IDENTIFICATION AND SELECTION
OF BEST-FITTING CANDIDATE DISTRIBUTION
FOR RAINFALL FREQUENCY ANALYSIS
IN CAMERON HIGHLANDS

by

Amir Hashim b. Mohd. Kassim [*1]
Choi Lim Fatt [*2]
Dept. of Hydraulics and Hydrology
Faculty of Civil Engineering

ABSTRACT

In frequency analysis based on analytical method, there are quite a number of probability distributions to be used for quantile estimation. The selection of inappropriate one will lead to either overestimation or underestimation of the quantiles. Thus the identification and selection of the best fitting probability distribution should be given emphasis. The L-moment method offers advantages over the conventional method of moment and thus is more reliable in the distribution identification. The focus of this study is on the identification and selection of best fitting probability distribution, based on L-moment ratio parameters and L-moment ratio diagram. The results show that the GEV (Generalized Extreme Value) distribution fits quite well to data series at most of the homogeneous regions and rainfall intervals.

[*1] Head of Hydraulics and Hydrology Department, Faculty of Civil Engineering, Universiti Teknologi Malaysia.

[*2] Research Officer, Hydraulics and Hydrology Department, Faculty of Civil Engineering, Universiti Teknologi Malaysia.

INTRODUCTION

Practitioners usually need to estimate the recurrence of rainfall extreme event at particular magnitudes in certain periods of time, through frequency analysis. This is important as design variable for structures such as reservoirs, spillways, irrigation networks and drainage systems. The estimation involves interpolation and extrapolation of the rainfall records available. In the current practice, the graphical methods based on probability papers are very common among practitioners since several decades ago due to their simplicity. However, the reliability of these methods is in question because the solution has been oversimplified. This could lead to overestimation (which is a waste of money due to overdesign) or underestimation (which could be a threat to human lives as structures may damage due to underdesign) of the quantiles estimated.

Thus, in order to minimize the extent of these problems, the analytical methods should be used instead of graphical methods. For analytical frequency analysis, the best fitting probability distribution needs to be identified or selected. This part of study will look into this crucial aspect in frequency analysis, which will critically affect the results of quantile estimation later.

OBJECTIVES

This part of study is carried out with the objective of:
a)   identifying and selecting the best fitting probability distribution for rainfall frequency analysis in Cameron Highlands.

SCOPE OF WORK

In this part of study, the annual maximum data series (based on water year) of 1-day, 2-day, 3-day, 5-day and 7-day intervals from 14 rainfall stations (with 531 station-years of data) in Hulu Telom Catchment and Bertam Catchment, are used. APPENDIX A gives the details of the rainfall stations while APPENDIX C shows the locations. The homogeneous regions with the respective rainfall stations are stated in APPENDIX B, which has been determined in another part of study.

Fig. 1 shows the overall procedure for frequency analysis used in this study. The selection and identification of best fitting distribution is carried out after data series identification (annual maximum series based on water year), data screening (hypothesis testing on independence, trend, randomness and homogeneity), parameter estimation (using on L-moment method) and regionalization (cluster analysis based on Euclidean distance measure and Ward's clustering method on 7 selected variables) but before quantile estimation (adopting the overall best fitting distribution) and generalized maps plotting (using kriging method as measure of point rainfall interpolation).

In this part of study, the differences of regional L-skewness and L-kurtosis between samples of rainfall data series and the five candidate probability distributions are computed. This will provide a measure about the degree of fitness of those candidate distributions to the sample of data. The distribution with the lowest difference will be selected as the best-fitting one. Meanwhile, the L-moments ratio diagrams are constructed and the regional sample L-skewness and L-kurtosis are plotted into the diagrams for identifying the best fitting distribution. Nevertheless it should be emphasized here that this is only a complementary measure as visual inspection is more subjective.
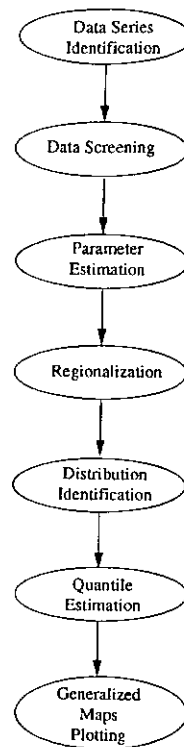

Fig. 1 : Overall Frequency Analysis Procedure in This Study

SOURCES OF DATA

The daily rainfall data is provided by the Department of Irrigation and Drainage (DID), Tenaga Nasional Berhad (TNB) and Malaysian Meteorological Service (MMS). Only 14 of the stations are selected for frequency analysis since their records are long enough for the purpose of this study.

4

LITERATURE REVIEW

Some studies have been done over the last few years on the application of L-moment method in hydrology, especially in frequency analysis. Vogel and Fennessey (1993) have compared the conventional product moment ratio and L-moment ratio diagram and concluded that product moment estimates of coefficient of variation and of skewness should be replaced by L-moment estimators for most goodness-of-fit applications in hydrology. Pilon et al (1991) concluded from their study and analysis of annual maximum precipitation in Ontario, Canada, for durations ranging from 5 minutes to 24 hours, that the variability in the L-skewness and L-coefficient of variation was primarily due to sampling variability. Pilon and Adamowski (1992) also came to similar conclusion for the study in the province of Nova Scotia, Canada, that by using the L-statistics and simulation, the variability of L-skewness is due in large part to sampling error. From the studies, it is found that the L-moment method offers some advantages over the product moment method. For instance, it is less sensitive to the effects of sampling variability and outliers, especially in small samples.

Loke (1994) clustered the Klang River Basin (with 20 rainfall stations and 624 station-year of data) into homogeneous regions and computed the L-moment ratio estimators for every region at rainfall with durations of 1-day, 2-day, 3-day, 5-day and 7-day. He suggested that the GEV distribution can best fit the regions of different rainfall durations through the use of L-moment ratio diagram. However, the approach was too subjective because only the visual inspection on the diagrams was done. Instead, the goodness-of-fit should be judged based on some numerical values that can be computed.

METHODOLOGY

The identification of best fitting distribution is done by comparing the regional L-kurtosis (t_4), between the sample of rainfall data and the candidate distributions, with L-skewness (t_3) based on the sample L-skewness. The difference between them is computed and the distribution contributes the least difference in regional L-kurtosis will be selected. Another complementary approach is by plotting the regional sample L-moment ratio parameters into L-moment ratio diagrams. The identification is done by selecting the candidate distribution with the curve nearest to the sample point.

Under this method, the best fitting frequency distribution is selected from among the following candidate frequency distributions :
a) Generalized Extreme Value (GEV) distribution
b) Log-Normal (LN) distribution
c) Pearson Type 3 (P3) or Gamma (GAM) distribution
d) Generalized Logistic (GLO) distribution
e) Generalized Pareto (GPA) distribution

Hosking and Wallis (1993) defined group average L-moment ratios, with N sites weighted proportionally to their record lengths and $t_r^{(i)}$ as sample L-moment ratios, as

$$\bar{t}_r = \frac{\sum\limits_{i=1}^{N} n_i t_r^{(i)}}{\sum\limits_{i=1}^{N} n_i} \tag{1}$$

where $r = 3, 4, ...$

For each of the candidate distributions, the L-kurtosis given by Hosking (1990) and Maidment (1993) are as in the equations below.

$$\tau_4^{GEV} = \frac{\left(1 - 6*2^{-k} + 10*3^{-k} - 5*4^{-k}\right)}{\left(1 - 2^{-k}\right)} \tag{2}$$

where $k = 7.8590\left[\dfrac{2}{(3+\tau_3)} - \dfrac{\log 2}{\log 3}\right] + 2.9554\left[\dfrac{2}{(3+\tau_3)} - \dfrac{\log 2}{\log 3}\right]^2$

$$\tau_4^{LN} = 0.12282 + 0.77518\tau_3^2 + 0.12279\tau_3^4 - 0.13638\tau_3^6 + 0.11368\tau_3^8 \tag{3}$$

$$\tau_4^{P3} = 0.1224 + 0.30115\tau_3^2 + 0.95812\tau_3^4 - 0.57488\tau_3^6 + 0.19383\tau_3^8 \tag{4}$$

$$\tau_4^{GLO} = \frac{\left(1 + 5(-\tau_3)^2\right)}{6} \tag{5}$$

$$\tau_4^{GPA} = \frac{(1-k)(2-k)}{(3+k)(4+k)} \tag{6}$$

where $k = \dfrac{1 - 3\tau_3}{1 + \tau_3}$

RESULTS AND DISCUSSION

From the previous CFA analysis, the values of sample L-skewness (t_3) and L-kurtosis (t_4) for all rainfall intervals at each station are obtained. The weighted sample L-moment ratio parameters are calculated to take into account the length or duration of records for each station being used in the analysis.

APPENDIX D shows the sample calculation for 1-day rainfall. The value n is the sample size, that is the length of rainfall record (in years) being used. For the same L-skewness of both sample and candidate distribution, the L-kurtosis for each candidate distribution is computed and compared to the sample L-kurtosis. The candidate distribution with the lowest difference in L-kurtosis is selected as the best fitting distribution for each region respectively. The positive value indicates that the regional weighted average sample point (with t_3 and t_4) is below the curve of the candidate distribution, and vice versa. **APPENDIX E** exhibits the respective L-moment ratio diagrams plotted for 1-day rainfall interval, with regions 1A, 1B and 1C accompanied by the scatter plots of L-moment ratio parameters of all rainfall stations.

For instance for Region 1B, from **APPENDIX D**, the differences in L-kurtosis ranked accordingly, are 0.0131 (GEV), 0.0182 (LN), -0.0273 (GLO), 0.0317 (P3) and 0.1026 (GPA). From **APPENDIX E**, we can see that the sample point is nearest to GEV curve and furthest from GPA curve. Thus the GEV distribution is the best fitting distribution for Region 1B.

Region	Best Fitting Distribution		2nd Best Fitting Distribution	
	Distribution	$\tau_4 - \tau_4^{distribution}$	Distribution	$\tau_4 - \tau_4^{distribution}$
1A	P3	-0.0017	GEV	-0.0121
1B	GEV	0.0131	LN	0.0182
1C	P3	0.0170	LN	-0.0623
2A	GEV	0.0158	GLO	-0.0227
2B	P3	-0.0131	GEV	-0.0185
2C	LN	-0.0267	P3	0.0514
3A	LN	0.0020	GEV	-0.0084
3B	GEV	-0.0037	LN	-0.0043
3C	LN	-0.0167	GPA	-0.0487
5A	GLO	0.0054	GEV	0.0401
5B	GEV	-0.0293	P3	-0.0355
5C	LN	-0.0058	GPA	-0.0365
7A	GLO	0.0637	GEV	0.0885
7B	LN	-0.0024	P3	0.0036
7C	GPA	0.0077	GEV	-0.0186

Table 1. Best-Fitting Probability Distribution

As a whole, from **Table 1**, it is shown that LN distribution is the best fitting distribution for 5 regions while GEV, P3, GLO and GPA are best for 4, 3, 2 and 1 regions respectively. The LN distribution should be ranked as the overall best fitting distribution. However, when the differences in L-kurtosis are evaluated carefully, 3 of the 5 regions with the LN distribution as the best are regions C which have only 1 rainfall station. Thus it is misleading to select the LN distribution as the overall best. Although the GEV distribution dominates only 4 regions, but the regions have more stations. Besides, from **Table 1** it is found that the GEV distribution is the second best fitting for 6 regions, compared to the LN distribution with only 3 regions. Furthermore, for some regions, the difference of GEV distribution is very close to the best fitting distribution. For example, for Region 2B and 3A, the best fitting distributions are P3 (-0.0131) and LN (0.0020) but the second best distribution is GEV (with -0.0185 and -0.0084) respectively. The GEV distribution is selected for regions of most rainfall intervals.

Thus it is more reasonable to say that the overall best fitting probability distribution for rainfall frequency analysis in Cameron Highlands is the GEV distribution. In another study, Loke (1994) clustered the rainfall frequencies in Klang River Basin based on 20 rainfall stations (with 624 station-year data) and concluded that the GEV distribution could fit quite well to all the regions of 1-day, 2-day, 3-day, 5-day and 7-day rainfall. Thus, from the results of both the studies, the GEV distribution has the potential of being adopted as the standard probability distribution for rainfall frequency analysis in Malaysia. Other countries such as United Kingdom, United States of America, Canada and China has adopted Generalized Extreme Value, Log Pearson Type III, 2-Parameter Log-Normal and Pearson Type III distributions respectively. However, the adoption of certain probability distribution for frequency analysis in Malaysia can only be further verified after more studies covering the whole Malaysia being conducted.

CONCLUSION

From the analysis, the best-fitting probability distribution for each region has been identified. Among them the GEV distribution is selected as the overall best fitting probability distribution for rainfall frequency analysis in Cameron Highlands. The quantile estimation will be based on the GEV distribution for all the regions.

ACKNOWLEDGEMENT

8

REFERENCES

Greenwood, J.A., Landwehr, J.M., Matalas, N.C. and Wallis, J.R. (1979). "Probability Weighted Moments : Definition and Relation to Parameters of Several Distributions Expressable in Inverse Form", *Water Resources Research*, 15(5) : 1049-1054.

Haan, C.T. (1977). *Statistical Methods in Hydrology*, Iowa State University Press, Ames, IA.

Hosking, J.R.M. (1986). "The Theory of Probability Weighted Moments", *Res. Rep. RC12210*, IBM Research, Yorktown Heights, N.Y. (As cited by Hosking and Wallis, 1993)

Hosking, J.R.M. (1989). "Some Theoretical Results Concerning L-Moments", *Res. Rep. RC14492*, IBM Research, Yorktown Heights, N.Y.

Hosking, J.R.M. (1990). "L-Moments : Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics", *J. R. Statist. Soc.*, 52(1) : 105-124.

Hosking, J.R.M. and Wallis, J.R. (1993). "Some Statistics Useful in Regional Frequency Analysis", *Water Resour. Res.*, 29(2) : 271-281.

Hosking, J.R.M., Wallis, J.R. and Wood, E.F. (1985). "Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments", *Technometrics*, 27(3) : 251-261.

Loke, K.W. (1994). *Clustering Rainfall Frequencies for Klang River Basin*, Master Thesis, Universiti Teknologi Malaysia.

Maidment, D.R. (Ed.) (1993). *Handbook of Hydrology*, McGraw-Hill Inc., New York.

Pilon, P.J. and Adomowski, K. (1992). "The Value of Regional Information to Flood Frequency Analysis Using the Method of L-Moments", *Can. J. Civ. Eng.*, 19 : 137-147.

Pilon, P.J., Adomowski, K. and Alila, Y. (1991). "Regional Analysis of Annual Maxima Precipitation Using L-Moments", *Atm. Res.*, 27 : 81-92.

Viessman, W., Lewis, G.L. and Knapp, J.W. (1989). *Introduction to Hydrology*, 3rd. Edition, Harper & Row, Singapore.

Vivekananda, M. (1981). *Analysis of Rainfall Characteristics for Hydrological Design*, Undergraduate Thesis, Universiti Teknologi Malaysia, Malaysia.

Vogel, R.M. and Fennessey, M. (1993). "L Moment Diagrams Should Replace Product Moment Diagrams", *Water Resour. Res.*, 29(6) : 1745-1752.

Yevjevich, V. (1972). *Probability and Statistics in Hydrology*, Water Resources Publications, Colorado, U.S.A.

APPENDIX A
DETAILS OF RAINFALL STATIONS IN CAMERON HIGHLANDS

No.	Station Name	Station No.	Location (DMS/MRSO)	Date Installed/ Closed	Operated By	Elevation Above MSL (m) #	Duration of Data Used
1.	Ladang Teh Blue Valley	9001	101 25 10 E, 04 35 10 N 380835.831 mE, 507436.129 mN	01/48	Estate	1450	1948-1993
2.	Pejabat TNB Kampung Raja	9002	101 25 00 E, 04 33 05 N 380516.340 mE, 503598.269 m N	01/62	TNB	1220	1962-1993
3.	Aturmasuk Sungai Telom	9003	101 25 30 E, 04 32 32 N 381437.857 m E, 502582.109 m N	02/64 / 05/91 05/91	TNB	1190	1964-1990
4.	Ladang Teh Sungai Palas	9004	101 25 00 E, 04 31 00 N 380505.009 mE, 499759.495 mN	01/54	Estate	1480	1954-1993
5.	Stesen Janaletrik Bintang	9006	101 25 30 E, 04 29 40 N 381422.337 m E, 497299.956 m N	02/54	TNB	1580	1954-1991
6.	Stesen MARDI Tanah Rata	9008	101 23 22 E, 04 27 54 N 377478.000 m E, 494054.000 m N	01/48	MARDI	1450	1948-1993
7.	Stesen Janaletrik Habu	9009	101 23 00 E, 04 25 00 N 376773.733 m E, 488714.759 m N	02/64	TNB	1120	1964-1993
8.	Pusat Peranginan Haiwan	9011	101 22 30 E, 04 26 45 N 375858.732 mE, 491942.125 mN	01/50	DID	1420	1950-1993
9.	Ladang Teh Shun Yip Leong	9012	101 21 00 E, 04 26 45 N 373084.804 m E, 491950.506 m N	01/55	Estate	1430	1955-1993
10.	Stesen Kajicuaca Tanah Rata	4413034	101 22 40 E, 04 28 25 N 376176.161 m E, 495012.240 m N	01/25	MMS	1460	1955-1992
11.	Ladang Boh (Kawasan Kilang)	4414036	101 25 27 E, 04 26 55 N 381315.006 m E, 492233.037 m N	11/47	Estate	1420	1947-1992
12.	Ladang Boh (Bahagian Boh)	4414037	101 25 38 E, 04 26 54 N 381653.948 m E, 492201.333 m N	04/48	Estate	1350	1948-1992
13.	Ladang Boh (Bahagian Selatan)	4414038	101 26 08 E, 04 26 53 N 382578.492 m E, 492167.920 m N	04/48	Estate	1210	1948-1992
14.	Stesen Telekom Gunung Brinchang	4513033	101 23 03 E, 04 31 05 N 376899.741 m E, 499923.768 m N	05/54 / 10/75 10/75	DID	2031	1975-1989

Estimated from topographic map

APPENDIX E
L-MOMENT RATIO DIAGRAMS

L-MOMENT RATIO DIAGRAM (1-DAY)

L-MOMENT RATIO DIAGRAM (1-DAY)

L-MOMENT RATIO DIAGRAM (2-DAY)

L-MOMENT RATIO DIAGRAM (2-DAY)

L-MOMENT RATIO DIAGRAM (3-DAY)

●	G (Gumbel)
□	E (Exponential)
△	N (Normal)
■	U (Uniform)
◇	L (Logistic)
——	GEV
—·—·	P3 @ GAM
—··—·	LN
········	GPA
— —	GLO
×	Region 3A
::	Region 3B
●	Region 3C

L-MOMENT RATIO DIAGRAM (3-DAY)

●	G (Gumbel)
□	E (Exponential)
△	N (Normal)
▣	U (Uniform)
◇	L (Logistic)
——	GEV
—·—·	P3 @ GAM
—··—·	LN
········	GPA
— —	GLO
●	Cameron Highlands

L-MOMENT RATIO DIAGRAM (5-DAY)

L-MOMENT RATIO DIAGRAM (5-DAY)

L-MOMENT RATIO DIAGRAM (7-DAY)

●	G (Gumbel)
▫	E (Exponential)
△	N (Normal)
▣	U (Uniform)
◇	L (Logistic)
——	GEV
———	P3 @ GAM
—·—	LN
······	GPA
— —	GLO
✕	Region 7A
∷	Region 7B
●	Region 7C

L-MOMENT RATIO DIAGRAM (7-DAY)

●	G (Gumbel)
▫	E (Exponential)
△	N (Normal)
▣	U (Uniform)
◇	L (Logistic)
——	GEV
———	P3 @ GAM
—·—	LN
······	GPA
— —	GLO
●	Cameron Highlands