized Extreme Value	0.4412
Generalized Pareto	0.3796
Gumbel	0.3321

The flood frequency analysis is conducted using several extreme distributions which is commonly used by previous researcher. Among the distribution, Weibull distribution is selected to determine the magnitude at selected return period. The result is tabulated in Table 6. Weibull distribution practically depends on the shape parameter which is also known as the slope and scale parameter which implies the stretchiness of the shape. The analysis conducted using small samples from 2004 to 2017, hence, this might be the reason Weibull is the most fitted distribution for this data set, in comparison with other available extreme distribution which usually appropriate for larger sample size.

Table 6: Magnitude at selected return period

Return Period (Years)	Magnitude (m)
5	5.5432
10	5.6008
20	5.6501
50	5.7069
100	5.7455

Selection of appropriate distribution is based on the data structure. Annual maximum tidal data shows quiet consistent data with no significant outliers. [8] shows that Weibull is among best fitted distribution in maximum data set at Tana River Basin.

V. CONCLUSION

Flooding involving high tide data is known as "nuisance flooding". Factors which cause the occurrence of high tide are rise in sea level, deterioration of natural and man-made barrier and land subsidence. Flood frequency analysis of tidal need to be updated from time to time especially in Pelabuhan Klang area because of the fluctuation trend in the area which is caused by several active activities since it is the main port used in Malaysia

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] Md Yusof NH, Mahmud MR, Abdullah MH. "*Effect of long term tidal constituents on mean sea level trend during El-Nino and La-Nina phenomena*". Int Arch Photogramm Remote Sens Spat Inf Sci - ISPRS Arch. 2013;42:225–234.doi:10.5194/isprs-archives-XLII-4-W5-225-2017

[2] Mohamad MF, Kamarul M, Samion H, Hamzah SB. "*Physical Modelling for Flood Evaluation of Selangor River Under Tidal Influence*". International Conference Data Mining, Civil and Mechanical Engineering. Bali; 2014. pp. 102–106. doi:10.15242/IIE.E0214013

[3] Ducruet C. "*A metageography of port-city relationships. Ports, cities, and global supply chains*", Ashgate. Ashgate; 2007. pp. 157–172. Available: https://halshs.archives-ouvertes.fr/halshs-00458067

[4] Mohamad MF, Hamid MRA, Awang NA, Shah AM, Hamzah AF. "*Impact of Sea Level Rise Due to Climate Change : Case Study of Klang and Kuala Langat Districts*". Int J Eng Technol. 2018;10: 59–65. doi:10.7763/IJET.2018.V10.1035

[5] Jiang S, Kang L. "*Flood frequency analysis for annual maximum streamflow using a non-stationary GEV model*". ARFEE 2018. Wuhan; 2019. doi:10.1051/e3sconf/20197903022

[6] Romali NS, Yusop Z. "*Frequency Analysis of Annual Maximum Flood for Segamat River*". MATEC Web Conf. 2017;103: 04003. doi:10.1051/matecconf/201710304003

[7] Mohamad Hamzah F, Mohd Yusoff SH, Jaafar O. "*L-Moment-Based Frequency Analysis of High-Flow at the Sungai Langat, Kajang, Selangor, Malaysia*". Sains Malaysiana. 2019;48: 1357–1366. doi:10.17576/jsm-2019-4807-05

[8] Philip Kibet Langat, Lalit Kumar, Richard Koech. "*Identification of the most suitable probability distribution models for maximum, minimum and mean streamflow*". WATER. 2019;11: 1–23. doi:10.3390/w11040734

[9] Bílková D. Alternative Tools of Statistical Analysis: "*L-moments and TL-moments of Probability Distributions*". Pure Appl Math J. 2014;3: 77–94. doi:10.11648/j.pamj.20140302.11

[10] Schlögl M, Laaha G. "*Extreme weather exposure identification for road networks – a comparative assessment of statistical methods*". Nat Hazards Earth Syst Sci. 2017;17: 515–531. doi:10.5194/nhess-17-515-2017

[11] Alahmadi F, Abd Rahman N, Abdulrazzak M. "*Evaluation of the best fit distribution for partial duration series of daily rainfall in Madinah, western Saudi Arabia*". Proc Int Assoc Hydrol Sci. 2014;364: 159–163. doi:10.5194/piahs-364-159-2014

[12] Ummi Nadiah A, Ani S, Zahrahtul Amani Z. "*An analysis of annual maximum streamflows in Terengganu, Malaysia using TL-moments approach*". Theor Appl Climatol. 2013;111: 649–663. doi:10.1007/s00704-012-0679-x

[13] Katz RW, Parlange MB, Naveau P. "*Statistics of extremes in climatology*". Adv Water Resour. 2002;25: 1287–1304. doi:10.1111/j.1467-9574.1989.tb01244.x

[14] Shin H, Jung Y, Jeong C, Heo JH. "*Assessment of modified Anderson-Darling test statistics for the generalized extreme value and generalized logistic distributions*". Stoch Environ Res Risk Assess. 2012;26:105–114. doi:10.1007/s00477-011-0463-y