

Rainfall Modelling using Generalized Extreme Value Distribution with Cyclic Covariate

Jasmine Lee Jia Min, Syafrina Abdul Halim*

Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, 43400, Selangor, Malaysia

Received October 11, 2020; Revised November 21, 2020; Accepted December 13, 2020

Cite This Paper in the following Citation Styles

(a): [1] Jasmine Lee Jia Min, Syafrina Abdul Halim , "Rainfall Modelling using Generalized Extreme Value Distribution with Cyclic Covariate," *Mathematics and Statistics*, Vol. 8, No. 6, pp. 762-772, 2020. DOI: 10.13189/ms.2020.080617.

(b): Jasmine Lee Jia Min, Syafrina Abdul Halim , (2020). *Rainfall Modelling using Generalized Extreme Value Distribution with Cyclic Covariate*. *Mathematics and Statistics*, 8(6), 762-772. DOI: 10.13189/ms.2020.080617.

Copyright ©2020 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract Increased flood risk is recognized as one of the most significant threats in most parts of the world, resulting in severe flooding events which have caused significant property and human life losses. As there is an increase in the number of extreme flash flood events observed in Klang Valley, Malaysia recently, this paper focuses on modelling extreme daily rainfall within 30 years from year 1975 to year 2005 in Klang Valley using generalized extreme value (GEV) distribution. Cyclic covariate is introduced in the distribution because of the seasonal rainfall variation in the series. One stationary (GEV) and three nonstationary models (NSGEV1, NSGEV2, and NSGEV3) are constructed to assess the impact of cyclic covariates on the extreme daily rainfall events. The better GEV model is selected using Akaike's information criterion (AIC), bayesian information criterion (BIC) and likelihood ratio test (LRT). The return level is then computed using the selected fitted GEV model. Results indicate that the NSGEV3 model with cyclic covariate trend presented in location and scale parameters provides better fits the extreme rainfall data. The results showed the capability of the nonstationary GEV with cyclic covariates in capturing the extreme rainfall events. The findings would be useful for engineering design and flood risk management purposes.

Keywords Covariate, Cyclic, Generalized Extreme Value, Rainfall, Stationary

1 Introduction

Extreme value analysis of hydrological data allows interpreting historical data and making inference on future probabilities

of occurrence of extreme events, such as floods due to extreme rainfalls. Extreme values are often represented by the maximum value of a given variable over a time period such a year [1]. As climate change became an increasingly important issue over the past two decades, technical terms like "stationarity" and "nonstationarity" also became more obvious. Nonstationarity can simply be defined as processes that are constant in time and that have statistical properties that are deterministic functions of time. Statistical inference for hydrological time series such as extreme rainfall events normally relied on the assumption of stationarity [2]. Nevertheless, under the combined influences of climate change and variability or human intervention [3], the series often exhibits nonstationary features and will not likely satisfy stationary assumptions [4, 5]. Nonstationarity may affect both the severity and frequency of these extreme events [6]; therefore, the reality of nonstationary hydrometeorological extremes needs to be properly addressed. Accurate distribution models are required to accurately capture the nonstationary climate. Nonstationary extreme value models with climatic covariates could be the valuable tools to assess future changes in extreme rainfall distribution and quantiles for engineering design and flood risk management purposes [7].

One of the well-known distributions that were used in hydrological analysis is known as generalized extreme value (GEV) distribution. GEV distribution is one of the most commonly used for the analysis of extreme rainfall events [8]. Estimation of return levels is usually based on three extreme value distributions: Gumbel, Fréchet and Weibull which were suggested by [9] for stochastic behaviour of large samples. Quite a number of studies have been performed in Malaysia for determining the most suitable probability distribution of extreme rainfall events. Gamma, Generalized Normal, Generalized Pareto, GEV, Gumbel, Log Pearson Type III, Pearson Type III

and Wakeby have been compared to determine the most accurate and appropriate distribution for the maximum rainfall estimates. It has been reported that GEV shows better descriptive and predictive abilities for the annual extreme rainfall as compared to other distributions [10]. From the analysis, it was found that 94% of rainfall stations in Peninsular Malaysia was best fitted by GEV. The findings are consistent with Nguyen [11] where several distributions namely Beta-K, Beta-P, GEV, Generalized Logistic distribution, Generalized Normal distribution, Generalized Pareto, Gumbel, Log Pearson Type III, Pearson Type III and Wakeby have been compared. It was found that GEV, Generalized Normal distribution and Pearson Type III were the best distribution for the annual maximum rainfall series. However, GEV was preferred as it gives a better description of annual maxima rainfall series where it has more solid theoretical basis which makes it preferable in determining the daily annual maximum rainfall. GEV was also capable in deriving the Intensity-Duration-Frequency (IDF) curve for rainfall series in [12] over other distributions. Recently, extreme precipitation data were fitted with four distributions namely, Generalized Pareto, Gumbel, GEV, and Exponential in Peninsular Malaysia. Results have shown that GEV distribution was found to best fit the time series at most of the stations in Peninsular Malaysia based on Kolmogorov Smirnov test [13].

It is very common to allow trends in the GEV parameters to consider nonstationarity in the frequency analysis of hydrological variables [14] as constant parameters may no longer be valid under nonstationary conditions. Several studies have developed nonstationary GEV models in hydrological area. Adlouni et al. [1] developed a nonstationary GEV with the linear and quadratic dependence of the location parameter on covariates, and the linear dependence in both location and scale parameters. Maraun et al. [15] proposed a study that investigated the annual cycle of extreme events across the United Kingdom by developing a statistical model and fitting the nonstationary GEV to data from 689 rain gauges. GEV distribution was used to fit to the time series of monthly maxima, across all months of the year simultaneously. The annual cycles of the location and scale parameters are approximated by harmonic functions, while keeping the shape parameter constant throughout the year. The study revealed that the approximation of the monthly maxima distribution by the GEV and the annual cycle by a sine wave appeared to be reasonable. The sinusoidal model proved to be a good compromise between a bias due to a stationarity assumption and the uncertainty owing to too many parameters.

Agilan and Umamahesh [16] observed that urbanization and local temperature changes were the best covariate factors for short duration rainfall in India. However, the covariate time was not concluded as the best covariates. Golroudbary et al. [17] found that the nonstationary GEV model gives a better estimate by considering the impact of the North Atlantic Oscillation and the annual seasonal cycle on the parameters. Meanwhile, this study uses a harmonic function model for all monthly maxima during the year with seasonal variations instead of individual models for every month. Sharma and Mujamdar [18] concluded that local mean temperature was found

to be significant covariate of nonstationary GEV for extreme rainfall of the studied region. A nonstationary GEV model with cyclic covariate structure is proposed in this work. All extreme events were calculated assuming that maximum annual daily precipitations follow the GEV distribution. The parameters are estimated by using maximum likelihood estimation (MLE). To account for nonstationarity, the parameters of the extreme value distribution are modified with a set of predictors or covariates [19, 20]. Accordingly, the sine and cosine functions are developed by considering the impact of cyclic covariate on the location and scale parameters. Therefore, the objectives of the study are to fit the stationary (GEV) and nonstationary (NSGEV) model with cyclic covariate on extreme daily rainfall series using the MLE, to identify the best fitted GEV model for extreme daily rainfall series using the Akaike's information criterion (AIC), bayesian information criterion (BIC) and likelihood ratio test (LRT) and to estimate the return level of extreme rainfall events using the best fitted GEV model.

2 Materials and Methods

Klang Valley covers the area of Kuala Lumpur and all the adjoining cities and towns in the state of Selangor. Over the years, the cities of Klang Valley have seen significant development and growth, leading to a good economy and a strong real estate market. The Klang Valley river basin is located on the west coast of Peninsular Malaysia and surrounds the federal territory of Kuala Lumpur and some parts of the state of Selangor. Since the Klang Valley river basin is surrounded by a heavily populated area with more than 4 million people, therefore it is considerably polluted. Huge development has narrowed certain stretches of the river and eventually resembles a large storm drain in some places. Hence, this contributes to the occurrences of flash floods in Kuala Lumpur, especially after heavy rain [21]. JPS Ampang rainfall station is selected in this study to represent Klang Valley. This is because, every year, higher rainfall amount is being recorded at this station.

2.1 Data

Daily rainfall data between the years 1975 and 2005 were sourced from the Drainage and Irrigation Department. JPS Ampang station as in Fig. 1 will be selected to represent daily rainfall in Klang Valley. Daily rainfall data consists of missing data out of 39 of 11324 total daily rainfalls. A missing rate of 5% or less in the datasets is inconsequential [22].

2.2 Stationary test

In order to fulfil the stationarity assumption of the GEV distributions, the Mann-Kendall (MK) test is performed on the data to check for trends over different selection periods. The MK trend test is a widely used non-parametric test for evaluating the presence of monotonic trends in time series data [23]. The MK test is a widely used technique for detecting monotonic trends in hydrological and meteorological time series [28]. The MK test which does not require normally distributed data and is well suited for analysing datasets that have

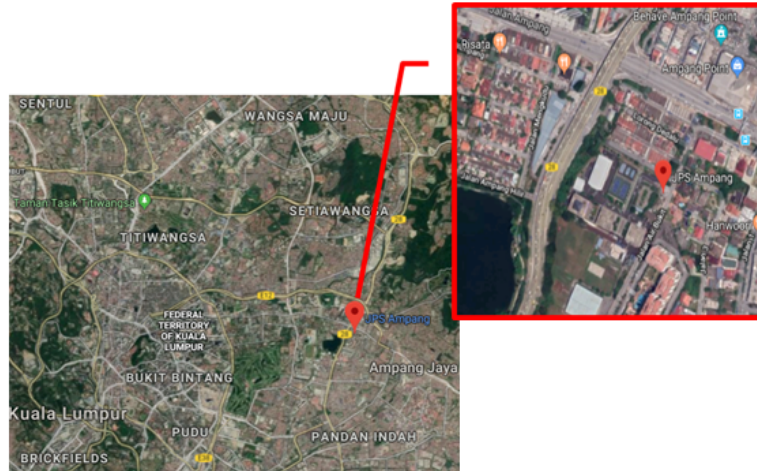


Figure 1. Location of JPS Ampang station.

missing or tied data [25]. The annual maximum daily rainfall series are concerned. The variable m , with $t = 1, \dots, T$, to denote these annual maxima, and with T being the total number of years of data in the record. The differences in the data values between different time steps as

$$d(t_1, t_2) = m_{t_2} - m_{t_1} \tag{1}$$

with $t_2 > t_1$, then the test statistic becomes

$$N = \sum_{t_1=1}^{T-1} \sum_{t_2=t_1+1}^T \text{sgn}[d(t_1, t_2)] \tag{2}$$

with $\text{sgn}[d(t_1, t_2)]$ denoting the sign of $[d(t_1, t_2)]$. The test statistic N represents the number of m_{t_2} is greater than m_{t_1} , minus the number of times m_{t_1} is greater than m_{t_2} , for all possible combinations of m_{t_2} and m_{t_1} with $t_2 > t_1$. A positive value of N implies that the time series increased more frequently than it decreased (and vice versa for a negative value of N), and the value of N is bounded by $\pm T(T - 1)/2$. Kendall's τ is a normalized version of this statistic, which is obtained by dividing by this upper bound

$$\tau = \frac{2N}{T(T - 1)} \tag{3}$$

So that τ is bounded by $[-1, 1]$. Assuming the data are serially independent, the null hypothesis can be approximated by a normal distribution, with more details in [23]. Using a 5% two-sided significance level, H_0 is rejected when the test statistics is less than the critical value which is 0.05. The Mann-Kendall analysis is conducted using the R package.

H_0 : There exists no trend in the extreme rainfall series

H_1 : There exists trend in the extreme rainfall series

2.3 Stationary generalized extreme value distribution (GEV)

The GEV distribution is a family of continuous probability distributions developed within extreme value theory (EVT) which comprises of Gumbel, Fréchet and Weibull families. The GEV distribution is parameterized with a shape parameter (ξ), location parameter (μ) and scale parameter (σ). Each type corresponds to the limiting distribution of block maxima from a different class of underlying distributions. If the tails of distribution decrease exponentially, it leads to the type I (Gumbel) ; if the tails of distribution decrease as a polynomial, it leads to type II (Fréchet) and if the distribution has finite tails, it leads to the type III (Weibull). GEV combines three distributions into a single form, allowing a continuous range of possible shapes that include all three of the distributions. One of the distributions is used to model a particular dataset of block maxima. Based on the EVT, the GEV distribution is the limit distribution of properly normalized maxima of a sequence of independent and identically distributed random variables. Data are blocked into sequences of observations of length n , for some large value of n , generating a series of block maxima, $M_{n,1}, \dots, M_{n,m}$, to which the GEV distribution can be fitted. Often the blocks are chosen corresponding to a time period of length one year, in which case n is the number of observations in a year and the block maxima are annual maxima. In this case, x is defined as the extreme rainfall (mm/day) in monthly maxima. Like the extreme value distribution, the GEV is often used to model the smallest or largest value among a large set of independent, identically distributed random values representing measurements or observations. The cumulative distribution function of the GEV distribution is

$$F(x; \mu, \sigma, \xi) = \begin{cases} \exp(-(1 + \xi \frac{x-\mu}{\sigma})^{-\frac{1}{\xi}}), & \sigma > 0, \\ 1 + \xi \frac{x-\mu}{\sigma} > 0, & \xi \neq 0 \\ \exp(-\exp(-\frac{x-\mu}{\sigma})), & \sigma > 0, \xi = 0 \end{cases} \tag{4}$$

The parameters of the GEV distribution are estimated by the method of maximum likelihood. A standard approach to determining the better fitting model is the maximum likelihood ratio test. Let the values $X = x_1, x_2, \dots, x_n$ be the n years of annual maximum series. The log likelihood is given as

$$\log L(\mu, \sigma, \xi | X) = -n \log \sigma - (1 + \frac{1}{\xi} \sum_{i=1}^n \log [1 + \frac{\xi(x_i - \mu)}{\sigma}]) - \sum_{i=1}^n \frac{\xi(x_i - \mu)}{\sigma} > 0, \xi \neq 0 \quad (5)$$

$$\log L(\mu, \sigma | X) = -n \log \sigma - \sum_{i=1}^n \log [\frac{\xi(x_i - \mu)}{\sigma}], \xi = 0 \quad (6)$$

The maximum likelihood estimate with respect to the entire GEV family can be estimated through maximizing (5) and (6) with respect to the parameter vector (μ, σ, ξ) [19].

2.3.1 Nonstationary generalized extreme value distribution (NSGEV)

Nonstationary processes have characteristics that change systematically through time. Nonstationarity is often apparent because of seasonal effects, perhaps due to different climate patterns in different months, or in the form of trends, more probably due to long-term climate changes. In this study, nonstationary GEV (NSGEV) distribution models will be developed to investigate whether GEV parameters exhibit any dependency with representative indices of cyclic covariate factors that have significant effects on daily rainfall in Klang Valley. Nonstationarity is introduced by revealing more than one parameters of the GEV distribution as function of cyclic covariate indices. The cumulative distribution function of the NSGEV distribution is,

$$F(x; \mu(t), \sigma(t), \xi) = \exp(-1 + \frac{\xi(x - \mu(t))}{\sigma(t)})^{-\frac{1}{\xi}}, \quad 1 + \frac{\xi(x - \mu(t))}{\sigma(t)} > 0 \quad (7)$$

Consequently, the constant GEV parameters μ (or σ or ξ) are replaced by the new parameters, μ_n and μ_1 (or the corresponding parameters for σ and ξ) [15]. On the other hand, it is quite difficult to estimate the shape parameter of the extreme value distribution with precision when it is time dependent, thereby; it is not realistic to attempt to estimate the scale parameter as a smooth function of time [19]. To obtain a model which is linear in the parameters, μ and σ are modeled as a combination of sine and a cosine with a 1-year period, evaluated at the centre-day of each month, c_i .

For the location parameter,

$$\mu(t) = \mu_0 + \mu_1 \sin(\omega c_i + \phi), i = 1, \dots, 12 \quad (8)$$

where, μ_1 denotes the amplitude of the sinusoidal component

(estimated free parameter) ϕ denotes the phase $\omega = \frac{(2\pi)}{352.25}$ which denotes the angular frequency c_i denotes t . However, it is not linear in the parameters due to phase π . Hence, 8 is reformulated to obtain an equivalent linear model as

$$\mu(t) = \mu_0 + \mu_1 \sin(\omega c_i) + \mu_2 \cos(\omega c_i), i = 1, \dots, 12 \quad (9)$$

In this study, the location and scale parameters with the seasonal variation described by sinusoidal functions and the shape parameter is assumed to be a constant in 10 and 11.

$$\mu(t) = \mu_0 + \mu_1 \sin(\frac{2\pi}{12}) + \mu_2 \cos(\frac{2\pi}{12}) \quad (10)$$

$$\sigma(t) = \sigma_0 + \sigma_1 \sin(\frac{2\pi}{12}) + \sigma_2 \cos(\frac{2\pi}{12}) \quad (11)$$

where, the terms above are the same as described in 8, 12 means 12 months in a year and ξ is treated as constant.

2.3.2 Selection of models

The significance of the models will be tested using AIC[26] and BIC [27] and LRT. Both the AIC and BIC are used to select the best nonstationary fitting models. The best fitted GEV model is selected following these three steps: (1) the model with the lowest AIC value is identified; then, (2) the model with the lowest BIC value, (3) the greatest deviance statistics value which indicates that NSGEV model explains more variability of data than GEV model. But how large is large for deviance statistics value should be determined from asymptotic distribution of deviance function. These steps imply that if one of the NSGEV model is the best fitted model, then the cyclic covariate trend is said to be significantly influenced the extreme daily rainfall at the 5% significance level based on LRT [28].

The AIC is

$$AIC = -2 \log L(\hat{\theta}) + 2k \quad (12)$$

The BIC is

$$BIC = -2 \log L(\hat{\theta}) + k \log n \quad (13)$$

where, θ denotes the vector of model parameters, $L(\hat{\theta})$ denotes the likelihood of the candidate model given the data when evaluated at the maximum likelihood estimate of θ k denotes the number of estimated parameters in the candidate model n denotes the number of observations.

Statistical significance of nonstationary parameters in the model M_1 against GEV model is commonly evaluated with LRT. If NSGEV models outperform the stationary GEV models; then the NSGEV models should be used instead of the GEV models for frequency analysis of extreme daily rainfall. Let M_0 and M_1 be stationary and nonstationary models respectively. LRT is employed to compare the superiority of

M_1 over M_0 using the log-likelihood difference (D) [1, 19].

$$D = 2[l(\hat{\theta}_0) - l(\hat{\theta})] \tag{14}$$

where, $l(\hat{\theta}_0)$ denotes the maximized log-likelihood under non-stationary model $l(\hat{\theta})$ denotes the maximized log-likelihood under stationary model. Large values of D indicate the model M_1 explains substantially more of the variation in the data than M_0 . Hypothesis is tested based on Chi-Square distribution with critical value of degree of freedom of 1 is 3.84 at the 0.05 significance level thus the model H_0 can be rejected which mean model H_1 is favoured.

H_0 : NSGEV model is not superior to GEV model

H_1 : NSGEV model is superior to GEV model

2.4 Return level

Once the best fitted GEV model for the data has been determined, the return level of extreme daily rainfall is derived. The probability P of the occurrence of the extreme events is defined as the chance of the event occurring at least once on average in T years; hence, $P = \frac{1}{T}$ [17]. More precisely, it is the level exceeded y the monthly maxima in any month with probability, p . Estimation of the return level, z_p is obtained from the stationary models by inverting the cumulative function of GEV distribution.

For GEV model,

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi}[1 - y_p^{-\xi}], & \xi \neq 0 \\ \mu - \sigma \log(y_p), & \xi = 0 \end{cases} \tag{15}$$

where, $G(z_p) = 1 - p, y_p = -\log(1 - p)$, and $0 < p < 1$.

For NSGEV model,

$$z_p = \mu(t) - \frac{\sigma(t)}{\xi}[1 - y_p^{-\xi}], \xi \neq 0 \tag{16}$$

where, $\mu(t)$ and $\sigma(t)$ are described in (7) and (8)

$$G(z_p) = 1 - p, y_p = -\log(1 - p), 0 < p < 1 \tag{17}$$

20-year, 50-year and 100-year return levels of extreme rainfall for the year 1975 and 2005 are predicted using (12) and (13) derived from the corresponding GEV distribution.

3 Results

A summary of daily rainfall amount for JPS Ampang station is summarized in Table 1 where the mean, standard deviation, median, first quartile, third quartile, minimum value and maximum value are given. Total daily rainfall in January showed

the least standard deviation whereas the highest standard deviation is recorded in April which indicates the largest variation in their daily rainfall amount series. This could possibly be affected by the extreme values (maximum value). Based on Fig. 2, the station recorded the highest amount rainfall in April and November due to heavier rainfall that occurred and least rainfall in January. Due to the northeast monsoon and southwest monsoon season, there are some infrequent outliers lies in the daily rainfall data series. These outliers are treated as maximum values where the GEV will be applied to fit these values.

4 Discussion

The presence of trend in the extreme rainfall series is tested using the MK trend test in which the null hypothesis indicates no trend. The daily rainfall is blocked into monthly as in Table 2 suggests that the annual maximum rainfall shows that there is no trend while the monthly maximum rainfall shows the existence of a trend as time increases. The seasonal monthly trend is more pronounce than annual trend as Malaysia’s surface climate is influenced by two monsoon seasons namely northeast (November until February) and southwest (May until August) monsoon seasons. In between these two monsoon seasons are the inter-monsoon seasons occurring in March – April and September – October, which brings convective rainfall. The annual average rainfall is 2,420 mm for Peninsular Malaysia, 2,630 mm for Sabah and 3,830 mm for Sarawak. Malaysia has consistently hot and humid weather throughout the year as it is situated near the equator. This explains the significance of the test statistics of MK where the monthly trend is more significant than annual trend.

4.1 GEV and NSGEV models

The total daily rainfall amount per month for JPS Ampang station during the period January-December for all available year is considered. The time blocks of fixed lengths (months) is considered and the maximum value is taken for each block. There are four GEV or NSGEV distribution models presented in this study: GEV0 is a stationary model where the parameters are treated as constant; NSGEV1 is a nonstationary model where the location parameter is treated as time dependent parameter; NSGEV2 is a nonstationary model where the scale parameter is assumed as time dependent parameter; NSGEV3 is a nonstationary model where both scale and location parameters are treated as time dependent parameters. The covariate functions in each model are presented in Table 3.

4.2 Parameter estimations and model selection

GEV distribution is parameterized with a shape parameter ξ , location parameter μ and scale parameter σ . Nonstationarity was introduced by revealing more than one parameters of the GEV distribution as function of cyclic covariate. μ and σ are described by sine and cosine functions while ξ is assumed to be a constant. In Table 4, the nonstationary models used in this study incorporated unfixed location and scale parameters and the shape parameter is fixed. The “isnev” package in

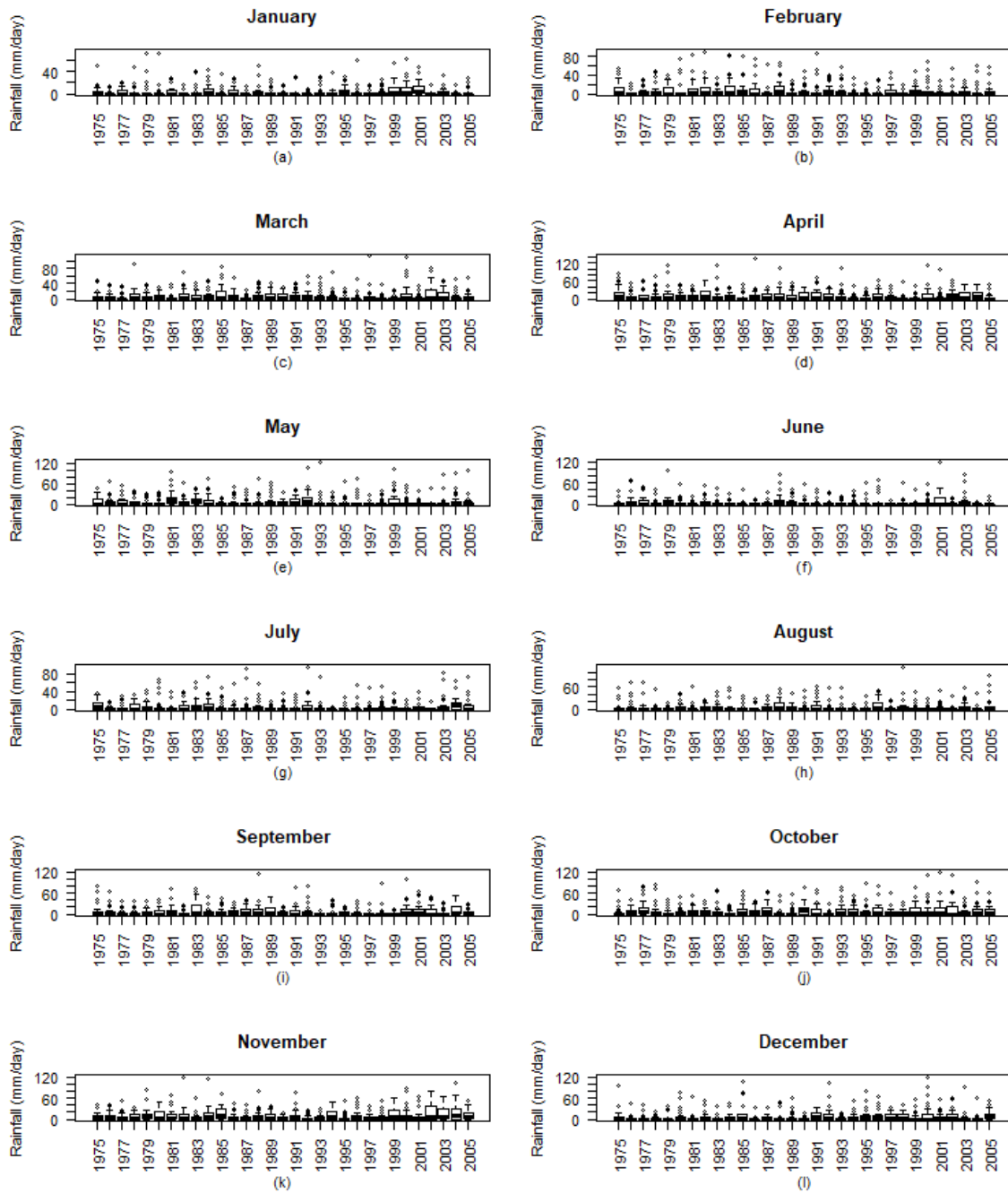


Figure 2. Boxplot of amount of daily rainfall from 1975 until 2005 at JPS Ampang station, Klang Valley.

Table 1. Statistics of rainfall amount (mm/day) from 1975 until 2005.

Month	Mean	Median	Standard deviation	First quartile	Third quartile	Minimum value	Maximum value
January	3.6931	0.00	9.0076	0	2.000	0	72.5
February	6.5467	0.00	14.0105	0	4.825	0	89.0
March	6.9974	0.40	14.0512	0	7.500	0	113.5
April	9.3381	1.00	16.9345	0	11.200	0	135.0
May	6.8075	0.00	14.9625	0	5.500	0	124.5
June	4.5095	0.00	12.204	0	1.500	0	118.5
July	4.8863	0.00	11.718	0	3.500	0	93.0
August	5.2059	0.00	12.4387	0	3.200	0	112.5
September	6.7801	0.40	13.699	0	7.050	0	115.5
October	8.7626	1.20	16.183	0	10.500	0	0.0169
November	9.5776	2.40	15.925	0	12.500	0	0.0046
December	6.4453	0.50	14.336	0	0.500	0	0.0022

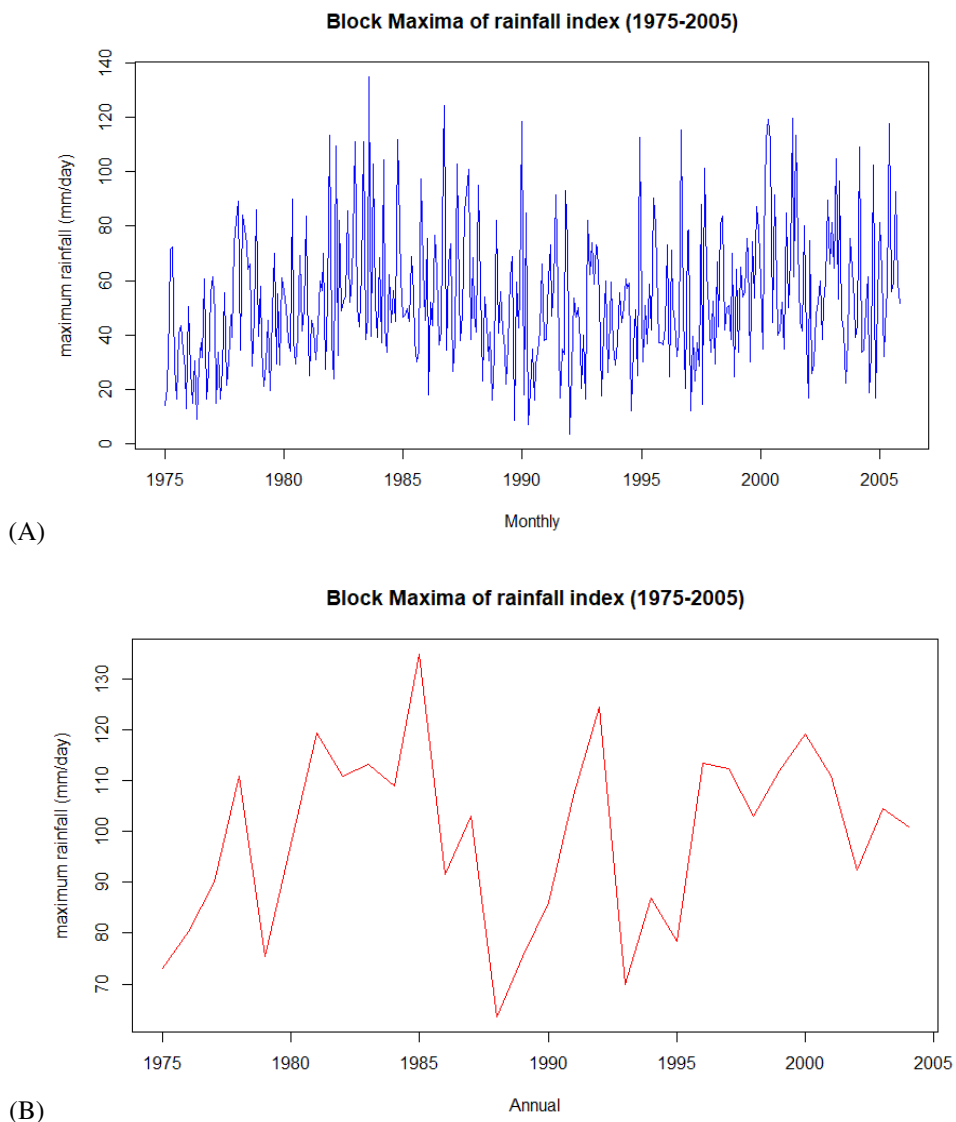


Figure 3. Time series plot of maximum rainfall at JPS Ampang station, Klang Valley over the period 1975-2005 for the (A) monthly and (B) annual periods.

Table 2. MK trend test for trend.

Period	Test statistics	(2-sided) P-value	Description
Monthly	0.104	0.002737*	Reject H_0 , there is existence of trend in the series
Annual	0.132	0.375	Do not reject H_0 , there is no existence of trend in the series

Table 3. Developed GEV and NSGEV models for monthly maxima rainfall data.

Model ID	Parameters of model	Remarks
GEV0	μ (constant), σ (constant), ξ (constant)	Stationary model
NSGEV1	σ (constant), ξ (constant), $\mu(t) = \mu_0 + \mu_1 \sin(2\pi t/12) + \mu_2 \cos(2\pi t/12)$	Nonstationary model with time dependent location parameter
NSGEV2	μ (constant), ξ (constant), $\sigma(t) = \sigma_0 + \sigma_1 \sin(2\pi t/12) + \sigma_2 \cos(2\pi t/12)$	Nonstationary model with time dependent scale parameter
NSGEV3	ξ (constant), $\sigma(t) = \sigma_0 + \sigma_1 \sin(2\pi t/12) + \sigma_2 \cos(2\pi t/12)$ $\mu(t) = \mu_0 + \mu_1 \sin(2\pi t/12) + \mu_2 \cos(2\pi t/12)$	Nonstationary model with time dependent location and scale parameter

hood under the different models is listed to enable the LRT test statistics to be computed. The result of model analysis with monthly maxima rainfall series of Klang Valley is given in Table 5. NSGEV3 is found to be the best NSGEV model for monthly maximum rainfall in which the trend is introduced in location and scale parameter with cyclic covariate. The AIC and BIC values indicate that NSGEV3 is the best or adequate model among the four models. Furthermore, the likelihood ratio test is used to evaluate between the choices of model which is $D(\theta) = 2[l(\hat{\theta}_n) - l(\hat{\theta})]$ whereby l is the log likelihood function, based on Chi-Square distribution with critical value of a degree of freedom of 1 is 3.84 at the 0.05 significance level. The deviation value between NSGEV3 and GEV0 is 4.416 which is greater than the critical value (3.84). This indicates that model NSGEV3 is the most superior to the stationary model.

4.3 Model diagnostics

Figure 4 show the model diagnostics from fitting the extreme monthly rainfall (mm/day) for one GEV model and three NSGEV models. The probability and quantile plots in Fig. 4(A) shows points are close to the linear line which means the fitted model is valid. The negative estimates of the ξ cause the return level curve to asymptote to a finite level. Thus, the estimated curve for the return level plot is approximately linear since the values of ξ is close to zero. For NSGEV1, NSGEV2, NSGEV3, only the residual probability and quantile plots are displayed with the quantile plot on the Gumbel scale in Fig. 4(B), 4(C) and 4(D). Adequacy of best selected model is checked with two graphical diagnostic plots. Inspection on NSGEV 1 and NSGEV2 resulted in similar results. These plots reveal that all models are well-fitted.

4.4 Return level estimate

Return values contain two quantities; return period $1/p$ and return level (recurrence level) z_p . Finally, the best fitting GEV model is used to estimate the 20-year, 50-year and 100-year return levels of extreme rainfalls at JPS Ampang Station between the years 1975 and 2005. The highest daily rainfall for the 30-year observation period is 135mm/day. Return levels are used to predict the probability that a daily maximum rainfall exceeding 135mm/day will occur in a longer period. The estimation of the T -year return levels for $T = 20, 50$ and 100 return periods are estimated as shown in Table 6. Equation (16) is used to estimate the return levels for return period = 20, 50 and 100 years as confidence of shape parameter ξ is not including 0. The return levels for monthly selection periods with their 95%

confidence intervals which are obtained by profile likelihood are displayed in brackets in Table 6. From this table, it can be observed that return level estimates increase as the return periods increase. The results also prove return level estimation of the best fitted GEV model indicates the rainfall does not exceed extreme rainfall value for in the next 20, 50 and 100 years in terms of monthly extreme rainfall which is 135mm/day of the observation rainfall data.

5 Conclusions

The purpose of this study is to provide statistical knowledge for the dependence of nonstationary effects of trend to be conveniently accounted for risk management. GEV distribution is used to model extreme rainfall using data obtained from JPS Ampang station for the period from 1975 to 2005. Due to the presence of the seasonal monsoon season in Malaysia, cyclic covariate trend is applied in the GEV distribution in this study. In this study, NSGEV distribution model is presented. The focus is on nonstationarity in the location and scale parameters of the GEV distribution in order to counter the seasonality factor in rainfall series. In the case of extreme value variables, the NSGEV model is better in capturing the mean and variance than GEV0. In particular, the models presented correspond to the stationary GEV model, the case in which there is a cyclic covariate trend in the location parameter, the case where the scale parameter is a cyclic function of the covariate and the case with cyclic covariate trend in both location and scale parameters. Parameter estimation for these models is generally carried out using the MLE method. The Block Maxima with GEV is applied on 30-year daily rainfall data in Klang Valley during the year of 1975 until the year of 2005. The data stationarity is also a concerning issue for this study, thus MK test is used to assess the stationarity of the data. For this study, the data is blocked into monthly and annually then the MK test is applied on the data. Results have shown that there is an existence of significance trend in the monthly extreme rainfall whereas there is an absence of significance trend in the annual extreme rainfall series. Overall, by considering the model selection criteria, NSGEV3 produces the smallest AIC value and BIC value. The deviation value between NSGEV3 and GEV0 is greater than the critical value therefore null hypothesis is rejected at 95% significance level which indicates that model NSGEV3 is the most superior to the GEV0. It can be thus concluded that NSGEV3 better fits the extreme rainfall data in JPS Ampang station. Based on the best fitted GEV model, the return levels are then computed in the period of 20 years, 50 years

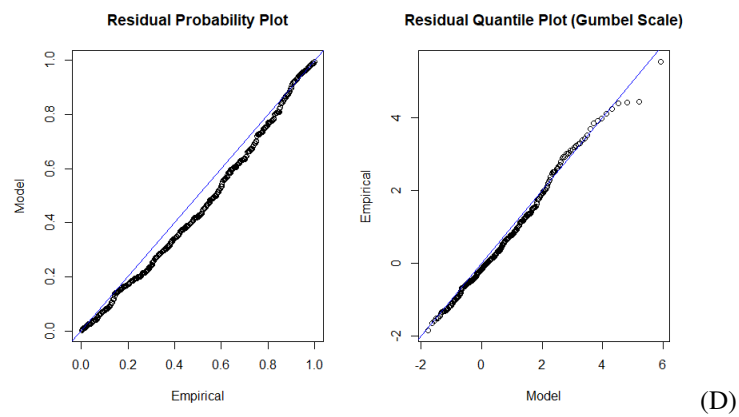
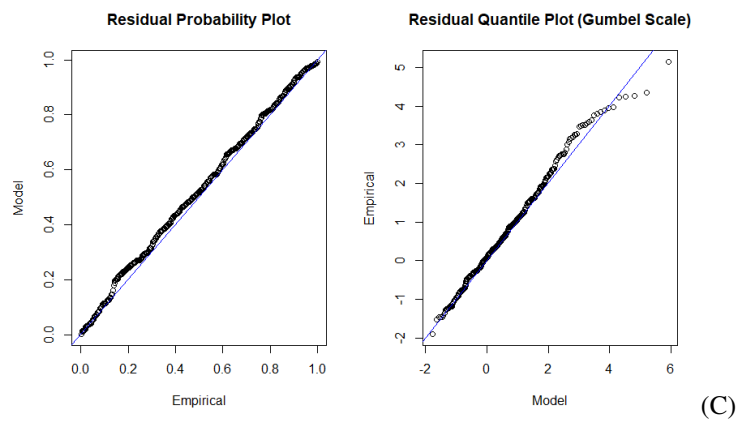
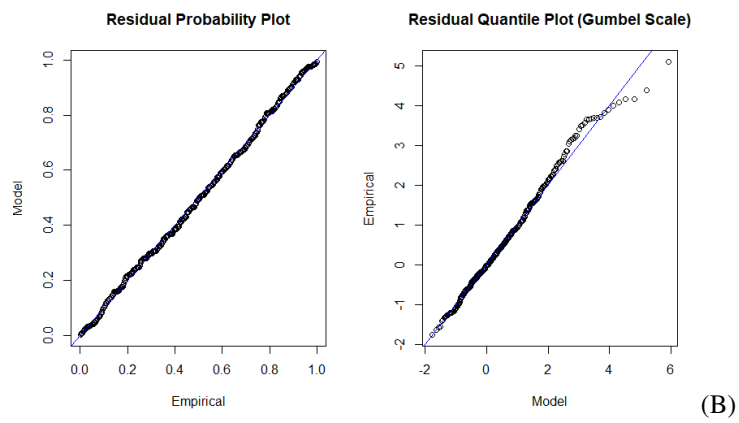
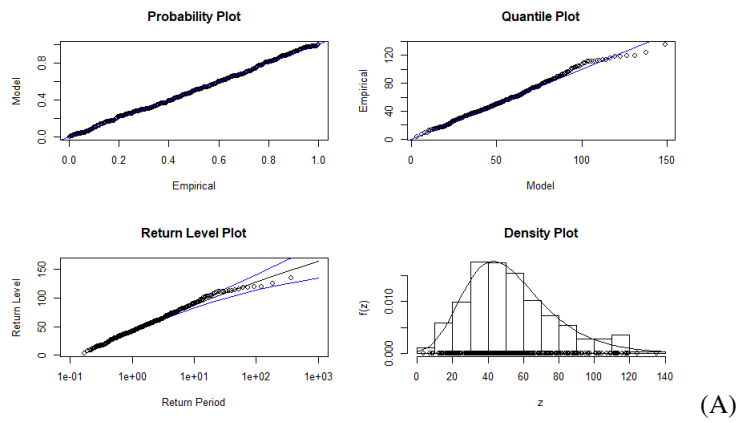


Figure 4. Model diagnostics for model (A) GEV0, (B) NSGEV1, (C) NSGEV2 and (D) NSGEV3.

Table 4. Developed GEV and NSGEV models for monthly maxima rainfall data.

Parameters	GEV0 (s.e) (LCL, HCL)	NSGEV1 (s.e) (LCL, HCL)	NSGEV2 (s.e) (LCL, HCL)	NSGEV3 (s.e) (LCL, HCL)
μ	42.0002 (1.2129) (39.6230, 44.3775)	-	41.9749 (1.2158) (39.5919, 44.3579)	-
μ_0	-	42.0148 (1.2119) (39.6395, 44.3901)	-	45.2633 (1.2819) (42.7508, 47.7758)
μ_1	-	0.7119 (1.1047) (-1.4533, 2.8771)	-	2.6528 (1.1772) (0.3455, 4.9601)
σ	20.6879 (0.8134) (19.0936, 22.2822)	20.6857 (0.8717) (18.9772, 22.3942)	-	-
σ_0	-	-	0.6765 (0.8735) (18.9644, 22.3886)	3.0723 (0.0420) (2.9900, 3.1546)
σ_1	-	-	0.2287 (0.8133) (-1.3654, 1.8228)	0.038 (0.0417) (-0.0437, 0.1197)
ξ	-0.0472 (0.0398) (-0.1252, 0.0308)	-0.0479 (0.0395) (-0.1253, 0.0295)	-0.0462 (0.0399) (-0.1244, 0.0320)	-0.0832 (0.0395) (-0.1606, -0.0058)

Table 5. Model selection criteria for monthly extreme rainfall.

Criteria	Model	Value
AIC	GEV0	-3398.338
	NSGEV1	-3401.922
	NSGEV2	-3402.260
	NSGEV3	-3409.170*
BIC	GEV0	-3394.915
	NSGEV1	-3389.958
	NSGEV2	-3389.696
	NSGEV3	-3393.465*
LRT, $D(\theta)$	GEV0 vs NSGEV1	-0.416
	GEV0 vs NSGEV2	-0.078
	GEV0 vs NSGEV3	4.416*

Table 6. Return level estimate of monthly maximum for $T=20, 50, 100$ with 95% confidence interval.

Period	$T=20$	$T=50$	$T=100$
Monthly	98.73	115.09	126.89
NSGEV3	(96.96, 101.80)	(113.04, 118.64)	(124.64, 130.79)

and 100 years later. The return level estimation of best fitted GEV model indicates the rainfall does not exceed extreme rainfall value for $T = 20, 50$ and 100 in terms of monthly extreme rainfall which is 135mm/day of the observation rainfall data.

This study has proved the capability of EVT serves as a useful analysis tool in describing the extreme events. Hence, non-stationary extreme events models with cyclic covariates could become the valuable tools to assess the future changes in extreme rainfall distribution and quantiles for engineering design and flood risk management purposes. The limitation is that other covariate trend apart from cyclic trend should be considered for this research purpose such as linear trend or quadratic trend therefore it is easier to evaluate the performance of GEV model distribution in extreme rainfall. For recommendation, this paper suggests the use of GEV approach in modelling rainfall in Malaysia for future study as it is a widely used distribution tools for extreme events which is commonly found in the past literatures. This study also recommends that the most recent rainfall data of extreme rainfall should be accounted to

carry out the future research instead of taking the past 14 years for research purpose. On the other hand, to design the infrastructure in a changing climate, it is vital to develop a nonstationary model to fit the extreme daily rainfall by modelling present trend in the observed extreme daily rainfall. There is an urgent need for the public authorities in the country to improve monitoring, modelling, and forecasting techniques in extreme daily rainfall in order to develop more effective flood protection measures and management. Findings of this study are expected to make significant contribution to the study area by increasing the understanding of nonstationarity of extreme rainfall in Klang Valley, Malaysia.

Acknowledgements

This work was supported by the Universiti Putra Malaysia, Putra-IPM grant (9619400).

REFERENCES

- [1] El Adlouni, S and Ouarda, Taha BMJ and Zhang, Xuebin and Roy, René and Bobée, Bernard. Generalized maximum likelihood estimators for the nonstationary generalized extreme value model, *Water Resources Research*, Vol. 43, No. 3, 2007.
- [2] Villarini, Gabriele and Smith, James A and Napolitano, Francesco. Nonstationary modeling of a long record of rainfall and temperature over Rome, *Advances in Water Resources*, Vol. 33, No. 10, 1256–1267, 2010.
- [3] Salas, Jose D and Obeysekera, Jayantha. Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events, *Journal of Hydrologic Engineering*, Vol. 19, No. 3, 554–568, 2014.
- [4] Milly, Paul CD and Betancourt, Julio and Falkenmark, Malin and Hirsch, Robert M and Kundzewicz, Zbigniew W and Lettenmaier, Dennis P and Stouffer, Ronald J and Dettinger, Michael D and Krysanova, Valentina. On critiques of “Stationarity is dead: Whither water management?”, *Water Resources Research*, Vol. 51, No. 9, 7785–7789, 2015.
- [5] Ganguli, Poulomi and Coulibaly, Paulin. Does Nonstationarity in Rainfall Requires Nonstationary Intensity-Duration-Frequency Curves?, *Hydrology and Earth System Sciences*, Vol. 21, 6461–6483, 2017.
- [6] Radinović, Djuro and Čurić, Mladjen, Measuring scales for daily temperature extremes, precipitation and wind velocity, *Meteorological Applications*, Vol. 21, No. 3, 461–465, 2014.
- [7] Vasiliades, L and Galiatsatou, P and Loukas, AJWRM. Nonstationary frequency analysis of annual maximum rainfall using climate covariates, *Water Resources Management*, Vol. 29, No. 2, 339–358, 2015.
- [8] Lana, X and Martínez, MD and Burgueño, A and Serra, C and Martín-Vide, J and Gómez, L. Distributions of long dry spells in the Iberian Peninsula, years 1951–1990, *International Journal of Climatology: A Journal of the Royal Meteorological Society*, Vol. 26, No. 14, 1999–2021, 2006
- [9] Fisher, Ronald Aylmer and Tippett, Leonard Henry Caleb. Limiting forms of the frequency distribution of the largest or smallest member of a sample, *Mathematical Proceedings of the Cambridge Philosophical Society*, Vol. 24, No. 2, 180–190, 1928.
- [10] Zalina, Mohd Daud and Desa, Mohd Nor Mohd and Nguyen, VTA and Kassim, Amir Hashim Mohd. Selecting a probability distribution for extreme rainfall series in Malaysia, *Water science and technology*, Vol. 45, No. 2, 63–68, 2002.
- [11] Nguyen, Truong-Huy and El Outayek, Sarah and Lim, Sun Hee and others. A systematic approach to selecting the best probability models for annual maximum rainfalls—A case study using data in Ontario (Canada), *Journal of hydrology*, Vol. 553, 49–58, 2017.
- [12] Al Mamoon, Abdullah and Rahman, Aatur. Selection of the best fit probability distribution in rainfall frequency analysis for Qatar, *Natural hazards*, Vol. 86, No.1, 281–296, 2017.
- [13] Nashwan, MOHAMED SALEM and Ismail, TARMIZI and Ahmed, KAMAL. Non-stationary analysis of extreme rainfall in peninsular Malaysia, *J. Sustain. Sci. Manag.*, Vol. 14, 17–34, 2019.
- [14] Yilmaz, AG and Hossain, I and Perera, BJC. Effect of climate change and variability on extreme rainfall intensity–frequency–duration relationships: a case study of Melbourne, *Hydrology & Earth System Sciences*, Vol. 18, No. 10, 2014.
- [15] Maraun, Douglas and Rust, Henning W and Osborn, Tim J. The annual cycle of heavy precipitation across the United Kingdom: a model based on extreme value statistics, *International Journal of Climatology: A Journal of the Royal Meteorological Society*, Vol. 29, No. 12, 1731–1744, 2009.
- [16] Agilan, V and Umamahesh, NV. What are the best covariates for developing non-stationary rainfall intensity-duration-frequency relationship?, *Advances in Water Resources*, Vol. 101, 11–22, 2017.
- [17] Golroudbary, Vahid Rahimpour and Zeng, Yijian and Mannaerts, Chris M and Su, Zhongbo Bob. Attributing seasonal variation of daily extreme precipitation events across The Netherlands, *Weather and climate extremes*, Vol. 14, 56–66, 2016.
- [18] Sharma, Shailza and Mujumdar, PP. On the relationship of daily rainfall extremes and local mean temperature, *Journal of Hydrology*, Vol. 572, 179–191, 2019.
- [19] Coles, Stuart and Bawa, Joanna and Trenner, Lesley and Dorazio, Pat. An introduction to statistical modeling of extreme values, Vol. 208, 2001.
- [20] Towler, Erin and Rajagopalan, Balaji and Gilleland, Eric and Summers, R Scott and Yates, David and Katz, Richard W. Modeling hydrologic and water quality extremes in a changing climate: A statistical approach based on extreme value theory, *Water Resources Research*, Vol. 46, No. 11, 2010.
- [21] Nasir, Siti Rashidah Mohd and others. Perspective of stakeholders on flash flood in Kuala Lumpur, *ISFRAM 2014*, 51–63, 2015.
- [22] Schafer, Joseph L and Graham, John W. Missing data: our view of the state of the art, *Psychological methods*, Vol. 7, No. 2, 147, 2002.
- [23] Chandler, Richard and Scott, Marian. Statistical methods for trend detection and analysis in the environmental sciences, 2011.
- [24] Villafuerte, Marcelino Q and Matsumoto, Jun. Significant influences of global mean temperature and ENSO on extreme rainfall in Southeast Asia, *Journal of Climate*, Vol. 28, No. 5, 1905–1919, 2015.
- [25] Gilbert, Richard O, *Statistical methods for environmental pollution monitoring*, 1987.
- [26] Akaike, Hirotugu. A new look at the statistical model identification, *IEEE transactions on automatic control*, Vol. 19, No. 6, 716–723, 1974.
- [27] Schwarz, Gideon and other. Estimating the dimension of a model, *The annals of statistics*, Vol. 6, No. 2, 461–464, 1978.
- [28] Villafuerte, Marcelino Q and Matsumoto, Jun. Significant influences of global mean temperature and ENSO on extreme rainfall in Southeast Asia, *Journal of Climate*, Vol. 28, No. 5, 1905–1919, 2015.