

Fitting statistical distribution of extreme rainfall data for the purpose of simulation

S.M. Shaharudin¹, N. Ahmad², N.S. Mohamed³, Hairulnizam Mahdin⁴

¹Department of Mathematics, Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris, Malaysia

²Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, Malaysia

³Technical Foundation, Malaysian Institute of Industrial Technology, Universiti Kuala Lumpur, Malaysia

⁴Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia

Article Info

Article history:

Received Sep 11, 2019

Revised Dec 12, 2019

Accepted Dec 30, 2019

Keywords:

Anderson darling test

Chi-square test

Daily torrential rainfall amount

Gamma distribution

Generalized pareto distribution

Lognormal distribution

Simulation

ABSTRACT

In this study, several types of probability distributions were used to fit the daily torrential rainfall data from 15 monitoring stations of Peninsular Malaysia from the period of 1975 to 2007. The study of fitting statistical distribution is important to find the most suitable model that could anticipate extreme events of certain natural phenomena such as flood and tsunamis. The aim of the study is to determine which distribution fits well with the daily torrential Malaysian rainfall data. Generalized Pareto, Lognormal and Gamma distributions were the distributions that had been tested to fit the daily torrential rainfall amount in Peninsular Malaysia. First, the appropriate distribution of the daily torrential rainfall was identified within the selected distributions for rainfall stations. Then, data sets were generated based on probability distributions that mimic a daily torrential rainfall data. Graphical representation and goodness of fit tests were used in finding the best fit model. The Generalized Pareto was found to be the most appropriate distribution in describing the daily torrential rainfall amounts of Peninsular Malaysia. The outputs can be beneficial for the purpose of generating several sets of simulated data matrices that mimic the same characteristics of rainfall data in order to assess the performance of the modification method compared to classical method.

Copyright © 2020 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

S.M. Shaharudin,

Department of Mathematics, Faculty of Science and Mathematics,

Universiti Pendidikan Sultan Idris,

35900 Tanjung Malim, Perak, Malaysia.

Email: shazlyn@fsm.upsi.edu.my

1. INTRODUCTION

Fitting distributions used to find a statistical distribution that best fits a data set are generated by random processes. A common problem that can be found in fitting distribution is to identify which distribution is suitable to a set of observations for a variable. One does this to be able to simulate data based on the chosen distribution. It is important to identify the best fitting distribution because inappropriate distribution will affect the generated data and onwards will certainly result in wrong decisions [1, 2].

In Malaysia, many studies had investigated on fitting the distribution of rainfall, either hourly, daily or annually [3-5]. For instance, [6] showed that the most appropriate distribution for hourly rainfall data is Mixed-Exponential distribution for one particular area, Wilayah Persekutuan. In [7] had made a report on exploring the statistical probability distributions namely, Generalized Pareto, Exponential, Beta and Gamma distributions of rainfall intensity for Peninsular Malaysia using hourly rainfall data. The results showed that the Generalized Pareto distribution is the best fit for the hourly rainfall data among the four probability distributions tested. Meanwhile, in the study of [8], Poisson Gamma distribution was used to model the daily

rainfall of four stations in Peninsular Malaysia while the study in [9] used Mixed Exponential and mixed lognormal distributions on the daily rainfall amount which covered Peninsular Malaysia area for the purpose of describing the daily rainfall patterns. However, little work had been carried out to choose the best fit distribution for annually maximum rainfall data in Peninsular Malaysia [10, 11]. Based on these studies, Generalized Extreme Value distribution seems more suitable in describing annually maximum rainfall data while Generalized Logistic and Generalized Pareto distributions are the most suitable for the annual data of maximum daily rainfall in Peninsular Malaysia.

These studies, however, did not aim to find the distribution based on the characteristics of rainfall data in Peninsular Malaysia. In analysing characteristics of rainfall data in Peninsular Malaysia, it is important to realize that the daily rainfall are inherently skewed, usually to the right as such data only take positive values and tend to be skewed towards higher values. Besides that, even if Peninsular Malaysia has two dry wet seasons, the amount of rainfall does not significantly vary much as compared to that in the four season regions [12]. Thus, distributions that exhibit these characteristics can be used to model the torrential rainfall.

Three distributions chosen which are gamma, Log-Normal and Generalized Pareto distribution (GPD) are tested on daily torrential rainfall data. These distributions are commonly used as potential candidates for the data generating mechanism of rainfall data [13]. The adequacy of the fitted model is validated through observing graphical representation and goodness of fit tests.

In Section 2, the rainfall data that used in this study is described. Then, in Section 3, the methodologies related to the fitting distribution and simulation data is presented. Next, in Section 4, the results and discussion for fitting statistical distribution of extreme rainfall is provided. Conclusion is given in the final section.

2. RAINFALL DATA

Daily rainfall data were obtained from Jabatan Pengairan dan Saliran (JPS) from the period of 1975 to 2007 over the 75 rain gauge stations in Peninsular Malaysia. In this study, the focus is on the occurrence of extreme rainfall event described as torrential rainfall. It is therefore necessary to choose some criteria that would lead to the establishment of a threshold in order to allow for a clear distinction between what constitutes a day of torrential rainfall in the Peninsular Malaysia region and what does not. The range of threshold for torrential rainfall data in Peninsular Malaysia is 60 mm/day [14]. This threshold is chosen based on the categorization of rainfall intensity by Jabatan Pengairan dan Saliran (JPS). By filtering days with rainfall more than 60 mm for at least 2% of overall stations, characteristics rainfall patterns were derived for the 250 torrential events which occurred during the 33-year period.

3. RESEARCH METHOD

3.1. Probability Distribution of Rainfall Data

The distributions of tropical rainfall data are generally skewed to the right and thus distributions that exhibit this characteristic can be used to model the torrential rainfall. For the rainfall data, the potential distributions are Generalized Pareto distribution, Log-Normal distribution and Gamma distribution as suggested by [15-17].

3.2. Generalized Pareto Distribution

Generalized Pareto distribution (GPD) is the distribution which is appropriate when the data consists of a set of maxima. This is a strong evidence why rainfall data fits the GPD due to the occurrence of episodes on extreme rainfall event described as torrential rainfall. GPD is heavy tailed, skewed and essentially, the model is used for extreme values as examined by [18].

The probability density function of Generalized Pareto distribution:

$$P(x) = \frac{1}{\sigma} \left(1 - \frac{\xi(x-\mu)}{\sigma} \right)^{\frac{\xi+1}{\xi}} \quad (1)$$

The GPD is characterized by ξ , a shape parameter in which the domain is negative infinity to positive infinity. σ is a scale parameter which measures the heaviness in the upper tail. μ is a location parameter that can be explained as the threshold. The cumulative distribution function of GPD is defined as,

$$F(x) = 1 - \left(\frac{\xi(x-\mu)}{\sigma} - 1 \right)^{\frac{1}{\xi}} \quad (2)$$

3.3. Log-Normal Distribution

Let W be a normal distributed random variable and X is assumed to follow a Log-Normal distribution if there exists a relationship between X and W such that $\ln(\mathbf{X}) = W$. Log-Normal distribution follows an asymmetrical and positively skewed behavior which has two parameters that are obtained by:

$$f(x; \theta, \omega) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln x - \theta)^2}{2\omega^2}\right], 0 < x < \infty \tag{3}$$

where θ and ω^2 are the mean and variance of normal distribution of W . The mean (μ) and variance (σ^2) of \mathbf{X} can be expressed as,

$$\mu = e^{\theta + \omega^2/2} \text{ and } \sigma^2 = e^{2\theta + \omega^2}(e^{\omega^2} - 1) \tag{4}$$

3.4. Gamma Distribution

Gamma distribution is widely used in hydrological data especially rainfall data. The probability density function of Gamma distribution is defined as,

$$f(x) = \frac{(x-\gamma)^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} \exp(-(x-\gamma)/\beta) \tag{5}$$

and cumulative density function of Gamma distribution as,

$$F(x) = \frac{\Gamma((x-\gamma)/\beta, \alpha)}{\Gamma(\alpha)} \tag{6}$$

where: α : shape parameter ($\alpha > 0$), β : scale parameter ($\beta > 0$), γ : location parameter ($\gamma = 0$ yields the two parameter Gamma distribution), $\Gamma(\alpha)$: Gamma function

3.5. Fitting Distribution

Fitting distributions are crucial in finding and selecting a statistical distribution that best fits a data set generated by random processes. A common problem in fitting distribution is to identify which distribution is fit to a set of observations for a variable. It is important to identify the best fitting distribution because inappropriate distribution will affect the generated data and onwards may develop wrong results. It has some guidelines in determining the best fitting distribution to a data set which is based on distribution graphs and goodness of fit tests.

3.6. Distribution Graph

QQ plot is a plot of the original data values plotted against the fitted distribution quantiles. The function of the QQ plot is to identify how well a specific distribution fits the observed data. If the observations are approximately linear, the resulting plot is the best fit distribution for the sample data.

The QQ plot is calculated as follow:

$$F^{-1}\left(F_n(x_i) - \frac{0.5}{n}\right) \tag{7}$$

where $F^{-1}(x)$: inverse cumulative distribution function, $F_n(x)$: empirical cumulative distribution function (CDF), n : sample size.

Apart from that, probability difference graph is one of the measures used to identify the best fit distribution for sample data set. Probability difference graph is a plot difference between the empirical cumulative distribution function and the fitted cumulative distribution function. The graph is calculated as follows:

$$Diff(x) = F_n(x) - F(x) \tag{8}$$

where $F(x)$ refers to CDF. The purpose of using this graph is to determine how well the theoretical distribution fits the observed data and to compare the goodness of fit of several distributions.

3.7. Goodness of Fit Tests

The Chi-Square test is one of the goodness of fit tests that measures how well the expected frequency of the fitted distribution compared to the observed frequency of a histogram of the observed data.

The probability distribution of chi-square (χ^2) test is given by

$$f(\chi^2) = \frac{1}{2^{v/2}\Gamma(v/2)} e^{-\chi^2/2} (\chi^2)^{(v/2)-1} \quad (9)$$

This test is known as the (χ^2) distribution with the number of degrees of freedom, v with a positive integer. $f(\chi^2)d(\chi^2)$ is the probability that a particular value of χ^2 falls between the range of χ^2 and $\chi^2 + d(\chi^2)$. The range of χ^2 is only positive values which is $0 < \chi^2 < \infty$.

Another goodness of fit test that is used in this study is Anderson-Darling where the function is to verify that a sample comes from a population with specific distribution. There are several steps when using Anderson-Darling method to test the fit of the distribution:

- Arrange the sample data set in increasing order: x_1, x_2, \dots, x_n where n is the observation of the sample data set.
- Evaluate the cumulative distribution function of the proposed distribution $F_x(x_i)$ at x_i , for $i = 1, 2, \dots, n$.
- Calculate the Anderson-Darling statistic.

$$A^2 = -\sum_{i=1}^n [(2i-1)\{\ln F_x(x_i) + \ln[1 - F_x(x_{n+1-i})\}]/n] - n \quad (10)$$

- Compute the adjusted test statistic A^* to account for the effect of sample size n .

$$A^* = A^2 \left(1.0 + \frac{0.2}{\sqrt{n}}\right) \quad (11)$$

- Choose a significance level, α and determine the corresponding critical value, c_α for the suitable distribution type.
- Based on the selected distribution, compare A^* with the appropriate c_α . If A^* is less than c_α , the selected distribution is acceptable at the significance level, α .

3.8. Simulation Procedure for Generating Synthetic Rainfall Data

Data sets are generated based on probability distributions that mimic a multivariate torrential rainfall data. The distributions of tropical rainfall data are generally skewed to the right and thus distributions that exhibit this characteristic can be used to model the torrential rainfall [19]. Three distributions are chosen which are gamma, Log-Normal and Generalized Pareto distributions (GPD) are tested on multivariate rainfall data. These distributions are commonly used as potential candidates for the data generating mechanism of rainfall data [20-22]. Estimation of the parameters for each of the above sampled probability distributions are based on the summary statistics from the daily torrential rainfall data in Peninsular Malaysia. Out of the three probability distributions which were sampled, GPD appears to fit the data set based on several assessments by distribution graphs and goodness of fit tests using Chi-square and Anderson Darling test. This distribution is remarkably good at significance level $\alpha = 0.01$ if $p - value > \alpha$, thus providing some evidence that the null hypothesis is true (i.e the GPD provides the correct statistical model for rainfall data).

Simulations were carried out on sample GPD distributions characterized by three parameters location (μ), scale (σ) and shape (ξ), obtained from the original torrential rainfall data of 33 years period in Peninsular Malaysia to construct an $n \times p$ matrix with $n = 250$ and $p = 15$ to represent 250 torrential rainfall days and 15 rainfall stations respectively. In order to vary the simulation tested, two different settings were used. Firstly, the scales (i.e. standard deviation) were varied below and above standard deviation of the original torrential rainfall data to assess the effect of preserving most of the variations in the data. All generated data clearly contain value of around 60 which reflects the 60mm/day threshold of torrential rainfall. Then, after data the generation completed, it had to be examined to assess whether the data came from the same distribution as the original data set.

4. RESULTS AND ANALYSIS

4.1. Descriptive Statistics for Torrential Rainfall Data

The descriptive statistics in terms of mean, standard deviation, the minimum and maximum amount of rainfall, skewness and coefficient of variation (CV) for each fifteen rain gauge stations are summarized in Table 1. Based on the values of descriptive statistics, the highest mean and standard deviation rainfall amounts among the stations were located in East region. This is expected since north-easterly monsoon wind tends to bring heavy rainfall to these locations [23].

In terms of CV, there were differences in the coefficient of variation between regions. The stations at the East region showed the largest variability of torrential rainfall amounts, with range from 46% to 70%. The lowest coefficient variation was found in the Northeast station with variation less than 39%. These results showed the irregularity of the daily torrential rainfall received by the stations [24]. On the other hand, Northeast and East region gave positive skewness with starting value from 1.8 mm to 5.8 mm. The results illustrated that the shape of rainfall distribution for the stations in these two regions were skewed due to the values of the skewness which were far enough from zero. Hence, we can conclude that the daily torrential rainfall data may not come from a normal distribution.

Table 1. Summary Statistics of Daily Torrential Rainfall Amount (mm) for Each Station Divided by Regions

Region	Station	Mean	Standard Deviation	Skewness	Coefficient Variation (%)
Northern	Pintu A.Bagan	85.0	26.8	2.6	32
	KlinikBkt. Bendera	88.5	34.5	5.8	39
	Selama	78.6	18.4	1.8	23
Eastern	Sek.Keb. Kemasek	107.6	52.6	1.8	49
	Kg.Merang, Setiu	113.8	60.6	1.9	53
	Sek.Men. Sultan Omar, Dungun	104.0	53.3	3.7	51
	Kg. Menerong	109.4	76.8	4.0	70
	Sek.Keb.Kg. Jabi	102.0	46.7	1.8	46
	JPS Kemaman	109.0	58.2	2.5	53
	Stor JPS Kuala Terengganu	118.1	68.4	2.4	58
	KlinikBidan, Jambu Bongkok	111.4	75.7	2.7	68
	Kota Bharu	115.9	73.4	3.0	63
	Endau	107.4	52.5	1.8	49
	Rumah Pam Pahang Tua,Pekan	114.0	63.3	2.4	56
	Kuantan	106.6	59.6	3.0	56

4.2. Fitting Distribution on Daily Torrential Rainfall Data in Peninsular Malaysia

Several graphs had been used in order to determine the best fitting distribution on torrential rainfall data such as QQ plot and probability difference graph. QQ plot revealed differences in location, spread and shape more clearly than density plots or histograms did. QQ plot was used to determine how well a theoretical distribution models the set of rainfall data provided. The QQ plot shows the best fit distribution when the quantiles from the rainfall data would lie close to where they might be expected and the points on the plot would straggle about the line $y=x$. Based on the Figure 1, Generalized Pareto distribution (GPD) shows that the best fitting distribution compared to other distributions with evidence that there was solid agreement in the body of the distribution with minimal fluctuations at the tails. From the plot, almost all of the points fall on the reference line and for the few that are not on the line, are still very close to it. Gamma and Log-Normal distributions depicted the worst fit because the tails of the plot vary far away from the reference line although there were few data points which fall onto or around the reference line.

Figure 2 represents the graph comparing the three distributions by seeing the difference between the empirical cumulative distribution function and the fitted CDF. The fitted distribution is considered good if the maximum absolute difference is less than 5% and the best fitting distribution considered ‘excellent fitted’ when the value is less than 1%. According to the Figure 2, the absolute value of the difference for GPD (the green line) is less than 1% meanwhile the Gamma (the pink line) and Log- Normal (orange line) is more than 10%.

Graphical techniques could be quite subjective and may produce misleading results. Hence, statistical measure is needed to support the results from the graphical techniques. The goodness-of-fit tests which are Chi-square and Anderson-Darling test at $\alpha(0.01)$ level of significance were used in this study for the selection of the best fit probability distribution. The goodness-of-fit tests are used to measure the “distance”(which refer χ^2 and A^2 test in this study) between the torrential rainfall data and the distribution that are testing and also compare that distance to some threshold value (critical value). The critical value depends on the sample size and the choice of significance level (α). The best fitting distribution of the torrential rainfall data is when the χ^2 or A^2 test is less than critical value.

The test statistic A^2 and χ^2 were computed for three probability distribution. Table 2 shows that GPD distribution was remarkably good due to the $\chi^2 < \text{critical value of } \chi^2$ and $A^2 < \text{critical value of } A^2$. It means that the rainfall data is not in the tail of distribution and H_0 is accepted. From the results of graphical implication and goodness-of-fit tests, GPD distribution is relatively concluded to have the best fit of torrential rainfall data in Peninsular Malaysia.

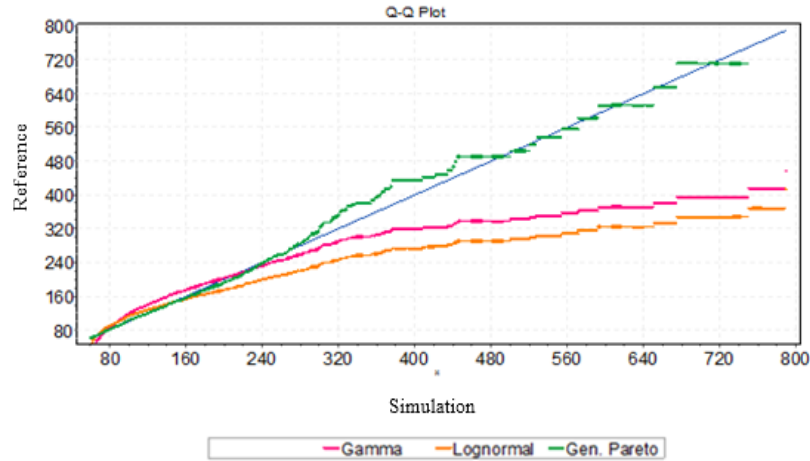


Figure 1. QQ plot displays the diagonal line along which the graph points should fall for three different distributions function

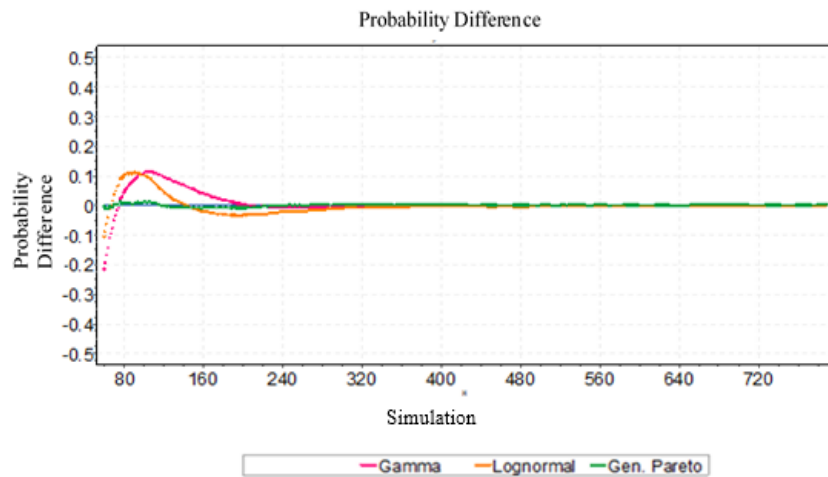


Figure 2. Plot of the difference between the empirical cumulative distribution function (CDF) and the fitted CDF for three different distributions

Table 2. Goodness of The Fit Test to Identify the Best Fitting Distribution to the Data Set.
Note That $\alpha = 0.01$

Distribution	Chi-Squared		Anderson Darling	
	χ^2	Critical Value	A^2	Critical Value
GPD	17.59	24.73	2.58	3.91
Log normal	485.42	24.73	118.65	3.91
Gamma	1037.5	24.73	213.44	3.91

4.3. Statistical Parameters for Simulation

Based on the results of fitting distribution, GPD had been chosen as the best fitting distribution for daily torrential rainfall data in Peninsular Malaysia. Hence, simulating the data matrices and estimation of the parameters for the sampled probability distributions were required. The required parameters are location parameter μ , scale parameter σ and shape parameter ξ . These parameters are based on the mean and standard deviation from the data set of 250 torrential rainfall days and 15 rainfall stations of the daily torrential rainfall data for the 33 years period in Peninsular as shown in Table 1. The shape parameter for this study is $\xi=0.2$ where it performed well for $0 \leq \xi \leq 1$ and very good for $\xi \leq 0.5$ [25].

5. CONCLUSION

The preferred best fitted distribution for daily torrential rainfall data in Peninsular Malaysia introduced in this study is useful for simulating rainfall data. Fitting distribution was used for selecting a statistical distribution that fits well to the generated data by comparing several distributions as potential candidates for the data generating mechanism. Several distributions had been tested and compared in this study to find the best fitting distribution. The distributions tested were Generalized Pareto, Lognormal and Gamma distributions. Based on distribution graphs (QQ plot and probability difference graph) and goodness of fit tests (Chi-square test and Anderson Darling Test), it was identified that the daily torrential rainfall data in Peninsular Malaysia seems to fit the Generalized Pareto distribution. This finding is important to generate synthetic rainfall data for rain gauge stations over Peninsular Malaysia with inadequate records. In addition, the purpose of using simulation data is to evaluate the performance of the proposed method against classical method in order to analyze the identifying rainfall patterns in Peninsular Malaysia.

ACKNOWLEDGEMENTS

The authors would like to thank Universiti Pendidikan Sultan Idris for their financial funding through GPU grant Vote No. 2018-0154-101-01.

REFERENCES

- [1] M.T. Amin, M. Rizwan, A.A. "Alazba, A Best-Fit Probability Distribution for the Estimation of Rainfall in Northern Regions of Pakistan", *Open Life Sci.* 11(2016), 432-440.
- [2] M.J. Mamman, O.Y. Martins, J. Ibrahim, M. I. Shaba, "Evaluation of Best-Fit Probability Distribution Models for the Prediction of Inflows of Kainji Reservoir, Niger State, Nigeria", *Air, Soil and Water Research.* 10(2017)1-7.
- [3] S. Dan'azumi, S. Shamsudin, A. Aris, "Probability Distribution of Rainfall Depth at Hourly Time-Scale", *World Academy of Science, Engineering and Technology.* 4(12)(2010), 670-674.
- [4] A.A. Jemain, J. Suhaila, "Fitting the Statistical Distribution for Daily Rainfall in Peninsular Malaysia based on AIC criterion", *Journal of Applied Science Research.* 4(12)(2008), 1846-1857.
- [5] S. Yue, M. Hashino, "Probability Distribution of Annual, Seasonal and Monthly Precipitation in Japan", *Hydrological Science Journal.* 52(5)(2007)863-877.
- [6] Y. Fadhilah, M.D. Zalina, V-T-V. Nguyen, S. Suhaila, Y. Zulkifli, "Fitting the Best-Fit Distribution for the Hourly Rainfall Amount in the Wilayah Persekutuan", *Jurnal Teknologi.* 46(C) (2007) 49-58.
- [7] S. Dan'azumi, S. Shamsudin, A. Aris, "Modeling the Distribution of Rainfall Intensity using Hourly Data", *American Journal of Environmental Sciences.* 6(3) (2010) 238-243.
- [8] R.M. Yunus, M.M. Hasan, N.A. Razak, Y.Z. Zubairi, *International Journal of Climatology.* 37 (2017) 1391-1399.
- [9] J. Suhaila, A. A. Jemain, "Fitting the Statistical Distributions to the Daily Rainfall Amount in Peninsular Malaysia", *Jurnal Teknologi.* 46(C) (2007) 33-48.
- [10] W.Z. Wan Zin, A.A. Jemain, "The Best Fitting Distribution of Annual Maximum Rainfall in Peninsular Malaysia based on Methods of L-moment and LQ-moment", *Theor. Appl. Climatol.* 96 (2009) 337-344.
- [11] T.M. Mohamed, A.A.A. Ibrahim, "Fitting Probability Distribution of Annual Rainfall in Sudan", *Journal of Engineering and Computer Sciences.* 17(2) (2016) 34-39.
- [12] S.M. Shaharudin, N. Ahmad, F. Yusof, "The Comparison of T-Mode and Pearson Correlation Matrices in Classification of Daily Rainfall Patterns in Peninsular Malaysia", *Matematika.* 29(1c)(2013), 187-194.
- [13] M.A. Alam, K. Emura, C.Farnham, J.Yuan, "Best-Fit Probability and Return Periods for Maximum Monthly Rainfall in Bangladesh", *Climate.* 6(9) (2018) 1-16.
- [14] S.M. Shaharudin, N. Ahmad, N.H. Zainuddin, "Modified Singular Spectrum Analysis in Identifying Rainfall Trend over Peninsular Malaysia", *International Journal of Electrical Engineering and Computer Science (IJECS).* 15(1)(2019), 289-293.
- [15] G.J. Husak, J. Michaelsen, C. Funk, "Use of the Gamma Distribution to Represent Monthly Rainfall in Africa for Drought Monitoring Applications", *International Journal of Climatology.* 27 (2007), 935-944.
- [16] F. Alahmadi, N. Abd Rahman, M. Abdulrazzak, "Evaluation of the Best Fit Distribution for Partial Duration Series of Daily Rainfall in Madinah, Western Saudi Arabia, Evolving water resources system: understanding", predicting and managing water-society interactions proceedings of ICWRS2014, Bologna, Italy. DOI: 10.5194/piahs-364-159-2014.
- [17] N.O. Alghazali, D.A. Alawadi, "Fitting Statistical Distributions of Monthly Rainfall for Some Iraqi Stations", *Civil and Environmental Research.* (6)(6)(2014), 40-46.
- [18] S.M. Shaharudin, N. Ahmad, N.H. Zainuddin, N.S. Mohamed, "Identification of Rainfall Patterns on Hydrological Simulation using Robust Principal Component Analysis, *Indonesian Journal of Electrical Engineering and Computer Science (IJECS).* 11(3) (2018) 1188-1193.
- [19] A.A. Mamoon, A. Rahman, "Selection of the best fit probability distribution in rainfall frequency analysis for Qatar", *Nat Hazards.* 86(2017) 281-296.
- [20] L. Ye, L.S. Hanson, P/ Ding, D. Wang, R.M. Vogel, "The Probability distribution of Daily Precipitation at the Point and Catchment Scales in the United States", *Hydrol. Earth Syst. Sci.* 22(2018), 6519-6531.

- [21] S. Lavanya, M. Radha, U. “Arulanandu, Statistical Distribution of Seasonal Rainfall data for Rainfall Pattern in TNAU1 Station Coimbatore, Tamil Nadu, India”, *International Journal of Current Microbiology and Applied Sciences*. 7(4) (2018), 3053-3062.
- [22] Mierlus-Mazilu, “On Generalized Pareto Distributions”, *Romanian Journal of Economic Forecasting*. 8 (2010), 107-117.
- [23] S. Jamaludin, S. MohdDeni, W.Z. Wanzin, A.A. Jemain, “Trends in Peninsular Malaysia Rainfall Data during the Southwest Monsoon and Northeast Monsoon Seasons”, *Sains Malaysiam*. 39(4)(2010), 533-542.
- [24] S. Jamaludin, F. Yusof, “Introducing the Mixed Distribution in Fitting Rainfall Data”, *Open Journal of Modern Hydrology*. 1(2011) 11-22.
- [25] P. De Zea Bermudez , S. Kotz, “Parameter Estimation of The Generalized Pareto Distribution-Part I”, *Journal of Statistical Planning and Inference*. 140(6) (2010), 1353-1373.

BIOGRAPHIES OF AUTHORS



Shazlyn Milleana was born in Johor Bahru, Malaysia, in 1988. She is a senior lecturer at the Department of Mathematics, Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris (UPSI). She graduated with a bachelor science degree in Industrial Mathematics from Universiti Teknologi Malaysia, in 2010. Upon graduation, she began her career as an Executive in banking institution. In the following year, she received an offer to continue her study as a fast-track PhD student at the same university. During her PhD journey, she developed an interest in multivariate analysis, specifically in finding patterns which deals with big data. Her research focuses on the area of dimension reduction methods specifically in climate informatics which involves analysis on huge climate-related datasets based on techniques in Data Mining. She had published her research in Scopus indexed journal and presented her work in various local and international conferences. She completed her PhD thesis at the end of 2016 and was conferred a doctorate degree in 2017.



Norhaiza is a senior lecturer at the Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia (UTM). She graduated with an honors degree in Mathematics, Statistics and Operational research from the University of Manchester, in 1996. She joined UTM in August 2000. In the following year, she continued her studies at the University of Sheffield for her master's degree. In Sheffield, she developed an interest in multivariate analysis, specifically in finding patterns which lead her to pursue a PhD degree at the University of Kent. She completed her PhD thesis at the end of 2007 and was conferred a doctorate degree in 2008. Finding patterns in any data have always been her research interests. She started the interests in profiling data – finding statistically distinctive and significant groups and features in the object of interest whilst at Sheffield. Currently, her research interests revolve around hydroinformatics particularly in investigating the streamflow variability of the local rivers.



Nur Syarafina Mohamed is a Senior Lecturer in Universiti Kuala Lumpur specialized in Optimization area. She graduated with a bachelor science degree in Industrial Mathematics from Universiti Teknologi Malaysia, in 2010. Upon graduation, she began her career as a lecturer in Universiti Teknologi Mara from 2010 till 2016 in a contract basis. Around 2013, she started her PhD journey in Universiti Sultan Zainal Abidin in Kuala Terengganu Malaysia. Her research interest is focused on Optimization where she modified the Conjugate Gradient Method by implementing new adjustment on the coefficient used in the algorithm. The parameter introduced is compared among the best parameters which was introduced previously. She had published her research in Scopus indexed journal and presented her work in various local and international conferences. She completed her PhD thesis in April 2017 and was conferred a doctorate degree in December 2017.



Hairulnizam Mahdin is an Associate Professor at Department of Information Security and Web, Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia. He completed his Ph.D. thesis in 2012 and was conferred a doctorate degree from Deakin University Australia in the same year. His research focuses on the area of data management, IoT, RFID, information security, software engineering and web technology. He has published his research in both ISI and Scopus-indexed journals.