Statistical Analysis of Rainfall in Kelantan and Pahang

NUR AIMI HAZIRA BINTI ATAN
Universiti Teknologi MARA
imehazira93@gmail.com

NUR SYAHEERA BINTI ATAN
Kolej Universiti Islam Antarabangsa Selangor
syaheera@kuis.edu.my

ABSTRACT

Analysis on extreme rainfall is very important in prediction of peak flood. This research aims to do statistical analysis on frequency of annual maximum daily rainfall that affects the occurrence of peak flood. The next objective highlighted in this study is to identify the most appropriate probability distribution for annual maximum daily rainfall. Gumbel Distribution and Weibull distribution are compared to get the most fitted probability distribution towards extreme rainfall data. Maximum Likelihood Method was used for determination of the parameter of the distributions involved. Goodness-of-Fit Test is applied under the Kolmogorov-Smirnov for checking the adequacy of fitting the model distribution selection to the observed extreme rainfall data. The result shows that Gumbel Distribution is the most fitted probability distribution towards extreme rainfall data.

Keywords: Ekstreme Rainfall, Gumbel Distribution, Weibull Distribution, Maximum Likelihood Method, Goodness-of-Fit Test

INTRODUCTION

Since prehistoric times, people lived near the oceans and rivers to get easy access in term of transportation, food as well as trade. It became the most important source towards people. However, things getting harder day by day. Human activities such as forest exploitation, burning of fossil fules have destroyed ecosystem. It changed earth's atmosphere faster than it has ever changed during its history. The current climate change earth is facing is caused by the increase in global temperatures. Indirectly, it gives high impact on the pattern of rainfall. Extreme rainfall or long periods of rain could be the factor of flood occurance.

According to National Geographic News in 2017, floods is the highest concentration of runoff water flow from its basin area that is over their natural bank. It is inundates land that is normally dry. In some country, they called river as a flood because the water level passing the Danger Level (DL) at particular site. Flood problems are becoming increasingly critical in several area at Peninsular Malaysia. The seasonal heavy continuous rains in the Malaysia since January 2017 have caused flooding in six states including Johor, Kelantan, Pahang, Perak, Selangor and Sabah. The worst hit states are Johor and Pahang where waters have rose up to 1.5 metres in certain areas. Flooding in states of Kelantan and Pahang has occurred for a very long period of time and it became worst due to geological aspects like causing widespread destruction of life and property.

Human capability are limited. There is no way to determine the exact amount of rainfall and rainfall intensity by which flood can be predicted. Various method have been used for flood estimation in order to reduce the risk. However, in prediction and estimate the floods, there is a lack of agreement among hydrologists as to which method or which distribution is the most reliable for defining the flood magnitude-frequency relation. Some effective analysis should be studied in order to get the most fitted flood frequency analysis in Malaysia by using basic rainfall data and to appraise the relative reliability of the methods of extreme value which are Gumbel Distribution and Weibull Distribution.

Therefore, flood frequency analysis is an important factor in flood risk assessment studies since this is a technique used by hydrologists to predict flow values corresponding to specific return periods or probabilities along a river. Frequency is the number of time that a given magnitude flood may occur in a given period. In order to understand how flood frequency analysis works, it is important to understand the concept of return period.

The theoretical definition of return period is the inverse of the probability that an event will be exceeded in a given year. In general, return period, which is also referred as recurrence interval, provides an estimate of the likelihood of any event in one year. These events include natural disasters such as floods or earthquakes. Return periods are used to convey the risks of rate events more effectively that simply stating the probabilities.

Therefore, this research aims to do statistical analysis on frequency of annual maximum daily rainfall that affects the occurrence of peak flood. The next objective highlighted in this study is to identify the most appropriate probability distribution for annual maximum daily rainfall. Gumbel Distribution and Weibull distribution are compared to get the most fitted probability distribution towards extreme rainfall data and the return period will be determined through this research.

LITERATURE REVIEW

According to Saurabh (2015), flood is one of the natural disasters which occur in India every year. It is generally occurred during the southwest monsoon season (June to October) and generated by the random coincidence of several meteorological factors. Flood also can be as overflow, which is actually a natural tragedy that can make a huge disaster which could destroy properties also can reduce substructures (Batool, 2017). Flood frequency analysis of the recorded annual peak discharge over number of the years at the site under consideration is perform as the one way of estimates the flood of specific recurrence interval.

Flood frequency analysis is the most important statistical technique in understanding the nature and magnitude of high discharge in a river (Kalpalatha & Sundar, 2017). It means that the objective is to relate the magnitude of events to their frequency of occurrence through probability distribution. Besides, the ability to estimate the magnitude of the flood in certain return period can help the hydrologist to design the hydrological projects (Ummi Nadiah Ahmad, *et al.*, 2011). Based on May et. al. (2015) reviewed, Log-Pearson type III distribution method used as the projection on the river basin discharges in Upper Klang River in order to perform flood frequency analysis. The study concluded that comparison between simulation of the climate change data is lower than the expected distribution of the estimation on flood frequency distribution. The outcomes of the flood frequency provide the useful information which is regarding the flood magnitudes and return periods of the selected river.

In addition, flood frequency analysis is a statistical technique that uses historical streamflow data to select and fit a probability distribution at a given location within a catchment (Haddad & Rahman, 2016). According to Kasiviswanathan *et. al*., (2017), flood frequency analysis also interested as a tool for the water resources management. In term of probability distribution, past studies of Mukherjee (2013), Gumbel Extreme Value Distribution is used for modelling the peak flood discharge and return period. The result gives the reasonable estimated peak flood discharge. This means that the Gumbel

Distribution is the best model to be used as modelling the extreme value. Flood frequency analysis also was carried out for the Osse river by using Gumbel distribution in order to model the annual maximum discharge of the river for several periods. This method shows that it is useful for storm management in the area of the study. Besides, it has proved that will be give the expected value of the discharges estimated (Solomon, 2013). Roy *et. al.*, (2015) said that the best method to analysing the flood frequency of Puthimari catchment is by using Gumbel distribution. It consists of fitting the theoretical extreme value probability distribution. The methods also used for planning and for infrastructural design.

In some cases of determining and comparing the method between the two extreme value distribution functions on modelling the forest, Weibull distribution were results as the most suitable model for describing the maximum diameters as compare to the Gumbel distribution which is suitable for the minimum diameters (Javier, 2015). However, as overall result, Gumbel distribution is the better ones as the purpose in modelling the distribution of the minimum and maximum values of the tree diameters.

Rainfall return period known as a recurrence interval or sometimes repeat interval is an estimate of the likelihood of rainfall event (Tanguy, 2017). It is a statistical measurement typically based on historic data denoting the average recurrence interval over an extended period of time and is usually used for risk analysis. Past article (Ismail, *et al.*, 2015) stated that in order to model long period flood event, a statistical distribution method is needed. Additional parameters such as flood volume and flood duration can be incorporated in flood frequency analysis by using copula technique. In addition, Saurabh (2015), mentioned that the annual flood series analysis has been carried out to estimate the flood quantiles at different return period at Kosi barrage site of river Kosi. The results used the probability distribution method in order to carried out the predicted discharge for the return period of 10, 25, 50, 100, 200, 500, 1000 in years and all the distribution used are fittings for Kosi River.

METHODOLOGY

Rainfall data recorded at daily intervals are obtained from the Department of Irrigation and Drainage (DID), Ampang Malaysia. The 43 years data from January 1975 to December 2017 for Pahang and Kelantan station are used in this study. Kelantan and Pahang are chosen due to high number of occurrence in floods. The analysis of the rainfall data used R statistical package software. Gumbel distribution and Weibull distribution were used for the rainfall analysis. The data is about fitting the parameters for daily rainfall data to find the best model in modelling flood analysis in Peninsular Malaysia. It is a daily rainfall data in millimeter (mm) measurement for every 5 stations in Pahang and Kelantan respectively. The return period will be determined. The raw data need to be checked through data cleaning before the analysis can be proceeded.

The PDF and Cumulative Distribution Function (CDF) of the Gumbel Distribution are given as:

$$\left.\begin{aligned} PDF: f(x) &= \frac{e^{-(x-\alpha)/\beta} e^{-e^{-(x-\alpha)/\beta}}}{\beta} \\ CDF: F(x) &= e^{-e^{-(x-\alpha)/\beta}}, \beta>0, where(x = x_1, x_2, \ldots, x_N) \end{aligned}\right\} \quad (1)$$

where, α and β are the location and scale parameters of the distribution. The parameters are computed by Maximum Likelihood Method through Equations (2) and (3).

$$\alpha = -\beta \ln\left[\sum_{i=1}^{N} \exp(-x_i / \beta) \Big/ N\right] \qquad (2)$$

$$\beta = \bar{R} - \left[\sum_{i=1}^{N} x_i \exp(-x_i / \beta) \Big/ \sum_{i=1}^{N} \exp(-x_i / \beta)\right] \qquad (3)$$

The PDF and Cumulative Distribution Function (CDF) of the Weibull Distribution are given as:

$$\left.\begin{aligned} PDF: f(x) &= \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-(x/\beta)^{\alpha}} \\ CDF: F(x) &= 1 - e^{-(x/\beta)^{\alpha}} \end{aligned}\right\} \qquad (4)$$

Where, α and β are the shape and scale parameters of the distribution. The parameters are computed by Maximum Likelihood Method through Equations (5) and (6).

$$\alpha = \left(\frac{\sum_{i=1}^{N} x_i^{\alpha} \ln x_i}{\sum_{i=1}^{N} x_i^{\alpha}} - \frac{1}{N}\sum_{i=1}^{N} \ln x_i \right)^{-1} \qquad (5)$$

$$\beta = \left(\frac{1}{N}\sum_{i=1}^{N} x_i^{\alpha} \right)^{1/\alpha} \qquad (6)$$

Return Period (T) is the average length of time in years for an event of given magnitude to be equaled or exceeded. Simply stated, a return period is an estimation of how long it will be between rainfall events of a given magnitude. A fundamental relationship is that between flood return period (T) and probability of occurrence (p).

EMPIRICAL RESULTS

Based on the result in the Table 1 below, it shows the summary of the Kelantan state. Kampung Aring have the lowest mean and JPS Kuala Krai have the highest mean with 81.64mm and 168.25mm, respectively. While, Dabong have the highest amount of standard deviation with 103.72mm as compared to the Kampung Aring that only 18.34mm. Results of the kurtosis for JPS Kuala Krai shows the value is around zero. This means that JPS Kuala Krai only have the data that is nearly distributed to the normal distribution.

Table 1: Summary Descriptive Statistics for Kelantan

Variable	Mean	Standard Deviation	Skewness	Kurtosis
Gua Musang	103.63	29.40	1.50	3.34
Dabong	143.89	103.72	4.35	22.13
JPS Kuala Krai	168.25	68.96	0.79	0.76
Kampung Aring	81.64	18.34	-1.93	6.78
Kampung Laloh	143.66	67.06	2.07	6.02

Table 2 below explained the summary descriptive statistics for Pahang state. Result shows Mardi Sungai Baging have the highest value of mean with 876.16mm while Kuantan shows the highest amount of standard deviation with 100.01mm compared to the Mardi Sungai Baging with 13.00mm. All stations have positive value in skewness. This means that all the station has positively skewed to the right and value of kurtosis shows that almost of the station closed to zero which means the data is nearly to normal distribution.

Table 2: Summary Descriptive Statistics for Pahang

Variable	Mean	Standard Deviation	Skewness	Kurtosis
Felda Bukit Tajau	96.81	31.01	0.19	0.20
Felda Kampong New Zealand	100.42	31.83	0.43	-0.34
Felda Sungai Pancing Selatan	145.45	50.65	1.05	0.73
Kuantan	209.90	100.01	0.91	0.64
Mardi Sungai Baging	876.16	13.00	0.12	-1.46

Gumbel Distribution shows the best model to be used in for Gua Musang since the value of the Akaike Information Criterion (AIC) is equal to 400.6454 which is the lowest one compared to other station. While, Weibull Distribution is best to be used for Kampung Aring with the value of AIC is equal to 402.5863. Therefore, by comparing these two models of distribution for Kelantan, Gumbel Distribution is the best to be used for rainfall analysis since the value of the Akaike Information Criterion (AIC) is lower. However, both distribution for Gumbel Distribution and Weibull Distribution shows the best model to be used is for Mardi Sungai Baging since the value of the Akaike Information Criterion (AIC) is equal to 345.5707 and 349.0763, respectively. Therefore, by comparing these two models of distribution for state of Pahang, Gumbel Distribution is the best to be used for rainfall analysis since the value of the Akaike Information Criterion (AIC) is the lowest.

The shape, α and scale, β is equal to 869.86603 and 11.03263, respectively which were determined by Maximum Likelihood Method. This value of parameters will be used as estimation 1-day extreme rainfall for different return periods vary from 2-year to 100-year adopting Gumbel Distribution. The adequacy of fitting of Gumbel Distribution to the series of Annual Maximum Daily Rainfall was performed by producing Kolmogorov-Smirnov test. From the result, it was observed that the p-value of the test statistic is 0.0003934, which is less than the 5% significance level. Therefore, Gumbel Distribution is the best distribution used in modelling the series of Annual Maximum Daily Rainfall.

Table 3: Peak Flood Discharge for Different Return Periods for Kelantan River

Return Period (year)	Peak Flood Discharge ($10m^3/s$)
2	22.78688
5	6.953836
10	23.39559
15	9.346768
20	10.07898
25	10.56639
50	12.05949
75	12.92403

Based on Table 3 above, the highest amount of peak flood discharge estimation is at the return period of 10-year and followed by 2-year of return period with amount of $23.39559 \times 10m^3/s$ and $22.78688 \times 10m^3/s$ respectively.

CONCLUSION

Overall, the descriptive statistic of the rainfall data for the two selected state shows the effectiveness of the rainfall data in dealing with the floods analysis. Gumbel distribution and Weibull distribution are corresponded with the rainfall data. Based on the best selected model, the estimation of the rainfall estimation was determined by using different years return periods value of 2, 5, 10, 15, 20, 25, 50, 100. The conclusion was drawn from the study is Gumbel Distribution is the best distribution to be used for modelling rainfall data in Peninsular Malaysia as compared to the Weibull Distribution. The Kolmogorov Smirnov test result confirmed the suitability of the Gumbel Distribution as the best model for Annual Maximum Rainfall (AMR). The different return period of rainfall estimation was estimated by using the best model in the previous successful objective which is Gumbel Distribution. Peak Flood Discharge (PFD) explained that the highest amount of peak flood discharge estimation is at the return period of 10-year and followed by 2-year of return period.

References

Alfieri, L., Burek, P., Feyen, L., & Forzieri, G. (2015). Global warming increases the frequency of river floods in Europe. *Hydrology and Earth System Sciences*, 19(5), 2247-2260. Retrieved September 23,2017, from https://www.hydrol-earth-syst-sci.net/19/2247/2015/hess-19-2247-2015.pdf

Batool, Z. (2017). Flood Frequency Analysis of Stream Flow in Pakistan Using L-Moments and TL-Moments. Retrieved October 10, 2017, from https://www.ijariit.com/manuscripts/v3i4/V3I4-1166.pdf

Haddad, K., & Rahman, A. (2016). Investigation on at-site flood frequency analysis in south-east Australia. Retreived August 21, 2017, from https://www.researchgate.net/publication/277233995_Investigation_on_at-site_flood_frequency_analysis_in_South-East_Australia

Ismail, A. Z., Yusop, Z., & Yusof, Z. (2015). Comparison of flood distribution models for Johor River basin. *Jurnal Teknologi*, 74, 123-128. Retrieved September 13, 2017, from https://www.researchgate.net/profile/Zulkifli_Yusop/publication/282311490_Comparison_of_flood_distribution_models_for_Johor_river_basin/links/56658cc108ae192bbf924eb2.pdf

J. Javier Gorgoso-Varela, Alberto Rojo-Alboreca. Use of Gumbel and Weibull functions to model extreme values of diameter distributions in forest stands. Annals of Forest Science, Springer Verlag/EDP Sciences, 2014, 71 (7), pp.741-750.

Kalpalatha, G. and Sundar, P. K. (2017). A Case Study on Flood Frequency Analysis, International Journal of Civil Engineering and Technology, 8(4), 2017, pp. 1762-1767.

http://www.iaeme.com/IJCIET/issues.asp?JType=IJCIET&VType=8&IType=4. Retrieved August 28, 2017, available from: https://www.researchgate.net/publication/316647068_A_case_study_on_flood_frequency_analysis

Kasiviswanathan, K. S., He, J., & Tay, J. H. (2017). Flood frequency analysis using multi-objective optimization based interval estimation approach. *Journal of hydrology*, 545, 251-262.

Mamoon, A., & Rahman, A. (2017). Selection of the best fit probability distribution in rainfall frequency analysis for Qatar. *Natural Hazards*, 86(1), 281-296.

May, R., Zainol, Z., & Shukor, M. F. B. (2015). Flood frequency analysis due to climate change condition at the Upper Klang River basin. In ISFRAM 2014 (pp. 271-281). Springer, Singapore.

Mohammadi, K., Alavi, O., Mostafaeipour, A., Goudarzi, N., & Jalilvand, M. (2016). Assessing different parameters estimation methods of Weibull distribution to compute wind power density. *Energy Conversion and Management*, 108, 322-335.

Mukherjee, M. K. (2013). Flood frequency analysis of River Subernarekha, India, using Gumbel's extreme value distribution. *International Journal of Computational Engineering Research*, 3(7), 12-19.

Nguyen, T. H., El Outayek, S., & Lim, S. H. (2017). A systematic approach to selecting the best probability models for annual maximum rainfalls–A case study using data in Ontario (Canada). *Journal of Hydrology*, 553, 49-58. Retrieved October 3, 2017, from http://www.sciencedirect.com.https.scihub.tw/science/article/pii/S0022169417305176

Odry, J., & Arnaud, P. (2017). Comparison of Flood Frequency Analysis Methods for Ungauged Catchments in France. *Geosciences*, 7(3), 88. Retrieved September 12, 2017, from http://www.mdpi.com/2076-3263/7/3/88

Rahman, A. S., Rahman, A., Zaman, M. A., Haddad, K., Ahsan, A., & Imteaz, M. (2013). A study on selection of probability distributions for at-site flood frequency analysis in Australia. *Natural hazards*, 69(3), 1803-1813. Retrieved October 6, 2017, from http://link.springer.com.https.sci-hub.tw/article/10.1007/s11069-013-0775-y

Roy, P., & De, S. K. (2015). A Comparative Approach To Flood Frequency Analysis Of The Puthimari River In Assam, India. *Asian Journal of Spatial Science*: Vol. III, 90.

Saurabh, S., and Prasad, J. (2015). Flood Frequency Analysis of River Kosi, Uttarakhand, India Using Statistical Approach. Retrived October 1, 2017, from https://s3.amazonaws.com/academia.edu.documents/42192518/flood_frequency_analysis_of_river_kosi__uttarakhand__india_using_statistical_approach.pdf?awsaccesskeyid=akiaiwowyygz2y53ul3a&expires=1513829067&signature=qyimo3keq3d7sv9um5ugpbd91ma%3d&response-content-disposition=inline%3b%20filename%3dflood_frequency_analysis_of_river_kosi_u.pdf

Solomon, O., & Prince, O. (2013). Flood frequency analysis of Osse river using Gumbel's distribution. *Civil and Environmental Research,* 3, 55-59.

Tanguy, M., Chokmani, K., Bernier, M., Poulin, J., & Raymond, S. (2017). River flood mapping in urban areas combining Radarsat-2 data and flood return period data. *Remote Sensing of Environment*, 198, 442-459. Retrieved August 29, 2017, from http://www.sciencedirect.com.https.scihub.tw/science/article/pii/S0034425717303012