Far East Journal of Mathematical Sciences (FJMS)

http://www.pphmj.com
http://dx.doi.org/10.17654/MS103081379
Volume 103, Number 8, 2018, Pages 1379-1398 **ISSN: 0972-0871**

REGIONAL FREQUENCY ANALYSIS ON PENINSULAR MALAYSIA USING *L*-MOMENTS

Sharainie Sahrin, Norazlina Ismail and Nor Eliza Alias

Faculty of Engineering and Built Environment
Universiti Kebangsaan Malaysia
Malaysia

Mathematics Department
Faculty of Science
Universiti Teknologi Malaysia
Malaysia

Department of Hydraulics and Hydrology
Faculty of Civil Engineering
Universiti Teknologi Malaysia
Malaysia

Abstract

Regional frequency analysis (RFA) is used to identify the homogeneous region using *L*-moments approach. At-site frequency analysis is unable to provide accurate results and gives erroneous results instead. One way to increase the precision of the data is by using RFA. The analysis combines data from sites with similar event frequencies into homogeneous regions. Cluster analysis is performed at first to determine the preliminary regions based on at-site characteristics. Then, it is tested using heterogeneity measure using

Received: March 29, 2017; Accepted: June 10, 2017

2010 Mathematics Subject Classification: 62P12.

Keywords and phrases: cluster analysis, homogeneous, regional frequency analysis, best-fitted distributions.

L-moments approach. The adjustment measures are taken to ensure that all regions satisfy the heterogeneity measure and clearly form the boundaries. The study area involving 83 rainfall stations of Peninsular Malaysia was separated into seven homogeneous regions. Each region was then evaluated using *L*-moments ratios to determine the best-fitted distribution. Three distribution functions were used which are generalized Pareto (GPA), generalized extreme value (GEV) and generalized logistic distributions (GLO). Results showed that GLO and GEV were the most suitable distributions for frequency analysis in Peninsular Malaysia.

1. Introduction

The annual maximum daily-rainfall is defined as an extreme instance with critical duration for a state or region with immediate consequences for agriculture, soil conservation, roads, dams and drainage (Willems et al. [40]). Information on the magnitudes and frequencies of extreme precipitations is essential for sustainable water resources management, planning weather-related emergencies, and designing hydraulic structures.

The frequency of heavy precipitation events increased over most areas and occurred especially in the late 20th century as reported by the Intergovernmental Panel on Climate Change (IPCC). In order to reduce the risk and maximize the efficiency in design, statistical and probabilistic methods are applied to past events to predict the exceedance probability of future events (Smithers and Schulze [32]). However, extreme events were rare and their records were often short, therefore, making an estimation of the frequencies of extreme events is difficult. Reliable estimations require very long station records if single station data are to be used. Therefore, regional frequency analyses are used to provide a framework for hazard characterization of the extreme events (Norbiato et al. [41]). Extreme-rainfall homogeneous regions refer to regions that contain sites with similar characteristics of extreme rainfall data such as means, skewness and kurtosis. This indicates that the areas within the homogeneous regions have similar conditions, climatic exposure, and source of extreme rainfalls.

L-moments method is the most popular method to identify these homogeneous regions as it uses an application that is statistically efficient and straightforward to implement (Hosking and Wallis [12]).

2. Data Screening

The study is focused on Peninsular Malaysia which has a total area of 131794km^2. The country's close proximity to the equator gives a hot and humid climate with daily temperature ranges from 25.5°C to 35°C. It is situated in the northern latitude between 1° and 6°N and the eastern longitude from 100° to 103° E. This study uses annual maximum daily-rainfall series to represent the extreme rainfalls. The data are obtained from 83 recording rain gauges located all over the Peninsular Malaysia and vary from 15 years to 38 years (1975-2012) relying on the establishment of the stations. The latitude and longitude of these stations are acquired from the Department of Irrigation and Drainage Malaysia.

For data screening, the discordancy measure *D* test is used to identify point sample *L*-moments that are marked differently from most of the other sites. It will determine any high discordant site (outlier). Sites with great errors in data will stand out from others and be flagged as discordant. The discordancy test, D_i, for site *i* is defined as

$$D_i = \frac{1}{3} N(u_i - \bar{u})^T S^{-1}(u_i - \bar{u}),$$

where u_i is a vector containing the three sample *L*-moment ratios for site *i*, *N* is the number of sites in the region, $\bar{u}$ represents the unweight regional average of *L*-moments ratio for each region

$$u_i = (t^{(i)} t_3^{(i)} t_4^{(i)})^T,$$

$$\bar{u} = N^{-1} \sum_{i=1}^{N} u_i,$$

and S is the sample covariance matrix expressed by

$$S = \sum_{i=1}^{N} (u_i - \bar{u})(u_i - \bar{u})^T.$$

Generally, a site is declared as discordant from the group if D_i value is greater than a critical value. $D_i \geq 3$ will be the critical value for $N \geq 15$ sites. If the D-statistic of a site exceeds 3, then the data is considered to be discordant from the rest of the regional data (Hosking and Wallis [12]).

3. Cluster Method

3.1. Ward's clustering method

Cluster analysis is a standard method of statistical multivariate analysis for dividing a data set into groups and successfully used to form regions for regional frequency analysis. Ward's method was selected for the determination of homogeneous regions in regional frequency analysis which is a hierarchical clustering method based on minimizing the Euclidean distance in site characteristics space within each cluster (Ramin [26]). Ward's hierarchical cluster method is also known as "minimum variance method". This method is distinct from other methods because it uses an analysis of variance approach to evaluate the distances between clusters. The Ward's clustering method is very efficient for clusters as membership is assessed by calculating the total sum of squared deviations from the mean of a cluster. The criterion for fusion is that it should produce the smallest possible increase in the error sum of squares (Satyvan and Sananse [29]).

The site characteristics used in this analysis were latitude, longitude, elevation, mean annual precipitation, and coefficient of variation. All the data were transformed to get the rescaled data range of between 0 and 1. Then, Ward's hierarchical clustering method was implemented.

3.2. Normalize data

Most clustering algorithms are very sensitive to the Euclidean distance or scale of the variables used in the analysis (Hosking and Wallis [12]).

Therefore, the variables with large absolute values are normalized so that their ranges are comparable. The variables are rescaled so that their values would lie between 0 and 1 to avoid dominance of site-characteristics (Malekinezhad and Zare-Garizi [17]):

$$X_{ij}^{N} = \frac{X_{ij} - X_{i,\min}}{X_{i,\max} - X_{i,\min}},$$

where

X_{ij} is the *i*th attribute of *j*th station,

$X_{i,\min}$ is the minimum *i*th attribute in all stations,

$X_{i,\max}$ is the maximum *i*th attribute in all stations,

X_{ij}^{N} is normalized *i*th attribute of *j*th station.

3.3. Determine distance measure

Distance is a measure of how far apart two objects are. For cases that are alike, the distance measures are small. In this Ward's clustering method, the standard version to form the distance measure is by using squared Euclidean distance (Lee and Willcox [15]). The formula of squared Euclidean distance is as follows:

$$\sum_{j=1}^{k} (a_j - b_j)^2,$$

where k denotes the number of variables, and a and b are two different clusters.

At every beginning step of the agglomerative hierarchical clustering process, each observation is its own cluster. The distance measure between two clusters which had only one case, respectively, was calculated using squared Euclidean distance. Every distance measure was the distance in the proximity matrix calculated using SPSS. As the data used 83 stations, the proximity matrix was a 83×83 matrix. The smallest distance between two clusters was merged together into a new cluster.

After forming new clusters with more than one case, the distance between pairs of clusters was defined by using linkage method that will be explained in the next step.

3.4. Linkage ward method

This method was assessed by calculating the total sum of squared deviations from the mean of a cluster. The number of clusters was reduced by one, by merging the two clusters that will produce the smallest possible increase in the error sum of squares (Satyvan and Sananse [29]). The same method was repeated by merging with other clusters until only one cluster was left. The error sum of squares is defined below:

$$ESS = \sum_i \sum_j \sum_k | X_{ijk} - \bar{x}_{i\bullet k} |^2,$$

where X_{ijk} denotes the value for variable k in observation j belonging to cluster i.

Here, the individual observations for each variable were compared against the cluster means for that variable. When the error sum of squares is small, this suggests that the data are close to their cluster means, implying that there is a cluster of like units.

3.5. Number of clusters

The appropriate number of groups (clusters) was determined using the silhouette widths (Borcard et al. [4]). The silhouette width is a measure of the degree of membership of an object to its cluster, based on the average distance between this object and all objects of the cluster to which it belongs, compared to the same measure computed for the next closest cluster (Borcard et al. [4]). The highest average silhouette width was selected as the optimum number of clusters. Silhouette widths were ranged from –1 to 1 (Borcard et al. [4]). The formula of silhouette width (SW_i) of every point is given below:

$$SW_i = \frac{(b_i - a_i)}{\max(a_i,\, b_i)},$$

where

a_i is the average distance from point i to all other points in the same cluster.

b_i is the minimum average distance from point i to all points in another cluster. The largest average silhouette width (*ASW*) value will be selected as the optimum number of cluster.

ASW was calculated using this formula:

$$ASW = \frac{1}{n_i}\sum_{i=1}^{n} SW_i.$$

4. Regional Homogeneity Test

Homogeneity tests compare the inter-site variations in sample *L*-moments for the group of sites with what would be expected of a homogeneous region. This test fits a four-parameter kappa distribution to the regional data set (Hosking [11]). Then, it will generate 500 equivalent region's data by Monte Carlo simulation. The variability of the statistics of the actual region is compared to the simulated series. The heterogeneity test is computed as

$$H_1 = \frac{(V - \mu_V)}{\sigma_V},$$

where μ_V and σ_V represent the population mean and standard deviation of simulated *V* values, respectively,

$$V = \left(\sum_{i=1}^{N} n_i (t^{(i)} - t^R)^2 \Big/ \sum_{i=1}^{N} n_i \right)^{1/2}.$$

Alternatively, dispersion measure could also be used based on *L*-skewness (t_3) and *L*-kurtosis (t_4) using these equations:

$$H_j = \frac{(V_j - \mu_{V_j})}{\sigma_{V_j}}, \quad j = 2, 3,$$

where

$$V_2 = \left(\sum_{i=1}^{N} n_i ((t^{(i)} - t^R)^2 + (t_3^{(i)} - t_3^R)^2)^{1/2} \middle/ \sum_{i=1}^{N} n_i \right),$$

$$V_3 = \left(\sum_{i=1}^{N} n_i ((t_3^{(i)} - t_3^R)^2 + (t_4^{(i)} - t_4^R)^2)^{1/2} \middle/ \sum_{i=1}^{N} n_i \right).$$

However, H_1 was recommended to be used for checking homogeneity since H_2 and H_3 can give false indication of homogeneity. They lack the power to discriminate between homogeneous and heterogeneous regions. They rarely yield H values larger than 2 even for grossly heterogeneous regions. Thus, the H_1 statistics has much better discriminatory power (Hosking and Wallis [12]).

Denote that the proposed region has N sites, with site i having size n_i and sample L-moment ratios $t^{(i)}$ (L-CV of i site), $t_3^{(i)}$ (L-skewness of i site), and $t_4^{(i)}$ (L-kurtosis of i site). Here, t^R, t_3^R and t_4^R are the regional average L-moments ratios (regional average L-CV, regional average L-skewness and regional average L-kurtosis) calculated using the following formula:

$$t_j^R = \frac{\sum_{i=1}^{N} n_i t_j^{(i)}}{\sum_{i=1}^{N} n_i}.$$

To test whether the region is homogeneous, the regional averaged L-moment ratios $(1, t^R, t_3^R$ and $t_4^R)$ are used to fit four-parameter kappa distributions suggested by Hosking and Wallis (Shahzadi et al. [42]). Large number of simulations N_{sim} is generated using Monte Carlo simulation. The simulated regions have sites with similar record length as the real samples. V is calculated for each simulated region. Then, the mean and standard deviations of the N_{sim} of V are determined. They are μ_V and σ_V, respectively. Following Hosking and Wallis [12], the region under analysis can be regarded as: (i) if $H < 1,$ then the region is “acceptably homogeneous”, (ii) if $1 \leq H < 2,$ then the region is “possibly heterogeneous”, and (iii) if $H \geq 2,$ then the region is “definitely heterogeneous”.

5. Refinement of Regions

Adjustment of the regions can be done based on Hosking and Wallis [12] in order to obtain “acceptably homogeneous” classification for all regions. The adjustment options are as the following: (i) dividing a region to form two or more new regions, (ii) transferring one or more sites from a region to others, or (iii) eliminating or deleting one or more sites from the data set.

6. Goodness-of-fit Measure

Z^{DIST} is used to establish the best-fitted distribution for each region. In this study, three distributions which are generalized logistic (GLO), generalized Pareto (GPA) and generalized value (GEV) were used to develop the regional frequency analysis procedures (Zakaria et al. [38]). The goodness-of-fit test measures how well the theoretical *L*-kurtosis of a fitted distribution matches the regional average *L*-kurtosis of the observed data. Z^{DIST} has been used to determine the best distribution for each region (Hosking and Wallis [12]):

$$Z^{DIST} = \frac{t_4^{DIST} - t_4^R}{\sigma^4},$$

where

$$t_4^R = \frac{\sum_{i=1}^{N} n_i t_4^{(i)}}{\sum_{i=1}^{N} n_i},$$

$$\sigma_4 = \sqrt{\frac{\sum_{m=1}^{N_{sim}} (t_4^{(m)} - t_4^R)^2}{N_{sim} - 1}}.$$

σ_4 can be obtained by repeated simulations of a kappa distribution. The simulations were exactly like the one that was used in the calculation of the heterogeneity measure. t_4^{DIST} is the kurtosis value of a distribution function in interest. It has been constructed in the form of

$$t_4^{DIST} = \sum_{k=0}^{8} A_k (t_3^R)^k.$$

A calculated value of zero for $|Z^{DIST}|$ indicates a perfect fit. The Z-statistics is considered to be acceptable at 90% confidence level if $|Z^{DIST}| \leq 1.64$. If more than one distribution is acceptable, then the one with the lowest $|Z^{DIST}|$ is chosen as the best-fit distribution.

7. Results and Discussion

7.1. Discordancy measure and cluster analysis

All the discordancy values for every site in each region are less than 3 meaning that there is no outlier in this analysis. The appropriate number of groups (clusters) was determined using the silhouette method (Malekinezhad and Zare-Garizi [17]). The optimal number of clusters for this analysis was seven, which has the largest value of 0.4052. The heterogeneity measures based on cluster method are stated below.

Table 1. Heterogeneity measures based on cluster method

Region	Heterogeneity measures		
	H_1	H_2	H_3
1	0.72	0.15	0.36
2	1.69	1.88	2.65
3	–0.35	2.13	2.02
4	0.62	1.84	1.55
5	0.77	1.85	2.34
6	–0.20	1.22	0.69
7	0.63	1.43	2.25

From Table 1, results show that Region 2 was "possibly heterogeneous" because its value was between 1 and 2. Regions 1, 3, 4, 5, 6 and 7 were "acceptably homogeneous" as their values were less than 1.

7.2. Refinement of regions

From Figure 1, there are some sites which are homogeneous based on the cluster technique but the boundaries of the cluster cannot be defined. Some sites of Regions 3 and 4 were overlapping whereby five sites of Region 3 were in Region 4. Therefore, it is difficult to form two boundaries. In addition, Table 2 shows that Region 2 is possibly heterogeneous. Adjustment options should be implemented for all regions to be considered as "acceptably homogeneous" so that the partitions between the regions can be easily formed as well.

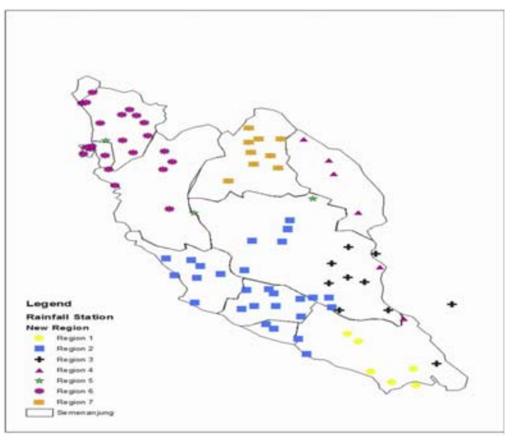


Figure 1. The homogeneous regions based on cluster analysis.

Based on Figure 1, there are a few regions that remained homogeneous such as Region 6 which is located at the north area of Peninsular Malaysia, comprising the states of Kedah, Perak, Pulau Pinang and Perlis. All three sites of Region 5 also satisfied the homogeneity test. All nine sites from Region 7 which are located at the Kelantan state remained homogeneous. All six sites from Region 1 located at the Johor state also fulfilled the homogeneity test.

There are some adjustments that need to be taken for several particular regions which are Regions 2, 3 and 4. Five sites of Region 3 were transferred to Region 4. Thus, this will form a region located in the eastern side of Peninsular Malaysia. Based on the literature by Zalina et al. [39] and Wong et al. [37], sites that are located in the eastern side experience the same climatic, geographic, and topographic characteristics.

Region 2 is divided into two regions which are 18 sites for the west coast area and nine sites for Pahang state. Then, nine sites of Region 2 which are in the Pahang state were move to Region 3 that is also accumulated in Pahang state, and combined to form one region. Based on Lim [16], the Pahang state has similarity in rainfall pattern because it is located at the middle of Peninsular Malaysia called the *Central Interior*. The new homogeneous regions are shown in Figure 2.

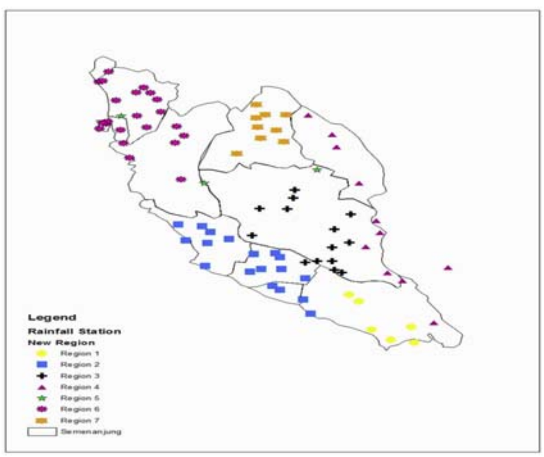


Figure 2. Homogeneous region after adjustment.

These are the heterogeneity measures for every region after adjustment.

Table 2. Heterogeneity measures after adjustment

Region	Heterogeneity measures		
	H_1	H_2	H_3
1	0.72	0.15	0.36
2	0.78	1.27	1.79
3	0.83	2.15	2.81
4	0.35	2.02	1.60
5	0.77	1.85	2.34
6	–0.20	1.22	0.69
7	0.63	1.43	2.25

From the table above, the values of heterogeneity measures (H_1) for all regions were less than 1. It means that all these regions are "acceptably homogeneous". After the adjustment options have been implemented, all the seven regions formed now satisfy the homogeneity test.

Figure 2 shows the boundaries of homogeneous regions after the adjustments were made. Seven homogeneous regions were identified. Even though the number of homogeneous regions obtained is seven which is more than the results in previous studies by Lim [16], Zalina et al. [39] and Baki et al. [20], it is more comprehensive because the mathematical method (cluster technique) was used to determine the preliminary regions instead of merely assuming the preliminary regions based on physical characteristics as was conducted in previous studies. Another variable was also added in this analysis which is elevation.

Local climates are also affected by the presence of mountain ranges particularly, the Titiwangsa Range throughout Peninsular Malaysia. Region 1 is located at the southern area of Malaysia which is in the Johor state based on topographic characteristics (Wong et al. [37]). Topographic characteristics refer to the surface features of land, which include the mountains, hills and creeks. The Titiwangsa Range divides Peninsular Malaysia between its east and west coasts. Region 2 is situated at the west coast area whereas Region 4 is located at the east coast area (sea coast region). The northeast monsoon brings heavy rainfall to the east coast of Peninsular Malaysia and more rainfall compared to the southwest monsoon. During the northeast monsoon period, rain can sometimes fall for days, but places on the western side of Peninsular Malaysia are not affected by this monsoon because they are sheltered by the Titiwangsa Range.

Sites in Region 6 located at the northern area of Peninsular Malaysia experience the largest daily variation of relative humidity. They also experience similar climatic aspects. Climate is the weather conditions of a region such as precipitation, temperature, humidity and winds (Baki et al. [21]). Region 3 which is accumulated at the Pahang state has different event frequencies from the coastal area because it is located in the middle of Peninsular Malaysia or the Central Interior (Lim [16]). Region 7 which is accumulated at the Kelantan state has a higher elevation in the interior parts of the north-eastern coast of Peninsular Malaysia, playing a crucial role in providing additional lifting and enhancing convection and rainfall (Juneng et al. [13]). Sites located at a high altitude (mountains) are more than 1,000 metres in height in Region 5. The rainfall stations are Gunung Berinchang and Gunung Gagauin Pahang, and the Kedah Peak in Kedah. The highlands are cooler and wetter because of the temperatures which are usually noticeably lower.

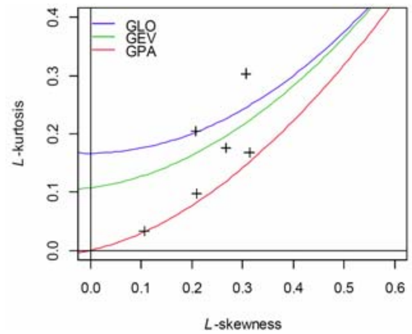
7.3. Goodness-of-fit test

The weighted mean of *L*-skewness and *L*-kurtosis are plotted into the *L*-moment ratio diagram. The diagrams are plotted in Figure 3 in order to view the appropriate distributions as guidance through graphical instrument by comparing the fit of several distributions which are GPA, GEV and GLO distributions to many samples of data in a group of regions (Baki et al. [21]). Further analysis to confirm the true distribution was carried out by using goodness-of-fit test.

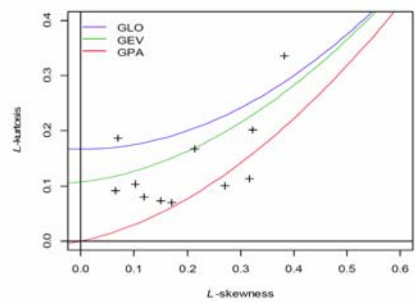
Based on Figures 3 and 4, the sample points should be distributed above and below the theoretical line of a suitable distribution. From the above *L*-moment diagrams, it appears that the GEV distribution is suitable for

Regions 1, 4 and 6, whereas GLO distribution is fitted to Regions 2, 3, 5 and 7. Further analysis was used to confirm the true distribution by using the goodness-of-fit test.

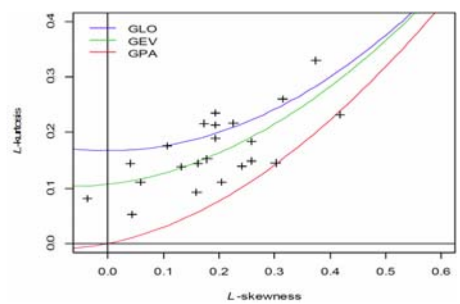
The Z-statistics is considered to be acceptable at 90% confidence level if $|Z^{DIST}| \leq 1.64$. If more than one distribution is acceptable, the one with the lowest $|Z^{DIST}|$ is chosen as the best-fit distribution. The table below shows the Z^{DIST} values for every region.



Region 1

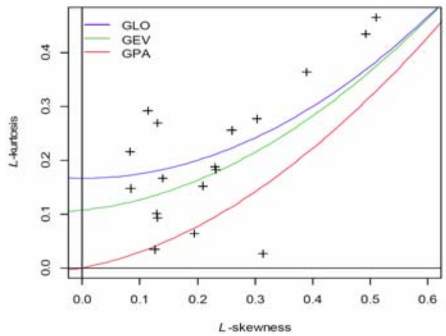


Region 4

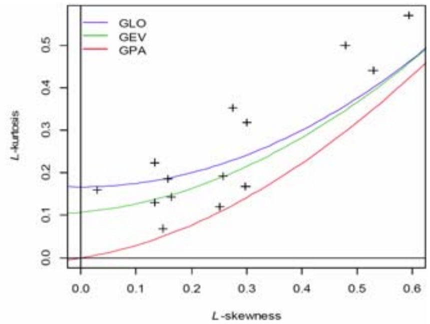


Region 6

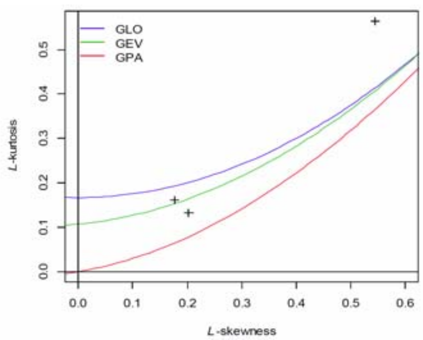
Figure 3. *L*-moments ratio diagram (a).



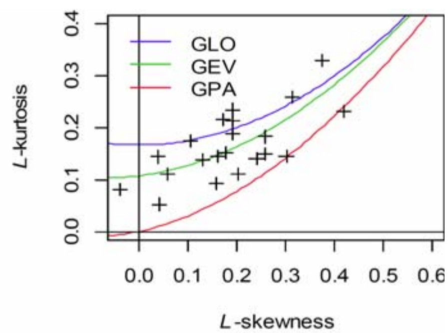
Region 2



Region 3



Region 7



Region 5

Figure 4. *L*-moments ratio diagram (b).

Table 3. Goodness-of-fit test for eight regions

Distribution	R1	R2	R3	R4	R5	R6	R7
GLO	1.53	–1.07	–2.62	3.17	–1.77	1.61	–0.91
GEV	0.41	–2.53	–3.82	1.26	–2.18	–0.88	–1.58
GPA	–2.30	–6.04	–6.83	–3.08	–3.33	–6.49	–3.48

From Table 3, results show that for Region 1, Region 4 and Region 6 revealed that the GEV distribution provides the best approximation of annual maximum daily-rainfall while for Region 2, Region 3, Region 5 and Region 7, the GLO distribution appears to be the best-fitted distribution.

8. Conclusion

Ward's hierarchical clustering method is used in order to determine the preliminary regions. Using this cluster technique, Peninsular Malaysia can be divided into 7 regions based on at-site characteristics. Then, they were tested using homogeneity test. Only six of them satisfied the homogeneity test. In order to ensure all the regions are acceptably homogeneous, the adjustment options are taken which are dividing a region to form two or more regions and transferring site to the other region. These newly formed regions are tested again using heterogeneity test. This study successfully formed 7 homogeneous regions. Even though the number of homogeneous regions obtained is seven which is more than the previous studies results such as Lim [16], Zalina et al. [39] and Baki et al. [20] but it is more comprehensive because mathematical method (cluster technique) is used to determine the preliminary regions instead of just assuming the preliminary regions based on physical characteristics like previous studies have done. This research also added one more variable in this analysis which is elevation. The results satisfy the condition made by previous researchers indicating that GLO and GEV are the most suitable distributions for frequency analysis in Peninsular Malaysia.

Acknowledgments

The authors are grateful to Department of Irrigation and Drainage Malaysia for providing the data conducted for this study. Gratitude is also for JASTIP fund support through grant number R.J130000.7322.4B298 as registered by UTM research management center (RMC). The JASTIP research was supported by Japan Science and Technology Agency (JST), Collaboration Hubs for International Research Program (CHIRP) within the framework of the Strategic International Collaborative Research Program (SICORP).

References

[1] S. Ahmad, M. H. Noorazuan and M. J. Yaakob, FenomenaPulauHaba Bandar danisualamsekitar di Bandaraya Kuala Lumpur, Geografia: Malaysian Journal of Society and Space 5(3) (2009), 57-67.

[2] K. Amin, Spatio-temporal variability of seasonal rainfall in western equatorial Africa, Theor. Appl. Climatol. 104 (2011), 57-69.

[3] D. Arti and C. Parthasarthi, Extreme rainfall frequency analysis for meteorological sub-division 4 of India using l-moments, International Journal of Environmental, Chemical, Ecological, Geological and Geophysical Engineering 7(12) (2013), 945-950.

[4] D. Borcard, F. Gillet and P. Legendre, Numerical Ecology with R, Springer, New York, Dordrecht, London, Heidelberg, 2011.

[5] W. L. Dale, The rainfall of Malaya, Part I, J. Trop. Geogr. 13 (1959), 23-37.

[6] F. Dikbas, F. Mahmut, K. Cem and M. Gungor, Classification of precipitation series using fuzzy cluster method, Int. J. Climatol. 32 (2012), 1596-1603.

[7] S. R. Durrans and J. Kirby, Regionalization of extreme precipitation estimates for the Alabama rainfall atlas, J. Hydrol. 295 (2004), 101-107.

[8] J. Greenwood, J. Landwehr, N. Matalas and J. Wallis, Probability-weighted moments: definition and relation to parameters of several distributions expressable in inverse form, Water Resource Res. 15 (1979), 1049-1054.

[9] B. Hassan, I. Atiem, L. Jianzhu and F. Ping, At site and regional frequency analysis for Sudan annual rainfall by using the *L*-moments and nonlinear regression techniques, International Journal of Engineering Research and Development 3(6) (2012), 13-19.

[10] J. R. M. Hosking and J. R. Wallis, Some statistics useful in regional frequency analysis, Water Resour. Res. 29 (1993), 271-281.

[11] J. R. M. Hosking, The four-parameter kappa distribution, IBM J. Res. Develop. 38(3) (1994), 251-258.

[12] J. R. M. Hosking and J. R. Wallis, Regional Frequency Analysis, Cambridge University Press, United Kingdom, 1997.

[13] L. Juneng, F. T. Tangang and C. J. C. Reason, Numerical case study of an extreme rainfall event during 9-11 December 2004 over the east coast of peninsular Malaysia, Meteorol. Atmos. Phys. 98 (2007), 81-98.

[14] M. G. Kendall, Rank Correlation Methods, 4th ed., London: Charles Griffin. 1975.

[15] A. Lee and B. Willcox, Minkowski generalizations of Ward's method in hierarchical clustering, Journal of Classification 31(2) (2014), 194-218.

[16] J. T. Lim, Rainfall minimum in peninsular Malaysia during the northeast monsoon, Mon. Weather Rev. 104 (1976), 96-99.

[17] H. Malekinezhad and A. Zare-Garizi, Regional frequency analysis of daily rainfall extremes using *L*-moments approach, Atmósfera 27(4) (2014), 411-427.

[18] K. G. Manish and G. Vivek, Identification of homogeneous rainfall regimes in northeast regions of India using fuzzy cluster analysis, Water Resour. Manage. 28 (2014), 4491-4511.

[19] R. Modarres, Regional Precipitation climates of Iran, Journal of Hydrology 45(1) (2006), 13-27.

[20] A. Mohd Baki, I. Atan and N. F. Mohd Halim, Frequency analysis of annual runoff in Peninsular Malaysia using a regional approach based on *L*-moments, Journal-The Institution of Engineers, Malaysia 73(3) (2012), 33-38.

[21] A. Mohd Baki, D. A. Mat Yusof, I. Atan and N. F. Mohd Halim, Regional flow frequency analysis on Peninsular Malaysia using *L*-moments, Jurnal Intelek 9(1) (2014), 63-68.

[22] R. P. C. Morgan, Rainfall of West Malaysia: a preliminary regionalization using principal components analysis, The Royal Geographical Society 3(4) (1971), 222-227.

[23] C. Ngongondo, C.-Y. Xu, L. Tallaksen, B. Alemaw and T. Chirwa, Regional frequency analysis of rainfall extremes in Southern Malawi using the index rainfall and *L*-moments approaches, Stoch. Environ. Res. Risk Assess. 25(7) (2011), 939-955.

[24] J. S. Park and H. S. Jung, Modelling Korea extreme rainfall using a kappa distribution and maximum likelihood estimate, Theor. Appl. Climatol. 72 (2002), 55-64.

[25] A. N. Pettitt, A two-sample Anderson-Darling rank statistic, Biometrika 63 (1976), 161-168.

[26] R. Ramin, Regional flood frequency analysis based on *L*-moment approach (case study west Azarbayjan basins), Journal of Civil Engineering and Urbanism 3(3) (2013), 107-113.

[27] R. Rao and V. Srinivas, Regionalisation of watersheds by hybrid-cluster analysis, J. Hydrol. 318 (2006), 37-56.

[28] P. Satyanarayana and V. V. Srinivas, Regional frequency analysis of precipitation using large-scale atmospheric variables, Journal of Geophysical Research Atmospheres 113 (2008) (No. D24).

[29] Y. Satyvan and S. L. Sananse, Comparisons of different methods of cluster analysis with application to rainfall data, International Journal of Innovative Research in Science, Engineering and Technology 4(11) (2015), 10861-10872.

[30] F. W. Scholz and M. A. Stephens, *K*-sample Anderson-darling tests, J. Amer. Statist. Assoc. 82(399) (1987), 918-924.

[31] A. Shabri, Z. M. Daud and N. M. Ariff, Regional analysis of annual maximum rainfall using TL-moments method, Theor. Appl. Climatol. 104(3-4) (2011), 561-570.

[32] J. C. Smithers and R. E. Schulze, A methodology for the estimation of short duration design storms in South Africa using a regional approach based on *L*-moments, Journal of Hydrology 241 (2001), 42-52.

[33] J. Suhaila, S. M. Deni, W. Z. W. Zin and A. A. Jemain, Trends in peninsular Malaysia, rainfall data during the southwest monsoon and northeast monsoon seasons: 1975-2004, Sains Malaysiana 39(4) (2010), 533-554.

[34] T. H. Teklu, T. T. Sveinn and A. Knut, Regional frequency analysis of extreme precipitation with consideration of uncertainties to update IDF curves for the city of Trondheim, Journal of Hydrology 498 (2013), 305-318.

[35] A. Viglione, F. Laio and P. Claps, A comparison of homogeneity tests for regional frequency analysis, Water Resour. Res. 43(3) (2007) (No. W03428).

[36] W. Z. Wan Zawiah, J. Abdul Aziz and I. Kamarulzaman, The best fitting distribution of annual maximum rainfall in peninsular Malaysia based on methods of *L*-moments and *LQ*-moment, Theor. Appl. Climatol. 96 (2008), 337-344.

[37] C. L. Wong, J. Liew, Z. Yusop, T. Ismail, R. Venneker and S. Uhlenbrook, Rainfall characteristics and regionalization in peninsular Malaysia based on a high resolution gridded data set, Water 8 (2016), 500.

[38] Z. A. Zakaria, A. Shabri and U. N. Ahmad, Regional frequency analysis of extreme rainfalls in the West Coast of Peninsular Malaysia using partial *L*-moments, Water Resour. Manage. 26 (2012), 4417-4433.

[39] M. D. Zalina, M. K. Amir Hashim, M. D. Mohd Nor and V. N. Van Thanh, Statistical analysis of at-site extreme rainfall processes in Peninsular Malaysia, Regional Hydrology 274 (2002), 61-68.

[40] P. Willems, K. Arnbjerg-Nielsen, J. Olsson and V. T. V. Nguyen, Climate change impact assessment on urban rainfall extremes and urban drainage: method and shortcomings, Atmos. Res. 103 (2012), 106-118.

[41] D. Norbiato, M. Borga, M. Sangati and F. Zanon, Regional frequency analysis of extreme precipitation in the eastern Italian Alps and the August 29, 2003 flash food, J. Hydrol. 345 (2007), 149-166.

[42] A. Shahzadi, A. S. Akter and B. Saf, Regional frequency analysis of annual maximum rainfall in monsoon region of Pakistan using *L*-moments, Pak. J. Stat. Oper. Res. 9(1) (2013), 111-136.