ESTEEM Academic Journal
Vol. 13, Special Issue, August 2017, 107-117

ANALYSING EXTREME TEMPERATURE SCENARIOS IN PENINSULAR MALAYSIA USING GENERALIZED EXTREME VALUE (GEV)

Nur Hanim Mohd Salleh[1] and Husna Hasan[2]
School of Mathematical Sciences, Universiti Sains Malaysia, 11800, USM, Penang, Malaysia
**corresponding author: [1]nurhanim_sh@yahoo.com; [2]husnahasan@usm.my*

ARTICLE HISTORY

Received
22 May 2017

Received in revised form
16 June 2017

Accepted
30 June 2017

ABSTRACT

Climate change is considered to be one of the biggest crisis which affects human life and nature. The anthropogenic or human factors such as land conversion, industrialization and transportation release greenhouse gasses amplify warming in air temperature. The objective of this study is to analyze the extreme temperature events at nine selected states in Peninsular Malaysia using Generalized Extreme Value (GEV) distribution. It also aims to predict the return level of the maximum temperature at different selection period. The estimation of parameters is determined using Maximum Likelihood Estimator (MLE) method. In this study, stationary and non-stationary GEV models are considered. Mann-Kendall trend test is applied to detect stationarity in a series of maximum temperature data. The result indicates that non-stationary model is preferred for Kuala Terengganu, Muadzam Shah and Senai stations. By evaluating the return period of T-years for each station, the result of the estimated return levels showed that the temperatures for all stations are increasing over 125 years except for the non-stationary stations.

Keywords: *extreme maximum temperature; generalized extreme value; return level; maximum likelihood estimation; Mann-Kendall.*

1. INTRODUCTION

Today, climate change is considered to be one of the biggest crisis which affects human life and nature. In its Fifth Assessment Report, the Intergovernmental Panel on Climate Change (IPCC), a group of scientific experts concluded there is an increase in global average surface temperature from 1951 to 2010 which was caused by anthropogenic factors (IPCC, 2014). The anthropogenic or human factors such as land conversion, industrialization and transportation release greenhouse gasses amplify warming in air temperature. Continued climate change could result in more extreme events like droughts, heat waves and floods.

Extreme events or rare occurrences are usually indicated by the presence of an observation which is very low (minimum) and very high (maximum). The behavior of the extreme events can be studied using Extreme Value Theory (EVT). EVT concerns the stochastic behavior of the extreme values in a single process. As suggested by Fisher and Tippett (1928), the behavior of the maxima can be explained by the three extreme value distributions namely Gumbel, Frechet and negative Weibull. The first application of the extreme value distribution was probably done by Fuller in 1914. However, the study on the extreme value distributions could be traced back to the work done by Bernoulli in 1709 as suggested by Kotz and

p-ISSN 1675-7939; e-ISSN 2289-4934

Nadarajah (2000). Nowadays, The EVT approach has been widely applied in many areas of study such as insurance, economics, hydrology and climatology.

Malaysia is a country located in Southeast Asia with two distinct parts which are Peninsular Malaysia and East Malaysia. Although this country could be considered as a free zone climate-related disasters such as earthquakes and volcanoes, lately, Malaysia is also suffering from the effects of climate change such as droughts and floods. These effects are related to the emission of greenhouse gasses mainly carbon dioxide that increase significantly corresponds to the rapid industrialization and economic growth (Begum & Pereira, 2011). As a developing country, the socio-economic development activities in Malaysia depend on the climate conditions. According to MOSTE (2000), an increase in temperature could put national food security at greater risk as every 1°C temperature rise may cause 10% reduction in rice yields. The temperature in Malaysia is predicted to continue on an increasing trend. Studies on the absolute temperature indices indicate significant warming trends in both lowest annual minimum temperature (Hasan & Mohd Salleh, 2015) and lowest annual average temperature indices (Mohd Salleh et al., 2015). Furthermore, modeling results estimate that the extreme temperature in several states in Malaysia may become warmer by mid and end of the century (Hasan et al., 2014).

Appropriate temperature modeling and prediction are necessary to reduce the negative impact of climate change and variability in this country. One of the methods that can be used to minimize that impact is by modeling the extreme temperature events using Generalized Extreme Value (GEV) distribution. The application of GEV distribution method in modeling the maximum temperature events in Malaysia has been studied by Hasan et al. (2014). The study was dealing with the daily average temperature data obtained from National Climatic Data Center. In the year 2012, Hasan et al. (2012) was probably the first to fit the GEV distribution on the daily maximum temperature data obtained from Malaysian Meteorological Department. However, the study was limited to only Penang state, Malaysia. For more comprehensive analysis, this paper aims to model the annual maximum temperature of daily maximum temperature data at nine selected states in Peninsular Malaysia using GEV distribution. It also attempts to predict the return level of the maximum temperature at different selection period.

2. DATA AND STUDY AREA

The daily maximum temperature (TMax) data used in this study are obtained from Malaysian Meteorological Department. The data are recorded at nine meteorological stations located in Peninsular Malaysia. Three of the stations which are Chuping (CP), Alor Setar (AS) and Bayan Lepas (BL) stations are located in the northern part of Peninsular Malaysia while Kuala Terengganu (KT) station is located in the eastern region of Peninsular Malaysia. Only one station located at the central part of Peninsular Malaysia which is Kuala Lumpur International Airport (KLIA) station. The last four stations which are Malacca (MC), Muadzam Shah (MS), Mersing (MR) and Senai (SN) stations located in the southern region of Peninsular Malaysia. The temperature data for all stations are measured in Degree Celsius (°C) and recorded from 1994 to 2013 except for KLIA station which is observed from 1998 to 2013.

p-ISSN 1675-7939; e-ISSN 2289-4934

ESTEEM Academic Journal
Vol. 13, Special Issue, August 2017, 107-117

3. RESULT AND DISCUSSIONS

3.1 Preliminary Analysis

Movements of the extreme value can be identified by using block maxima method. The value of maximum observations are blocked into selected period such as annually, monthly, weekly and other selection periods before fitted to the gev distribution. As a first approach to study trends in the maximum temperature annually, the data are blocked into annual maximum. Then, the non-parametric mann-kendall (mk) trend test is applied to investigate the stationary assumption of the classical gev distribution. To avoid the problem roused by data skew, this non-parametric test is considered over the parametric once (smith, 2000).

3.2 Generalized Extreme Value (GEV) Distribution

The gev distribution is a three-parameter model that combines the gumbel, frechet and weibull distributions, also known as the extreme value distribution of type i, ii and iii. These three types of distribution have different forms of behavior in the tails. The gumbel distribution has a light tail, meaning that although the maximum can take on infinitely high values, the probability of obtaining such levels become small exponentially, as described by

$$G(z) = exp\left\{-exp\left[-\left(z-\mu(z-\mu)/\sigma\right)\right]\right\}, \qquad -\infty < z < \infty \; .$$

μ is the location parameter and σ is the scale parameter. In contrast, the fréchet distribution has a heavy tail that is,

$$G(z) = \begin{cases} 0 & z \le \mu \\ exp\left[-\left((z-\mu)/\sigma\right)^{-\xi}\right], & z > \mu \end{cases}$$

ξ is the shape parameter. The weibull distribution is bounded above meaning that there is a finite value which the maximum cannot exceed is as follow:

$$G(z) = \begin{cases} exp\left\{-\left[-\left((z-\mu)/\sigma\right)^{\xi}\right]\right\}, & z < \mu \\ 1, & z \ge \mu \end{cases}$$

The gev distribution with cumulative function (coles, 2001)

$$G(z;\mu,\sigma,\xi) = exp\left\{-\left[1+\xi\left((z-\mu)/\sigma\right)\right]^{-1/\xi}\right\}, \text{ for } \xi \ne 0$$

Where $\mu \in R$, $\sigma > 0$ and $\xi \in R$, defined on $\left[1+\xi(z-\mu)/\sigma\right] > 0$ has the ability to describe all three types of tail behavior; it will follow either the gumbel, fréchet or weibull distribution for $\xi = 0$, $\xi > 0$ and $\xi < 0$, respectively.

3.3 Model Selection and Likelihood Ratio Test

p-ISSN 1675-7939; e-ISSN 2289-4934

Two models, namely, stationary model 1 and non-stationary model 2 are considered. Model 1: μ,σ and ξ are constants is a classical gev model with all three parameters considered to be time-independent. Model 2: $\mu(t)=\beta_0+\beta_1 t$, σ and ξ are constants where t refers to units of the selection period, is a time-dependent model which variations in time are accounted for through time t through a linear trend.

To determine the best fitting model between model 1 and model 2, likelihood ratio (lr) test is used. The lr test statistic, defined as $\gamma=-2ln\left(\frac{L_0}{L_1}\right)$ has a chi-square distribution with 1 degree of freedom (since the number of the parameters differ by one). L_0 is the maximum likelihood for the three-parameter model 1 and the alternative model L_1 is the maximum likelihood for the four-parameter model 2. The three-parameter model (model 1) is preferred if $\gamma<\chi^2_{1,0.90}=2.706$ or else the four-parameter model (model 2) is preferred.

3.4 Maximum Likelihood Estimation (MLE) and Model Diagnostics

The maximum likelihood estimation (mle) provides a standard way to estimate the parameters of a gev distribution. It also offers a more consistent approach to parameter estimation problem and shows less bias (shukla, 2010). The parameters for both model 1 and model 2 are estimated in this study using the mle method from r (r development core team, 2009) and extreme package (gilleland, 2016).

The diagnostic plots employed for judging the goodness of fit for the fitted gev models are probability, quantile, return level and density plot. The data would line up on the diagonal of the probability and quantile plots in the case of a perfect fit. The return level plot exhibits the return period compared with the return level with an estimated 95% confidence interval. However, some modification is needed for the non-stationary cases (model 2) due to the lack of homogeneity in the distribution assumptions for each observation (coles, 2001). For the non-stationary case, the plots are applied to the residuals of the data.

3.5 Return Level Estimation

Return level of an extreme event, z_p is the level that is expected to be exceeded on average once every $1/p$-year. In extreme value terminology, z_p is the return level associated with $1/p$-year return period (garcía-cueto & santillán-soto, 2012). The return level is derived from the gev distribution by inverting its cumulative function and then solving for the return level. Estimates of extreme quantiles of the annual maximum distribution can be obtained by the equation:

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

Where p is the probability such that $G(z_p)=1-p$, $y_p=-\log(1-p)$ and $0<p<1$. The confidence interval for the return level can be attained via the profile likelihood.

p-ISSN 1675-7939; e-ISSN 2289-4934

3.6 Descriptive Statistics

The annual extreme temperature can be characterized by its minimum and maximum, mean and standard deviation. Table 1 provides a quantitative comparison between the meteorological stations and it can be concluded that the annual lowest maximum temperature is observed at Bayan Lepas station. Next, Chuping station records the highest value of maximum and mean of annual maximum daily temperature. Besides, the standard deviation for Alor Setar and Chuping stations are found to be higher as compared to the other stations. This result may indicate that the amounts of extreme rainfall in those two stations are relatively more spread as compared to the other regions of Peninsular Malaysia.

Table 1: Descriptive statistics for the annual maximum temperature

Station	N	Min	Max	Mean	SD
Chuping	20	35.4	39.1	36.84	1.825
Alor Setar	20	33.3	35.6	34.52	0.731
Bayan Lepas	20	35.9	40.1	37.15	1.014
Kuala Terengganu	20	33.5	35.8	34.42	0.622
KLIA	16	34.5	37.2	35.76	0.790
Malacca	20	34.2	38.0	35.52	0.994
Mersing	20	33.5	36.2	34.80	0.660
Muadzam Shah	20	35.3	37.3	36.10	0.569
Senai	20	34.1	37.2	35.24	0.664

Table 2: Stations with significant trend

Station	Test Statistic	*p*-value	Trend
Kuala Terengganu	-2.280	0.011	Decreasing
Mersing	1.337	0.091	Increasing
Muadzam Shah	-1.793	0.037	Decreasing
Senai	-1.566	0.059	Decreasing

p-ISSN 1675-7939; e-ISSN 2289-4934

3.7 Testing for Trend

The next analysis involves assessing the existence of trend using scatter plot and Mann-Kendall (MK) trend test. The MK trend test result illustrated in Table 2 revealed that there are four stations have a significant trend. Kuala Terengganu, Muadzam Shah and Senai stations have a significant downward trend with the test statistics of -2.280, -1.793 and -1.566 and the *p*-value of 0.011, 0.037 and 0.059, respectively. Conversely, Mersing station exhibits a significant upward trend with the test statistic of 1.337 and *p*-value of 0.091.

3.8 Parameter Estimation and Model Selection

The estimation of parameters of both models is conducted using the MLE method. The estimation values of the stationary model (Model 1) are given in Table 3. The values in the bracket are the parameter standard error. Since there are four stations shows non-stationarity in the data set, subsequent analysis involves the non-stationary model (Model 2) is performed.

Table 3: Parameter estimation for stationary model 1

Station	μ (se)	σ (se)	ξ (se)
Chuping	36.391(0.187)	0.736(0.140)	0.035(0.191)
Alor Setar	34.377(0.218)	0.810(0.207)	-0.601(0.309)
Bayan Lepas	36.660(0.181)	0.682(0.142)	0.126(0.226)
Kuala Terengganu	34.155(0.131)	0.513(0.095)	-0.070(0.182)
KLIA	35.455(0.208)	0.704(0.155)	-0.189(0.255)
Malacca	35.058(0.162)	0.641(0.124)	0.135(0.178)
Mersing	34.538(0.146)	0.587(0.100)	-0.163(0.149)
Muadzam Shah	35.824(0.109)	0.412(0.084)	0.076(0.255)
Senai	3.967(0.138)	0.564(0.095)	-0.090(0.123)

p-ISSN 1675-7939; e-ISSN 2289-4934

Table 4: Parameter estimation for non-stationary model 2

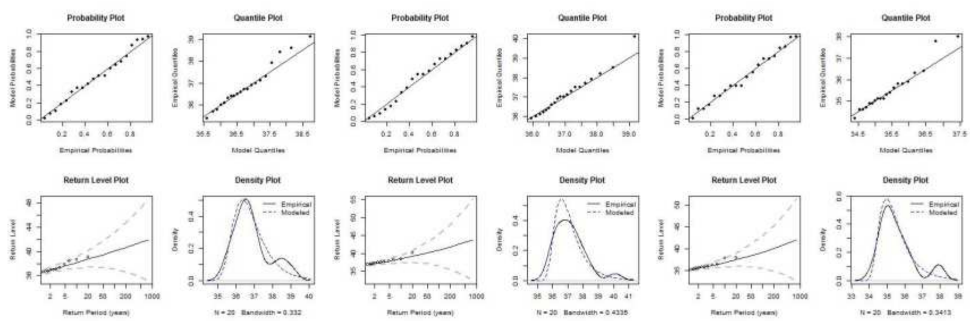
Station	β_0 (se)	β_1 (se)	σ (se)	ξ (se)
Kuala Terengganu	34.768 (0.298)	-0.052 (0.023)	0.483 (0.091)	-0.206 (0.199)
Mersing	34.377 (0.282)	0.016 (0.024)	0.589 (0.098)	-0.180 (0.132)
Muadzam Shah	36.075 (0.192)	-0.022 (0.013)	0.386 (0.076)	0.053 (0.209)
Senai	35.459 (0.234)	-0.047 (0.021)	0.482 (0.090)	-0.012 (0.177)

The parameter estimation values of the non-stationary model are presented in Table 4. The parameter β_1 in Model 2 corresponds to the annual rate of change in annual maximum temperature. Using Likelihood Ratio (LR) test, the best fitting model for Kuala Terengganu, Mersing, Muadzam Shah and Senai stations are determined.It is found that only three stations favor Model 2 which are Kuala Terengganu, Muadzam Shah and Senai stations with the *p*-value less than 0.10.

An analysis of the shape parameter obtained for both models and all meteorological stations show that this parameter is positive at four stations (Alor Setar, Chuping, Malacca and Muadzam Shah); thus, the distribution class corresponding to the data is the Frechet distribution. According to the negative value of the shape parameter at other 5 stations (Bayan Lepas, Kuala Terengganu, KLIA, Senai), the adequate distribution class is the Weibull distribution.

3.9 Model Diagnostics

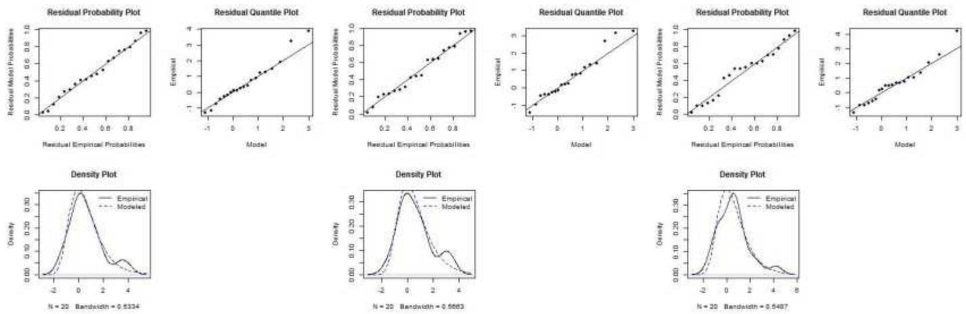
Figure 1 displays the probability, quantile, return-level and the density plots for some stations which belong to stationary (Model 1). When incorporating the location parameter as a function of time, the plots are applied to the residuals of the data. The residual probability and residual quantile plots for the non-stationary model are illustrated in Figure 2. The data mostly line up on the diagonal of the probability and quantile plots with small deviations from the straight line. This result suggests that the model assumption is valid for the data plotted.



Alor Setar Chuping Malacca

Figure 1: Diagnostic plots for stations stationary model

p-ISSN 1675-7939; e-ISSN 2289-4934



Kuala Terengganu Muadzam Shah Senai

Figure 2: Diagnostic plots for non-stationary model

3.10 Return Level Estimation

The estimation of the return level is needed for the purpose of design and risk assessment under environmental change. Using the same method applied by Hasan et al. (2014), for the stations with the non-stationary model, the data are transformed into stationary by removing the trend. Table 5 shows the results of estimated T-year return levels and 95% confidence intervals for T = 10, 50, 100 and 125 return periods. The estimation of the 95% confidence interval is done using profile likelihood.

It can be seen from Table 5 that the return levels for the annual maximum temperature at all stations gradually increase for higher and higher return periods except for the three stations (Kuala Terengganu, Muadzam Shah and Senai) which belong to the non-stationary model. Within the next 50 years, it is predicted that a maximum temperature event will reappear for Alor Setar, Bayan Lepas, Chuping, KLIA, Malacca, Mersing and Senai stations.

p-ISSN 1675-7939; e-ISSN 2289-4934

Table 5: Return level estimates for stationary model

Station	Return Period, T (years)				
	10	**25**	**50**	**100**	**125**
CP	38.11 (37.22, 39.01)	38.88 (37.36, 40.40)	39.47 (37.27, 41.66)	40.06 (37.01, 43.11)	40.26 (36.90, 43.62)
AS	35.37 (35.15, 35.60)	35.53 (35.31, 35. 75)	35.60 (35.30, 35.89)	35.64 (35.64, 36.01)	35.65 (35.26, 36.05)
BL	38.44 (37.40, 39.47)	39.35 (37.40, 41.30)	40.10 (37.09, 43.10)	40.91 (36.50, 45.33)	41.19 (36.24, 46.14)
KT	35.12 (34.79, 35.99)	34.6 (34.16, 35.74)	33.47 (32.9, 34.76)	31.01 (30.29, 32.44)	29.75 (30.29, 32.44)
KLIA	36.74 (36.17, 37.32)	37.14 (36.27, 38.02)	37.40 (36.19, 38.60)	37.62 (36.04, 39.19)	37.68 (35.97, 39.39)
MC	36.75 (35.77, 37.72)	37.62 (35.89, 39.36)	38.35 (35.76, 40.94)	39.15 (35.42, 42.88)	39.42 (32.26, 43.58)
MR	35.64 (35.21, 36.08)	36.00 (35.39, 36.61)	36.23 (35.45, 37.02)	36.44 (35.45, 37.43)	36.50 (35.45, 37.56)
MS	36.78 (36.3, 37.71)	36.88 (36.05, 38.23)	36.65 (35.45, 38.32)	35.89 (34.22, 37.9)	35.45 (33.6, 37.57)
SN	36.06 (35.55, 37.14)	35.80 (34.98, 37.31)	34.95 (33.81, 36.79)	32.92 (31.38, 35.09)	31.85 (30.17, 34.12)

Comparing these results with the analysis result from our previous study (Hasan et al., 2014), the return level estimates for annual maximum temperature based on daily average temperature are more varied than the daily maximum temperature data. Those result in the previous study exhibited that the maximum temperatures for the stations were expected to reappear within the next 50, 100 and 125 years return period. However, the maximum temperature events for all stationary stations investigated in this study are expected to return within the next 50 years.

For Bayan Lepas station, the estimation of the return levels in previous and current studies showed that this station will enter its maximum temperature state within the next 50 years. On the contrary, a maximum temperature event for Malacca station is believed to re-emerge within the next 50 and 125 years period based on the daily maximum temperature (current) and daily average temperature (previous) analysis respectively.

p-ISSN 1675-7939; e-ISSN 2289-4934

4. CONCLUSIONS

This study investigates the extreme temperature scenarios at nine meteorological stations in Peninsular Malaysia. In the study, Extreme Value Theory was successfully applied to the annual maximum of daily maximum temperature data. Two models; stationary and non-stationary GEV model was fitted using a block maxima approach. Likelihood Ratio Test indicated that three stations, namely, Kuala Terengganu, Muadzam Shah and Senai favor non-stationary model. The diagnostics plots confirm the adequacy of these models for the data analyzed.

As discussed before, the return period for the maximum temperature event obtained from this study is slightly different compared to our previous study which is based on daily average temperature (Hasan et al., 2014). Generally, the return level estimates in both studies showed that majority of the stations exhibit increasing trends over 125 years which. The increase in warming trend could be due to natural factors such as El Nino-Southern Oscillation (ENSO) and The Indian Ocean Dipole (IOD) (Tangang et al., 2007).

In this study, all six stations from the stationary model are expected to re-enter their maximum temperature state within the next 50 years. On the other hand, the stations analyzed in the previous study were predicted to enter their maximum temperature states within varies of return period. For further study, other covariates such as wind speed and rainfall may be included in modeling the non-stationary GEV model.

5. ACKNOWLEDGEMENT

This work was supported by School of Mathematical Sciences, Universiti Sains Malaysia. The authors would like to thank Malaysian Meteorological Department for providing the temperature data.

REFERENCES

Begum, R. A. & Pereira, J. J. (2011). Corporate perceptions of climate change in Malaysia. *African Journal of Business Management*, 5 (11), 4299-4305.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values.* Great Britain: Springer.

Fisher, R. A. & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 24, 180–290.

García-Cueto, O. R. & Santillán-Soto, N. (2012). Modeling Extreme Climate Events: Two Case Studies in Mexico, Climate Models, Dr. Leonard Druyan (Ed.). Retrieved November, 18. 2013 from http://www.intechopen.com/books/climate-models/modeling-extreme-climate-events-two-casestudies-in-mexico.

p-ISSN 1675-7939; e-ISSN 2289-4934

Gilleland, E. (February, 2016). Package 'extRemes'. Retrieved from https://cran.r-project.org/web/packages/extRemes/extRemes.pdf

Hasan, H., Ahmad Radi, N.F. & Kassim, S. (2012). Modeling Extreme Temperature Using Generalized Extreme Value (GEV) Distribution: A Case Study of Penang. *World Congress on Engineering,* 1, 181-186.

Hasan, H. & Mohd Salleh, N. H. (2015). Extreme temperature indices analyses: A case study of five meteorological stations in Peninsular Malaysia. *AIP Conference Proceeding,* 1682, 050011.

Hasan, H., Mohd Salleh, N.H. & Kassim, S. (2014). Stationary and Non-Stationary Extreme Value Modeling of Extreme Temperature in Malaysia. *AIP Conference Proceeding,* 1613, 355-367.

IPCC (2014). Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland. pp 151.

Kotz, S. & Nadarajah, S. (2000). Extreme Value Distributions: Theory and Applications. *London Imperial College Press*.

Ministry of Science, Technology and Environment (MOSTE) (2000). Malaysia Initial National Communication. pp. 131.

Mohd Salleh, N. H., Hasan, H. & Kassim, S. (2015). Trends in Temperature Extremes across Malaysia. *Advances in Environmental Biology,* 9 (27), 174-181.

R Development Core Team. (2009). *R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing*, Vienna, Austria.

Shukla, R. K. (2010). On the proficient use of GEV distribution: a case study of subtropical monsoon region in India. Anale. *Seria Informatica*, 8, 81-92.

Smith L. C. (2000). Trends in Russian Arctic river-ice formation and breakup, 1917 to 1994. *Physical Geography*, 21 (1), 46-56.

Tangang, F. T., Juneng, L. & Ahmad, S. (2007). Trend and Interannual Variability of Temperature in Malaysia: 1961–2002. *Theoretical and Applied Climatology*, 89, 127-141.

p-ISSN 1675-7939; e-ISSN 2289-4934