Jurnal Teknologi

Full Paper

COMPARISON OF FLOOD DISTRIBUTION MODELS FOR JOHOR RIVER BASIN

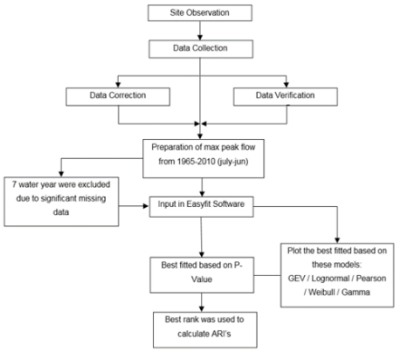
Ahmad Zuhdi Ismail*, Zulkifli Yusop, Zainab Yusof

Faculty of Civil Engineering, Resource Sustainability Research Alliance, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia

Article history
Received
24 April 2015
Received in revised form
4 May 2015
Accepted
9 May 2015

*Corresponding author
ahmadzuhdismail@utm.my

Graphical abstract



Abstract

One of the most useful and commonly used parameters to describe a flood event is peak flow or annual maximum flood. In many localities, storm water control facilities are required and their sizes are determined based on certain peak flow magnitude. This study aimed at estimating the average recurrent interval (ARI) of flood event for Johor River basin based on the distributions of annual peak flow. The analysis used annual maximum flow data from July 1965 to June 2010 recorded at the Rantau Panjang gauging station. Five distribution models, namely Generalized Extreme Value (GEV), Lognormal, Pearson 5, Weibull and Gamma were tested. The goodness fit test (GOF) of Kolmogorov-Smirnov (K-S) was used to evaluate and estimate the best-fitted distribution. The results reaffirm the current practice that GEV is still the best-fitted distribution model for fitting the annual peak flow data. On the other hand, gamma distribution showed the poorest result.

Keywords: General extreme value distribution, goodness of fit test, average recurrent interval, annual peak flow, Johor River basin

Abstrak

Aliran puncak atau banjir tahunan maximum adalah salah satu parameter penting dan sering diguna untuk memodelkan peristiwa banjir. Prasarana untuk kawalan dan pengurusan banir adalah diperlukan oleh pihak berkuasa dan saiznya ditentukan berdasarkan magnitud aliran puncak tertentu. Oleh itu, kajian ini bertujuan untuk menganggarkan purata kala kembali (ARI) bagi lembangan Sungai Johor dengan berdasarkan taburan model pembolehubah banjir. Analisis ini menggunakan data aliran maksimum tahunan daripada Julai 1965 hingga Jun 2010 yang telah direkodkan di stesen Rantau Panjang. Lima jenis model taburan iaitu Generalized Extreme Value (GEV), Lognormal, Pearson 5, Weibull dan Gamma, telah diuji. Uijan Kebagusan Penyuaian (GOF) dari Kolmogorov-Smirnov (KS) digunakan untuk menentukan model taburan terbaik. Keputusan analisis menyokong amalan sediada bahawa model GEV merupakan taburan terbaik untuk data banjir tahunan. Sebaliknya, taburan gamma menunjukkan prestasi yang paling lemah.

Kata Kunci: Taburan umum nilai lampau, uijan kebagusan penyuaian, purata kala kembali, aliran puncak tahunan, Lembangan Sungai Johor

1.0 INTRODUCTION

Being located near to the equator and surrounded by seas, Malaysia receives high annual rainfall which is always so intense. Over the past few years, there is evident of increase rainfall intensity and coupled with the expanding impervious area have resulted in more extreme and frequent flood occurrence [2]. Flood has

caused tremendous losses to properties and sometime life. There is a continuous interest in determining the most appropriate data distribution for flood frequency analysis. Information on average recurrent interval, derived from frequency analysis is crucial for hydraulic analysis and designing hydraulic structure. The design must consider metrological, geomorphologic, economics, land and topographic conditions [13].

In order to model long period flood event, a statistical distribution method is needed. [3], [4], [7], [8], [9], [10], [15], [17], [18], [19] used statistical distributions to model the long term flood characteristics. However, additional parameters such as flood volume (Qv) and flood duration (Qd) can be incorporated in flood frequency analysis by using copula technique, this technique still needs refinement before it can be used as a standard practice. The study focuses on single parameter of flood frequency, which is Qp. In Malaysia, a 100 year ARI has been used as a practice for designing hydraulic structure such as dams, channels and bridges [13]. Recently, this standard has been extended to 200 years return period for construction of urban drainage and flood control design [6]. There is always a finite probability risk in all hydraulics works especially for flood estimation. Even though the procedure for estimating rainfall design has been well documented, it is still subject to spatial variability.

Thus, a single probability distribution may not be applicable for different sites [13]. Generalized Extreme Value (GEV) is the most commonly used distribution in flood frequency analysis in Peninsular Malaysia [23]. However, based on annual flood data from 23 river basins in Sarawak, the peak flow data were best fitted by GEV and Generalized Logistic distributions [11]. Similarly, [1] suggests Generalized Logistic distribution for fitting annual flood data for rivers in Negeri Sembilan. The modelling results from this study were then used to estimate flood sizes of selected Average Recurrent Interval (ARI).

2.0 MODEL FORMULATION

2.1 Study Area

Johor river catchment covers an area of approximately 747.1 km^2 with 122.7 km in length of the main river (Figure 1). There are about 24 sub-catchments within Johor river basin. The delineation of sub catchment was based on the topography and river network. Sg. Sayong, Sg. Pengeli, Sg. Sebol, Sg. Linggu, Sg. Seluyut are the main tributary of Johor river. The average area of sub catchment was about 6226.1 hectare.

The average slope is about 0.1 %, which means, Johor River is a low lying and flat area. The presents study examines the performance of five probability distribution models, namely GEV, Lognormal, Pearson 5, Weibull and Gamma for modelling annual flood of Sungai Johor. These models were chosen because they are commonly recommended by many researchers [10], [15], [16]. GEV, Gamma, and Weibull models are classified as extreme event flood distribution model.

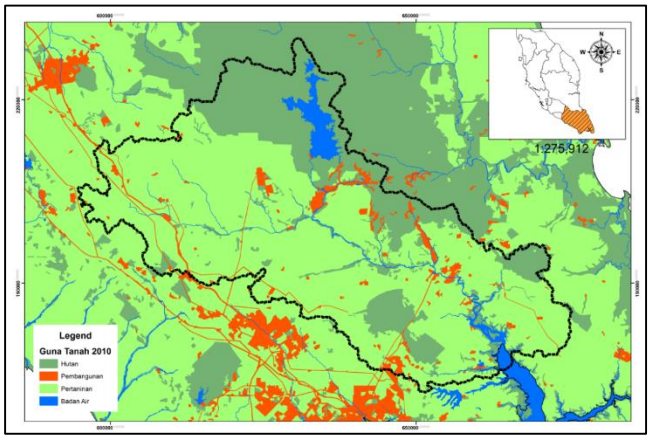


Figure 1 The land use cover for Johor river basin (forest: darker green; residential: orange; vegetation: light green; water: blue)

2.2 Frequency Analysis Models

There have been many studies on flood frequency analysis. [8] said flood frequency can be estimated using observed and simulated rainfall data in order to validate watershed models. [23] compared different flood frequency method and found that the hydro computer simulation program was the most successful in defining flood frequency curve. The simulated flow matched the observed data. [20] used TR-20 computer program to simulate flood, while Alexander (1993) discussed the method of storm transposition to estimate the frequency of huge flood.

Frequency analysis basically deals with statistical properties of rainfall or runoff (flow) series. In practice, these techniques are primarily used for larger catchments because they are more likely to be gauged and have longer record [11]. However, it is also applicable to midsize catchments, provided the record length is adequate. For ungauged catchment, frequency analysis can be used in a regional context to develop flow characteristic applicable to hydrological homogeneous regions. Nowadays, there are increasing concern on developing distribution models to best fit observed data using distribution such as, GEV, Lognormal, and Weibull. Table 1 explains the advantages of common model distribution for flood frequency analysis used in this study.

Table 1 The summary of common distribution models for flood frequency analysis

Model	Advantages
Log Pearson 5	Extrapolation can be made with values of events with return periods well beyond the observed flood events. It is a standard technique used by Federal Agencies in the United States.
GEV	Suitable for extreme event/ peak flow
Lognormal	A common choice if the data is positively skewed.
Gamma/ Weibull	The Gamma and Weibull distributions are two distributions that are closely related to the lognormal distribution

In general, a distribution with a larger number of flexible parameters such as GEV will be able to model the input data more accurately than a distribution with a less number of parameters such as Gumbel. Frequency analysis uses random variables and probability distributions. The former follows a certain probability distribution while the latter is a statistical function that describes the relative chance of occurrence for all possible outcomes of the random variables [22]. In statistical notation, $P\ (X = x_1)$ is the probability P that the random variable X takes on the outcome x_1. A shorter notation is P (x_1). Distribution analysis is an advanced statistical analysis to mainly repattern the flow rate for many years collected by the authority. The statistical frequency models are not only for routed flow, but also for maximum and minimum peak flow design.

Therefore, the design storm must consider the metrological, geomorphological, economics, land, soils, and topography factors. One of the common practices in hydrology is estimating the Annual Exceedance Probability (AEP) and Average Recurrent Interval (ARI) [20]. ARI refers to the return period in time between the events that have same magnitude, volume and duration.

The specify ARI is expressed as:

$$Tr = 1/P \qquad (1)$$

where T_r is return period, P is AEP in percent, hence 1% AEP has 100 years return period.

3.0 METHODOLOGY

Figure 2 shows general methodology for frequency analysis. The 1st phase determined the annual maximum flow from hourly data between July 1965 to Jun 2010. Continuous missing of record of more than three months were removed. As a result, only 38 annual maximum flood data were used in this study.

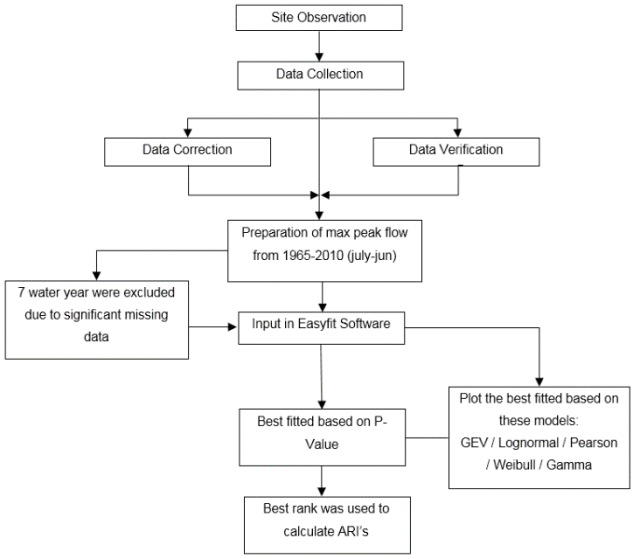


Figure 2 The methodology of frequency analysis in this study

In the 2nd phase, the 38 annual flood data were analyzed using EasyFit Software to determine the distribution models that can best fit the data.

3.1 Goodness of Fit Test

In order to determine the best fitted distribution model, GOF test was used. The GOF test is appropriate when the data is random and its distribution follows the theoretical probability distribution function. K-S at 5% level of significant was used to define the best fit ranking (2).

Kolmogorov-Smirnov (K-S) is given by (2),

$$F(x) = \frac{1}{n}\sum_{i=1}^{n} I(xi \le x) \qquad (2)$$

where i (Condition) = 1 if true and 0 otherwise. Given two cumulative probability functions Fx and Fy the Kolmogorov-Smirnov static test (D_+ andD_-) are given by:

$$D_+ = \max(Fx(X) - Fy(x)) \quad (3)$$

$$D_- = \max(Fy(x) - Fx(x)) \quad (4)$$

where x_1 and x_2 are the lower and upper limits for bin I, respectively. The probability difference model test (equation 5) is useful to assess how good a theoretical distribution fits into the observed data and to compare the performance of several fitted distributions.

$$Difference\ (x) = F_n(x) - F(x) \quad (5)$$

3.1 Quantile Estimation of GEV in ARI

A common approach of describing hydrologic events is by stating the Annual Exceedance Probability (AEP) or the Average Recurrent Interval (ARI). An ARI represents a statistical average number of years between similar events over long periods of record. The probability of P of a given ARI, T_r, occurs on an average once in N successive years can be determined using equation 6;

$$P = 1 - \left(1 - \frac{1}{Tr}\right)^N \quad (6)$$

where *P* is probability of a returning value, T_r is return period, and *N* is equivalent to interval of years. Specifically, the return period, T_r, is given by

$$T_r = \frac{1}{p} \quad (7)$$

where T_r is in years and *P* is the AEP in percent. Hence, a 1% of AEP has ARI 100 years. A design flood is probabilistic or statistical estimation being generally based on some form of probability analysis of flood and rainfall data. In hydrology, a design is not only for routine flow design, but more importantly is for maximum flood estimation or maximum peak flow for several calculated years. The design is intended to obtain the value with extremely low probability of exceedance.

The distribution function of x is given by [23]:

$$\mathrm{x} = u + \frac{a}{x_{avg}}[1 - (-logF)^K] \quad (8)$$

The probability of a flood to occur in any year given by:

$$P_x = 1 - \frac{1}{T_r} \quad (9)$$

where, T_r is return period, the T-year quantile can be estimated by equation 10;

$$x_T = u + \frac{a}{k}[1 - \{-\log(1 - \{-\log(1 - \frac{1}{T})\}^k] \quad (10)$$

4.0 RESULTS AND ANALYSIS

Figure 3 shows the annual flood variation from 1965 to 2010. The highest flow of 724.7 m³/s was recorded in 1985 and the lowest in 1973, which was 76.4m³/s. There are six biggest floods over the 45 years period, which occurred in 1969, 1979, 1982, 1989, and 1995 and 2006/07. In 2006/07 flood event, one of major contributors to flood was tidal effect at the downstream. During those big floods the tidal influence was pronounce.

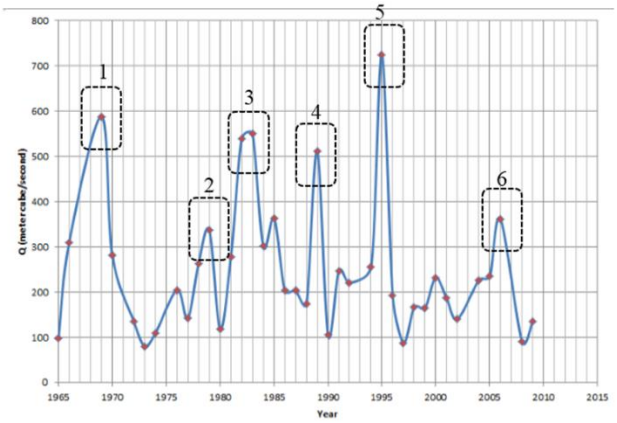


Figure 3 The annual peak flow at Rantau Panjang gauging station from 1965 to 2010

Table 2 presents the descriptive statistics of the annual flood data. The maximum, minimum, and mean of annual flood are 724.73 m³/s, 76.89 m³/s and 251.59 m³/s respectively. The data was positively skewed with coefficient of variation of 61%.

Table 2 Descriptive statistics of annual flood between 1965 and 2010 at Rantau Panjang gauging station

Statistics	Value
Sample size	38
Range	645.68
Min	76.89
Max	724.73
Mean	251.59
Variance	23527
Std.Dev	153.38
Coef.of Variation	60.9%
Std. Error	24.88
Skewness	1.43
Excess Kurtosis	1.77

Table 3 presents the best parameter estimates for the five distribution models which are shape parameters (∝, k), continuous scale parameters (σ, β), and continuous location parameters (μ, γ). Those parameters were generated using EasyFit software. Only results for three parameters are shown, which are better than two parameter models.

Table 3 Fitting results for probability distribution of annual flood

	Distribution	Parameters
1	Gamma	α=1.091 β=158.6 γ=78.569
2	Gen. Extreme Value	k=0.19646 σ=93.782 μ=175.09
3	Weibull	α=1.0908 β=178.53 γ=78.521
4	Lognormal	σ=0.74468 μ=5.0606 γ=46.792
5	Pearson 5	σ=3.8871 β=762.79 γ=-8.3373

Table 4 presents the performance ranking of various cumulative density based on the K-S GOF tests. GEV distribution is ranked the first, followed by Pearson 5, Lognormal, Weibull and the least for Gamma. The ranking is based on P-value. A P-value closer to one indicates a better-fit distribution. In this analysis, the GEV with P-value of 0.99 emerges as the best distribution model.

Table 4 Goodness-of-fit test ranking for various distributions of annual flood

Distribution	Kolmogorov Smirnov	
	P	Rank
GEV	0.99010	1
Pearson 5	0.98806	2
Lognormal	0.97408	3
Weibull	0.97101	4
Gamma	0.90748	5

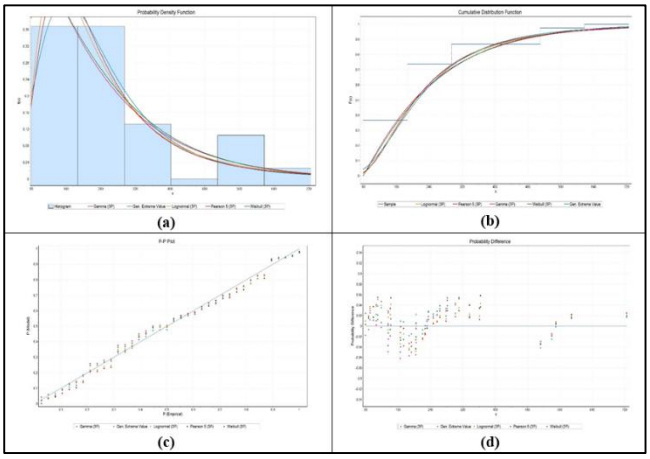


Figure 4 (a) Probability density functions, (b) cumulative distribution function, (c) probability-probability plots and (d) probability difference plot for the five distribution functions

The PDF for the five tested distribution models; GEV (blue), Gamma (maroon), Log Normal (yellow), Pearson-5 (red) and Weibull (green) are shown in Figure 4. The distribution that has the most number of points close to the line represents the best fitted distribution model. Based on the results, GEV model is the best fitted model distribution. The cumulative distribution function (CDF) in figure 4b shows the non-exceedance probability for a given magnitude. The P-P plot (Figure 4c) is a graph of the empirical CDF values against the theoretical CDF values. It's recommended that if the maximum absolute difference is less than 0.05 (5%) the fit can then be considered as good. Through all the patterns in figure 4, again the best fitted is GEV model distribution.

Table 5 The ARI based on GEV distributions model

I	Return period T (yr)	Probability P (%)	Flood discharge Q (m3/s)
1	1.05	95	7
2	1.11	91	45
3	1.25	80	94
4	2	50	209
5	5	20	363
6	10	10	466
7	25	4	595
8	50	2	691
9	100	1	786
10	200	0.5	856

Table 5 presents the estimated flood sizes for ARI between one to 200 years. However in view of limited long term data, the reliability of the estimates decrease as the ARI increase. Based on the results, the estimated peak flows for 50, 100 and 200 years return period, are 691 m^3/s, 786 m^3/s and 856 m^3/s respectively.

5.0 CONCLUSION

Five probability distributions namely GEV, Lognormal, Pearson 5, Weibull, and Gamma were tested on their ability to fit annual flood data measured at Sungai Johor. GEV distribution provided the best results for fitting the annual flood, followed by Pearson 5, Lognormal, Weibull and the least for Gamma. Missing data during big flood events were simulated in order to which otherwise the flow is underestimated due to bank overflow. The estimated peakflow for 50, 100 and 200 year return periods are 691 m^3/s, 786 m^3/s and 856 m^3/s respectively.

Acknowledgement

The authors are grateful to the Department of Irrigation and Drainage (DID) Malaysia for supplying the flow data. Thanks to UTM's Research Management Centre for facilitating this research through research vote number 00G08.

References

[1] Ahmad, U. N., A. Shabri, Z. A. Zakaria. 2011. Flood Frequency Analysis of Maximum Stream Flows Using L-Moments and TL-Moments Approach. *Applied Mathematic Science.* 5(5): 243-253.

[2] Bates, P. P., Horith, M. S., C. N. and Mason, D. C. 1997. Integrating Remote Sensing Observation of Flood Hydrology and Hydraulic Modelling. *Hydraulically Process.* 11: 1777-1795.

[3] Bobee, B., G. Cavidas, F. Ashkar, J. Bernier and P. Rasmussen, 1993. Towards A Systematic Approach To Comparing Distributions Used In Flood Frequency Analysis. *Journal of Hydrology.* 142: 121-136.

[4] Bobee, B. and P. F. Rasmussen. 1995. Recent Advances in Flood Frequency Analysis. *Rev. Geography.* 33(S2): 1111-1116.

[5] Danazumi, S. and S. Shamsudin. 2011. Modelling the Distribution of Inter-Event Dry Spells for Peninsular Malaysia. *Applied. Science. Research*. 7: 333-339.

[6] Deni, S. M. and A. A. Jemain. 2009. Fitting The Distribution of Wet and Dry Spells With Alternative Probability Models. *Atmosphere Physics.* 104: 13-27.

[7] Garde, R. J. and U. C. Kothyari. 1990. Flood Estimation in Indian Catchments. *Journal of Hydrology*. 113: 135-146.

[8] Gunasekara, T. A. G. and C. Cunnane. 1992. Split Sampling Technique for Selecting a Flood Frequency Analysis Procedure. *Journal of. Hydrology*. 130: 189-200.

[9] Haktanir, T. 1992. Comparison of Various Flood Frequency Distributions Using Annual Flood Peaks Data of Rivers in Anatolia. *Journal of Hydrology*. 136: 1-31.

[10] Haktanir, T. and H. B. Horlacher. 1993. Evaluation of Various Distributions for Flood Frequency Analysis. *Journal of Hydrology.* 2(1-2): 15-32.

[11] Lim, Y. H. and Lye, L. M. 2003. Regional Flood Estimation for Ungauged Basins in Sarawak, Malaysia. *Hydrological Science Journal*. 48(1): 79-94

[12] Malcom G. Anderson, Paul D. Bates. 2007. Model Validation Perspective in Hydrological Science. University of British, United Kingdom.

[13] MASMA. 2000. Urban Stormwheater Management Manual for Malaysia, Volume 4/Chapter 3, Department of Irrigation and Drainage.

[14] Mitosek, H. T. and W. G. Strupczewski. 2004. Simulation Results of Discrimination Procedures. Retrieved from: http://www.igf.edu.pl/.

[15] Mitosek, H. T., W. G. Strupczewski and V. P. Singh. 2002. Toward An Objective Choice of an Annual Flood Peak Distribution. Proceeding of the 5th International Conference on Hydro-science and-engineering, Published on CR ROM: Advances in Hydro-Science and Engineering,

[16] Mohsen S., Zulkifli Y., Fadhilah Y. 2012. Modelling the Distribution of Flood Characteristic for a Tropical River Basin. *Applied Sciences, Engineering and Technology.* 6(4): 733-738.

[17] Mohsen S., Zulkifli Y., Fadhilah Y. 2013. Comparison of Distribution Models for Peakflow, Flood Volume and Flood Duration. *Applied Sciences, Engineering and Technology.* 6(4): 733-738.

[18] Mutua, F. M. 1994. The Use of the Akaike Information Criterion in the Identification of an Optimum Flood Frequency Model. *Journal of Hydrology Science*. 39(3): 235-244.

[19] Rao, A. R., and Hamed, K. H. 2000. Flood Frequency Analysis. *International Journal of Climatology.* 29: 385-416.

[20] Suhaila, J. and A. A., Jemain. 2008. Fitting the Statistical Distribution for Daily Rainfall in Peninsular Malaysia Based On AIC Criterion. *Journal Applied Science.* 3: 1027-1036.

[21] Takara, K. T. and J. R. Stedinger. 1994. Recent Japanese Contributions to Frequency Analysis And Quantile and Quantile Lower Bound Estimators. *Journal of Hydrology Science*. 217-234.

[22] Vogel, R. M., W. O. Thomas and T. A. McMahon. 1993. Flood-Flow Frequency Model Selection in Southeastern United States. *Journal Water Resource Planning Management*. 119(3): 353-366.

[23] Zalina, M. D., M. N. M., Desa, V. V. Nguyen and A. H. M. Kassim. 2002. Selecting a Probability Distribution for Extreme Rainfall Series in Malaysia. *Journal Water Sciences Technology*. 45: 63-68.