Stationary and Nonstationary Generalized Extreme Value Models for Monthly Maximum Rainfall in Sabah

A H Syafrina¹, A Norzaida² and J Jannatul Ain³

¹,³ Department of Mathematics, Faculty of Science Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
² UTM Razak School of Engineering and Advanced Technology, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia

Abstract. Generalized Extreme Value (GEV) model is the combination of three types of distribution class namely Gumbel, Fréchet and Weibull distributions of the Extreme Value Theory (EVT). In hydrological studies, GEV model is widely applied in the modelling of extreme rainfall. The nature of hydrological variables is highly complex, especially with the changing climate and frequent occurrences of extreme events. As such, some rainfall models assume rainfall series as stationary, while some as nonstationary. In this study, GEV models based on stationary and nonstationary data are used to assess monthly maximum rainfall data within Sabah, Malaysia and performance comparison between both GEV models is conducted. Theoretically both stationary and nonstationary GEV models are based on the same foundation, however nonstationary GEV model allows the location and scale parameters to be expressed as cyclic function of time. In this study, the stationary and the nonstationary GEV models are individually fitted to rainfall data at selected stations. Monthly maximum rainfall is blocked, and the estimated location ($\mu$), scale ($\sigma$) and shape ($\xi$) parameters are estimated by using Maximum Likelihood Estimation method. Performance of both GEV models are compared based on the Akaike Information Criterion, Bayesian Information Criterion and the likelihood ratio goodness of fit tests. Results showed that nonstationary GEV model is the best fit. The inclusion of cyclic covariates in the GEV model gives improvement on the stationary GEV model at the study region. It is also concluded that, Gumbel is identified as the significant distribution for monthly maximum rainfall in Sabah at 5% significance level.

1. Introduction

Generalized Extreme Value (GEV) model is the combination of three types of distribution class named Gumbel distribution, Fréchet distribution and Weibull distribution of the Extreme Value Theory (EVT). GEV model is widely used in modelling the extreme weather events especially in determining the best fitting for extreme rainfall. Several studies were conducted to determine the best distribution for extreme events according to the specific locations. Several studies have also reported that GEV possesses better descriptive and predictive abilities for the annual extreme rainfall as compared to other distributions. For example, in a study conducted in Ontario, Canada [1], ten distributions namely Beta-K, Beta-P, GEV, Generalized Logistic, Generalized Normal, Generalized Pareto, Gumbel, Log Pearson Type III, Pearson Type III and Wakeby distributions were compared to determine the distribution of annual maximum rainfalls within the region. It was found that the performance of GEV, Generalized Normal distribution and Pearson Type III were about equally good in describing the annual maximum rainfall series [1]. However, among the three distributions, it was found that GEV was the best distribution as GEV is fundamentally based on statistical theory of extreme random variable.
Another study of extreme rainfall has also been conducted in South Korea [2]. The results showed that GEV was suitable for describing annual extreme rainfall series based on rainfall data collected from 1961 to 2001 [2]. Similarly, [3] have also concluded that GEV was the best distribution to derive the Intensity-Duration-Frequency (IDF) curve for rainfall series in Qatar based on rainfall data collected from 1962 to 2011. In addition, a study has been conducted in Tanzania by [4] in determining the best fitting model for annual maximum daily rainfall series for over three decades between 1961 to 2014 and 1984 to 2014. The results have shown that Gumbel distribution was the most appropriate distribution. Meanwhile, exponential distribution is considered as the preferable distribution for data above 99% threshold value. Both models were well fitted for annual rainfall maxima in Tanzania. A recent study conducted by [5] found a similar result. The annual maximum daily rainfall data in Northern Algeria was collected from 1939 until 2009. As a result, GEV distribution with Fréchet distribution was found as an appropriate model for most of the stations in Algeria.

Environmental variables such as hydrological variables however, are prone to the seasonal trend. Such trend is sometimes detected in rainfall data, hence some rainfall models assume that the observed rainfall series tend to exhibit nonstationary behaviour. Mainly, hydrologists are concern with the using of appropriate rainfall models in hydrological studies especially those involving infrastructure projects and flood mitigation works. [6] stated the importance of identifying the most accurate model, which may lead to making the wrong assumption on the probability of EVT itself. There are several studies that has been conducted in identifying the suitable model for the maximum rainfall data in various locations. For instance, [7] have fitted stationary and nonstationary GEV models at 18 rainfall stations in Taiwan. Daily rainfall data were recorded from 1961 until 2010 and were blocked into annual maximum rainfall. There were three models involved in this study; a stationary GEV model where time is assumed to be constant and two nonstationary models where both models were in the form of linear and quadratic equations. This study has shown that among the 18 stations, 4 stations were well fitted into the nonstationary model and the remaining stations were well fitted into the stationary model with Gumbel distribution. Recent study by [8] in modelling the nonstationary GEV model of annual maximum rainfall in China was conducted. Data were collected from 1951 until 2013.

In a study on distribution of rainfalls in Malaysia, GEV was found to be the best model for the series of annual maximum rainfall [9]. Particularly, eight distributions namely Gamma, Generalized Normal, Generalized Pareto, Generalized Extreme Value, Gumbel, Log Pearson Type III, Pearson Type III and Wakeby have been compared to determine the most accurate and appropriate distribution for maximum rainfall estimates within Peninsular Malaysia. In another study which was conducted in Penang, Malaysia involved two types of data, monthly and half-yearly maximum rainfall from the year 2000 until 2009 [10]. Two models being employed in this study were the stationary GEV model which assumed rainfall data as constant variable and the second model is nonstationary GEV model with only location parameter assumed to be time-varying. The results showed that the nonstationary GEV model was suitable for monthly maximum rainfall while the stationary GEV was suitable for half-yearly maximum rainfall. The seasonality tests were also conducted in this study to determine the existence of a particular trend. Using the Mann-Kendall test, it was shown that there was a trend involved in monthly maximum data, however no trend was detected for the half-yearly maximum rainfall data which may due to data insufficiency. The study concluded that half-yearly maximum data can be assumed as stationary while monthly maximum data is nonstationary.

Malaysia’s rainfall is characterized by two rainy seasons associated with the Southwest Monsoon (SWM) from May to September and the Northeast Monsoon (NEM) from November to March. Substantial rainfall also occurs in the transitional periods (usually occur in April and October) between the monsoon seasons [11,12]. Due to the difference in monthly rainfall characteristics in Malaysia, this study proposes to fit the GEV model for monthly maximum rainfall in Sabah, which is located in the East of Malaysia. Sabah is the second largest state in Malaysia, sprawling over 72,500 sq. kilometres.
Despite the large size of this state, studies on rainfall distribution in this region is still lacking. Thus, the present study aims to fit the stationary and nonstationary GEV model to the rainfall data in selected rainfall stations in Sabah and compare the performance of both GEV models.

2. Data
In this study, the stationary and nonstationary GEV models are constructed based on sixteen years of historical data (2000-2015) at two stations in Sabah. The input data required by the GEV models are daily scaled rainfall data. Rainfall data were sourced from the Malaysia Drainage and Irrigation Department (DID). The selected rainfall stations in Sabah are as in Table 1.

| Station ID | Name of the station | Longitude (°) | Latitude (°) | Elevation (m) |
|------------|---------------------|---------------|--------------|---------------|
| ID5156001  | Mesapol             | 115.63        | 5.1          | 51            |
| ID6868001  | Kudat               | 116.83        | 6.89         | 16            |

3. Method
3.1. Stationary GEV model
Generalized extreme value distribution (GEV) is the family of asymptotic distributions that describe the behaviour of extreme conditions. GEV distribution consists of Gumbel, Fréchet and Weibull families which are also known as type I, II and III extreme value distributions, respectively [13]. The cumulative probability distribution for GEV is given by

\[
F(x) = \begin{cases} 
\exp\left[-\left(1 - \frac{\xi}{\alpha}(x - \mu)\right)^{\frac{1}{\xi}}\right]; & \xi \neq 0 \\
\exp\left[-\exp\left(-\frac{x - \mu}{\alpha}\right)\right]; & \xi = 0 
\end{cases} 
\] (1)

where \(\mu\), \(\alpha\) and \(\xi\) are the location, scale and shape parameters, respectively. When \(\xi < 0\), the GEV corresponds to the Weibull distribution and can be fitted to heavy tailed behaviour. The \(\xi > 0\) represents the Fréchet distribution and is used to fit left skewed samples. The case, \(\xi = 0\) corresponds to the Gumbel distribution and has moderate right tail with two parameters, scale and location [13]. The GEV parameters are estimated with the Maximum Likelihood Estimator (MLE) method as follows,

\[
L(\mu, \sigma, \xi) = -n \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{n} \log \left[1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)\right] - \sum_{i=1}^{n} \left[1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}; \quad \xi \neq 0 
\] (2)

\[
L(\mu, \sigma, \xi) = -n \log \sigma - \sum_{i=1}^{n} \log \frac{x_i - \mu}{\sigma} - \sum_{i=1}^{n} \exp \left\{-\frac{x_i - \mu}{\sigma}\right\}; \quad \xi = 0 
\] (3)

3.2 Nonstationary GEV model
The nonstationary GEV model is the basically a modification of the stationary GEV model. In nonstationary GEV model, the location \(\mu\) and scale parameter \(\sigma\), of GEV distribution are allowed to vary over time. Nonstationarity in a time series is often plausible because of periodicity or a trend. In this study, the location and scale parameters are expressed as a cyclic function of time \(t\) as represented by (4) and (5), respectively. The covariates are

\[
\mu(t) = \mu_0 + \sin(2\pi t/T) \quad \text{and} \quad \sigma(t) = \sigma_0 + \cos(2\pi t/T) 
\]
\[ \mu(t) = \beta_0 + \beta_1 \sin \frac{2\pi t}{6} + \beta_2 \cos \frac{2\pi t}{6} \]  
(4)

\[ \alpha(t) = \phi_0 + \phi_1 \sin \frac{2\pi t}{6} + \phi_2 \cos \frac{2\pi t}{6} \]  
(5)

3.3 Goodness-of-fit tests

Three goodness-of-fit tests which are Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and likelihood ratio are in this study to determine the best GEV model between stationary and nonstationary GEV models. The AIC and BIC formula are

\[ AIC = -2l + 2p \]  
(6)

\[ BIC = -2l + p \log n \]  
(7)

where \( l \) is the maximized log-likelihood for the model, \( p \) is the parameters of the model and \( n \) is the number of samples. Thus, between stationary GEV model and nonstationary GEV model which having a small value of each of the criterion is chosen as the best model.

The deviance test statistic, \( D \), for the likelihood ratio test is used to test model the following hypothesis

\( H_0 : \) Model is stationary  
\( H_1 : \) Model is nonstationary

If \( D > \chi^2_{1,0.05} = 3.84 \), model, the null hypothesis is rejected which concludes that the nonstationary GEV model is suitable for the monthly maximum rainfall.

4. Results and Analysis

In fitting the GEV model, extreme rainfall series is blocked into monthly maximum rainfall. Table 2 shows the estimated parameter of stationary GEV distribution of every rainfall station. The 95% confidence interval of the shape parameter for station Mesapol includes zero which indicates \( \xi \) is not significantly different from 0 at the 5% significance level. In contrast, the 95% confidence interval of the shape parameter for station Kudat does not includes zero which indicates \( \xi \) is significantly different from 0 at the 5% significance level. This implies that GEV with Gumbel distribution is suitable for Mesapol station while GEV with Fréchet distribution is suitable for Kudat station.

Figure 2 shows the diagnostic plots of monthly maxima rainfall which can be used to validate the fitted model. Both the probability plot and quantile plot are plotted close to the linear line which can be concluded that GEV fitted model is suitable at every station. However, there are slight differences between these plots where the probability plot tends to have a weakness as probability plot will bound to 1 as the sample size is large. Therefore, quantile plot becomes the solution to the weakness from the probability plot. The return level plot shows the return levels of rainfall amount in the next 10, 25, 50, and 100 years. The plot is close to linear whereby it represents a satisfaction of empirical estimation and the estimated curve. Finally, the correspond density estimate seems consistent with the histogram of the data. Consequently, all four diagnostics plots support the fitted GEV model.

### Table 2. Parameter values of stationary GEV distribution.

| Station | Loglikelihood, \( l \) | Parameter | Parameter value | Standard error | 95% Confidence Interval |
|---------|------------------------|-----------|----------------|----------------|-------------------------|
| Mesapol | 917.6194               | \( \mu \) | 53.329         | 1.992          | (49.43, 57.23)          |
|         |                        | \( \sigma \) | 24.464         | 1.457          | (21.61, 27.32)          |
|         |                        | \( \xi \)  | 0.005          | 0.054          | (-0.10, 0.11)           |
The estimated parameters of the nonstationary GEV model are shown in Table 3. The shape parameter, $\xi$, for Mesapol and Kudat stations are (-0.1318, 0.0838) and (-0.06552, 0.17752), respectively.

![Diagnostic plots of stationary GEV model for annual maximum rainfall in (a) Mesapol and (b) Kudat stations.](image)

**Figure 1.** Diagnostic plots of stationary GEV model for annual maximum rainfall in (a) Mesapol and (b) Kudat stations.

However, environmental data is prone to seasonal cyclicity due to the seasonality trend. The monthly maximum rainfall for both rainfall stations are plotted in Figure 2. It has been shown that periodicities or cyclic changes do appear in the data at both of rainfall stations. Thus, the cyclic covariates are proposed in this study by fitting the nonstationary GEV to the observed rainfall data. In particular, both location, $\mu$, and scale parameters, $\sigma$, are treated as a cyclic function such in (2) – (3). Table 3 shows the estimated parameters of the nonstationary GEV model. The 95% confidence intervals of the shape parameter, $\xi$, for Mesapol and Kudat stations are (-0.1318, 0.0838) and (-0.06552, 0.17752), respectively.
respectively. Both intervals include zero which indicates that the shape parameter, $\xi$ is not significantly different from 0 at the 5% significance level. This implies that GEV with Gumbel distribution is suitable for Mesapol and Kudat stations.

![Monthly maximum rainfall of Mesapol](image)

(a) Mesapol

![Monthly maximum rainfall of Kudat](image)

(b) Kudat

**Figure 2.** The scatter plot of monthly maximum rainfall in (a) Mesapol and (b) Kudat stations.

| Station  | Loglikelihood $l$ | Estimated Parameter | Estimated value | Standard error |
|----------|------------------|---------------------|-----------------|----------------|
| Mesapol  | 923.0719         | $\hat{\beta}_0$    | 66.573          | 4.511          |
|          |                  | $\hat{\beta}_1$    | 14.416          | 6.256          |
|          |                  | $\hat{\beta}_2$    | -16.824         | 6.767          |

**Table 3.** Parameter values of nonstationary GEV distribution.
Figure 3 shows the diagnostic plot of nonstationary GEV model that consists of probability plot and quantile plot of Gumbel scale for annual maxima rainfall at two rainfall stations. The nonstationary annual maxima data are usually standardized by transforming to a common Gumbel distribution to assume that $X_t$ is independent and identical random variable across the years. The probability plot and quantile plot show that both are plotted close to the linear line, which can be concluded that the accuracy of nonstationary GEV fitted model is valid. Even though probability plot and quantile plot deliver the same information about the validity of the model fitted, probability plot tends to have a weakness which can be overcome by using quantile plot.

![Figure 3. Diagnostic plots of nonstationary GEV model with cyclic covariates fitted for monthly maximum rainfall in (a) Mesapol and (b) Kudat stations.](image-url)
Finally, the best fitted GEV model is chosen based on Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and likelihood ratio test. Lowest value of AIC and BIC indicates the better model. Table 4 shows the results of AIC, BIC and likelihood ratio test values for both stationary and nonstationary GEV models. The negative value indicates a lower degree of information loss than does a positive [14]. In order to make a conclusion on which of the model is good, GEV models that produce smaller AIC and BIC values will be chosen. The nonstationary model produces the lowest AIC and BIC as compared to the stationary model. Thus, nonstationary model is the best fit for each station. This is also supported by the likelihood ratio test where the test statistic, $D$ is greater than the critical value at 5% significance level $\chi^2_{1,0.05} = 3.84$ which proves that the nonstationary model gives improvement when fitting the monthly maximum rainfall at both stations. The results are also consistent with studies done by [15] where Gumbel distribution was the best fitted model for annual maximum rainfall in Sabah.

Hence, the nonstationary GEV model with cyclic covariates on the location and scale parameters are identified as such,

(a) Mesapol

\[
\begin{align*}
\mu(t) &= 66.573 + 14.416 \sin 2\pi t/6 + 16.824 \cos 2\pi t/6 \\
\alpha(t) &= 3.304 + 0.091 \sin 2\pi t/6 - 0.073 \cos 2\pi t/6
\end{align*}
\]

(b) Kudat

\[
\begin{align*}
\mu(t) &= 46.234 + 12.761 \sin 2\pi t/6 + 1.39 \cos 2\pi t/6 \\
\alpha(t) &= 3.323 + 0.263 \sin 2\pi t/6 + 0.063 \cos 2\pi t/6
\end{align*}
\]

Table 4. Results of AIC, BIC and likelihood ratio tests.

| Goodness of fit tests | Station | Stationary GEV model | Nonstationary GEV model with cyclic covariates | Likelihood ratio test, $D$ |
|-----------------------|---------|----------------------|-----------------------------------------------|--------------------------|
| AIC                   | Mesapol | -1829.24             | -1832.14                                     | 10.905                   |
| BIC                   | Mesapol | -1830.67             | -1841.58                                     |                          |
| AIC                   | Kudat   | -1841.19             | -1855.95                                     | 22.76                    |
| BIC                   | Kudat   | -1842.62             | -1865.38                                     |                          |

5. Conclusion

Stationary and nonstationary GEV models are theoretically based on the same foundation, however nonstationary GEV model allows the location and scale parameters to be expressed as cyclic function of time. This provides the avenue to accommodate the trend or seasonality that may exist within rainfall series. In this study, it was found that both stationary and nonstationary GEV models are well fitted at the two selected rainfall stations in Sabah. Nevertheless, the nonstationary GEV model is the best fit based on the results of goodness of fit tests. Allowing the inclusion of cyclic covariates in the GEV model gives improvement on the stationary GEV model at both stations. Results also show that Gumbel is identified as the significant distribution for monthly maximum rainfall in Sabah at 5% significance level.

This study assess rainfall series using both stationary and nonstationary rainfall models. In view of the significant increasing trends of extreme rainfall events in Malaysia, it is important to concern the nonstationary nature of rainfall data to achieve better estimations of rainfall distribution. Although this study was conducted at only two stations, the results and methodology could be further extended.
future work, other heavy tail distribution could be employed together with the incorporation of cyclic covariate to describe the maximum rainfall across Malaysia.

Acknowledgement

I would like to thank Malaysian Department of Irrigation and Drainage (DID) for providing the rainfall data. This study was fully supported by Geran Putra – Inisiatif Putra Muda (IPM), Vote No: 9619400.

References

[1] Nguyen T H El Outayek S and Lim S H 2017 A systematic approach to selecting the best probability models for annual maximum rainfalls–A case study using data in Ontario (Canada) Journal of Hydrology 553 49-58
[2] Nadarajah S and Choi D 2007 Maximum daily rainfall in South Korea Journal of Earth System Science 116 4 311-320
[3] Al Mamoon A Joergensen N E Rahman A and Qasem H 2014 Derivation of new design rainfall in Qatar using L-moment based index frequency approach International Journal of Sustainable Built Environment 3 1 111-118
[4] Ngailo J T Reuder J Rutalebwa E Nyimvua S and Mesquita D S M 2016 Modelling of extreme maximum rainfall using extreme value theory for Tanzania International Journal of Scientific and Innovative Mathematical Research 4 34-45
[5] Boudrissa N Cheraitia H and Halimi L 2017 Modelling maximum daily yearly rainfall in northern Algeria using generalized extreme value distributions from 1936 to 2009 Meteorological Applications 24 1 114-119
[6] Coles S Pericchi L R and Sisson S 2003 A fully probabilistic approach to extreme rainfall modelling Journal of Hydrology 273 1-4 35-50
[7] Chu L F McAleer M and Wang S H 2013 Statistical modelling of recent changes in extreme rainfall in Taiwan International Journal of Environmental Science and Development 4 1 52
[8] Gao M Mo D and Wu X 2016 Nonstationary modeling of extreme precipitation in China Atmospheric Research 182 1-9
[9] Zalina M D Desa M N M Nguyen V T A and Kassim A H M 2002 Selecting a probability distribution for extreme rainfall series in Malaysia Water Science and Technology 45 2 63-68
[10] Hasan H and Wai Chung Y 2010 Extreme Value Modeling and Prediction of Extreme Rainfall: A Case Study of Penang In AIP Conference Proceedings Vol. 1309, No. 1, pp. 372-393 AIP
[11] Suhaila J Jemain A A Investigating the impacts of adjoining wet days on the distribution of daily rainfall amounts in peninsular Malaysia Journal of Hydrology 2009 368 17–25
[12] Wong C Liew J Yusop Z Ismail T Venneker R and Uhlenbrook S 2016 Rainfall characteristics and regionalization in Peninsular Malaysia based on a highresolution gridded data set Water 8 11 500
[13] Coles S Bawa J Trenner L and Dorazio P 2001 An introduction to statistical modeling of extreme values Vol. 208 London: Springer
[14] Baguley T 2012 Serious stats: A guide to advanced statistics for the behavioral sciences. Palgrave Macmillan
[15] Nicklós Jefrín Nurmin Bolong Justin Sentian Ismail Abustan Thamer Ahmad Mohammad Janice Lynn Ayog 2018 Comparison of GEV and Gumble’s Distribution for Development of Intensity Duration Frequency Curve for Flood Prone Area in Sabah. Malaysian Journal of Geosciences 2 1 42-44