Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/100103

An Investigation of Parent Distributions and Long-Term Trends of Average Maximum and Minimum Temperature in the Limpopo Province of South Africa

*Anna M. Seimela and Daniel Maposa**

Department of Statistics and Operations Research, University of Limpopo, Polokwane, South Africa

Received: 14 Apr. 2020, Revised: 11 Jun. 2020, Accepted: 17 Jun. 2020
Published online: 1 Mar. 2021

Abstract: In studying natural hazards or disasters that occur due to temperature extremes such as heat waves and cold waves, it is crucial to understand the underlying distributions of the maximum and minimum temperatures at a particular site or region. The present study intends to investigate the parent distributions of maximum and minimum temperatures at various sites in the Limpopo province of South Africa. The parent distributions were investigated at four meteorological stations in Limpopo province, namely: Mara (1949-2018), Messina (1934-2009), Polokwane (1932-2018) and Thabazimbi (1994-2018). Four candidate parent distributions; normal, lognormal, gamma and Weibull distributions, were fitted to the average monthly maximum and minimum daily temperatures. Prior to the selection of the parent distributions, the data set at each station was subjected to normality test using the Shapiro-Wilk (SW) and Jarque-Bera (JB) tests. The normality tests have revealed that the maximum and minimum temperature data series at all the stations do not follow a normal distribution. Akaike information criterion (AIC) and Bayesian information criterion (BIC) were used to select the best fitting distribution at a particular site. The parent distribution with the lowest value of AIC and BIC was chosen as the best fitting distribution for the data. Goodness-of-fit diagnostic tests such as the Q-Q plots, P-P plots, empirical and theoretical density and cumulative distribution function (CDF) plots were conducted on the selected and/or competing candidate distributions. The findings reveal that short-tailed distributions in the Weibull domain of attraction, which include the Weibull distribution, are the best fitting parent distributions for both maximum and minimum temperature series at all the stations. Furthermore, a generalised extreme value (GEV) distribution was fitted to all the data set for each station in order to establish and validate whether the Weibull family is indeed a good fit to the data. The GEV distribution findings further confirmed the Weibull class as the parent distribution for all the stations in the study. The Mann-Kendall test and time series plots trend analysis findings have shown that there is a downward and upward long-term trend in minimum and maximum temperature data, respectively. Future studies will look into the possibility of applying both univariate and multivariate extreme value theory (MEVT) techniques to investigate further whether these climatic changes in mean monthly temperature can indeed be attributed to global warming and other natural modes of interdecadal variability.

Keywords: Limpopo province, long-term trend, maximum and minimum temperatures, parent distributions

1 Introduction

Over the last five decades, South Africa has experienced a significant increase in the mean annual temperatures with hot and cold extremes increasing and decreasing, respectively, in frequency across the country [1]. Temperature is one of the main climatic elements that can indicate climate change as climate change seems to be one of the most important issues in the recent two decades [2,3]. Climate change is a measurable reality posing significant social, economic and environmental risks and challenges globally [4]. Global warming and its associated increase in temperature extremes pose a substantial challenge for natural systems.

According to [5], human activities are the major causes of climate change because of the burning of fossil fuel which produces gases like carbon dioxide, methane and nitrous oxides which lead to global warming. Climate change can influence nature and threaten humans in different aspects of life economically and socially [6]. It is widely believed that the changing temperature due to global warming is permanently changing the earth's climate. That is, an increase in

* Corresponding author e-mail: danmaposa@gmail.com

© 2021 NSP
Natural Sciences Publishing Cor.

temperature is likely to lead to a global increase in drought conditions, increase in agricultural demand and decrease in water supplies due to evapotranspiration [7,6].

The temperature extremes and frequent flooding are affecting agricultural production leading to scarce food and water resources, which is a big threat to a country like South Africa, where the population is rapidly growing [8]. South Africa is also concerned about public health around extreme hot events such as heat waves rather than extreme cold events and how the impact of these events may change in the future [9]. In the past four decades (1980-2015), Southern Africa experienced 491 climate disasters (meteorological, hydrological, and climatological) that resulted in 110 978 deaths, left 2.49 million people homeless and affected an estimated 140 million people [10].

Kruger and Shongwe [11] investigated temperature trends for the period 1960-2003 in South Africa using time series analysis and found that the average temperature is increasing in South Africa. Tshiala et al. [12] analysed the temperature trends over Limpopo province of South Africa for the period 1950-1999 using Mann-Kendall test statistic. The seasonal trends showed variability in mean temperature with an increase of 0.18^0C per decade in winter and 0.09^0C per decade in summer. Kruger and Sekele [13] investigated the trends of daily maximum and minimum extreme temperature indices for 28 weather stations in South Africa for the period 1962-2009 using cluster analysis. The results for the maximum temperature indicated an increase in warm extremes, while the minimum temperature indicated decreases in cold extremes.

The present study is aimed at investigating the parent distributions and long-term trends of the average maximum and minimum temperature in the Limpopo province of South Africa. As the previous studies have indicated, the Limpopo province is the hottest province in South Africa and the changing climate attributed to global warming is increasing this burden of warm extremes in the province [14,15]. Nemukula et al. [15] conducted a study on bivariate threshold excess modelling of the extreme high temperatures in Limpopo province of South Africa in the extreme value theory (EVT) realisation. Their study did not attempt to establish the parent distributions of the temperature series in the Limpopo province. Literature on the parent family of limiting distributions and long-term trends of temperature extremes in the Limpopo province and majority of the provinces of South Africa is scarce. Therefore, it becomes imperative to study the parent distributions and long-term trends of these temperature extremes. Another gap intended to be closed in this study is the use of the generalised extreme value (GEV) distribution as a validation technique or goodness-of-fit test instead of its common use in the block maxima realisation [16]. The knowledge gained from this study will form a benchmark in the application of some statistical techniques to parent distributions model selection and assist in disaster risk reduction through climate change management and also help economically by reducing and saving the amount of energy supply required for air conditioning and cooling systems in the province.

The paper is organised as follows: Section 2 gives a detailed research methodology which includes the data source, study area and statistical techniques applied in the analysis of the data. Section 3 presents the summarised results of the study in the form of tables and figures, as well as a comprehensive discussion of the results. Section 4 gives a detailed conclusion of the study.

2 Research Methodology

This section presents the data source, study area and statistical techniques used to analyse the data.

2.1 Data Source and Study Area

The study uses time series secondary data. The average maximum and minimum temperature data measured in degrees Celsius (^0C) was obtained from the South Africa Weather Service (SAWS) database. The data covers the following stations of the Limpopo province of South Africa: Mara (1949-2018), Messina (1934-2009), Polokwane (1932-2018) and Thabazimbi (1994-2018).

2.2 Testing for Normality

This subsection presents the theoretical framework of the normality tests used in this study.

© 2021 NSP
Natural Sciences Publishing Cor.

2.2.1 Shapiro-Wilk Test

Shapiro-Wilk tests the null hypothesis that a sample $X_1,...,X_n$ come from a normally distributed population.

The test statistic is given by

$$W = \frac{(\sum_{i=1}^{n}(a_i)(x_i))^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}, \tag{1}$$

where x_i is the i^{th} order statistics. That is, the i^{th} smallest number in the sample,

$$\bar{x} = \frac{(x_i+...+x_n)}{n}. \tag{2}$$

The constants $a_i s$ in (1) are given by

$$(a_1,...,a_n) = \frac{m^t(V^{-1})}{(m^t(V^{-1}(V^{-1})m)^{1/2}}, \tag{3}$$

where

$$m = (m_1,...,m_n)^T, \tag{4}$$

$m_1,...,m_n$ are the expected values of the order statistics of independent and identically distributed (iid) random variables from the standard normal distribution, and V in (3) is the covariance matrix of those order statistics [17].

2.2.2 Jarque-Bera Test

Jarque-Bera (JB) test is defined as:

$$JB = \frac{n-k+1}{6}(s^2+\frac{1}{4}(c-3)^2), \tag{5}$$

where

n is the number of observations,
s is the sample skewness,
c is the sample kurtosis,
k is the number of regressors [18,19].

Additionally, the sample skewness s and sample kurtosis c are given by

$$s = \frac{\hat{\mu_3}}{\hat{\sigma_3}}, \tag{6}$$

$$c = \frac{\hat{\mu_4}}{\hat{\sigma_4}}, \tag{7}$$

where $\hat{\mu_3}$ and $\hat{\mu_4}$ are the estimates of the third and fourth central moments [18,19].

2.3 Parent Distributions

This subsection presents the theoretical framework of the parent distributions investigated in this study.

2.3.1 Normal Distribution

A normal distribution is symmetrical and has a bell-shaped density curve with a single peak [20]. The normal density function is given by

$$f(x|\sigma,\mu) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{-(x-\mu)}{2\sigma^2}\right), \quad \text{with} \quad \sigma > 0, \tag{8}$$

where μ is the mean, and σ is the standard deviation, known as the location and scale parameters of the distribution, respectively [21].

© 2021 NSP
Natural Sciences Publishing Cor.

2.3.2 Lognormal Distribution

The lognormal distribution probability density function (PDF) is given by

$$f(x|\sigma,\mu) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log x - \mu)}{2\sigma^2}\right), \tag{9}$$

with μ and σ as mean and standard deviation, known as the location and scale parameters of the distribution, respectively [21].

2.3.3 Weibull Distribution

The Weibull distribution PDF is given by

$$f(x|\alpha,\beta) = \left(\frac{\alpha}{\beta^{-\alpha}}\right)(x^{\alpha-1}) \exp\left(-\left(\frac{x}{\beta}\right)^{\alpha}\right), \tag{10}$$

with α and β as shape and scale parameters, respectively.

2.3.4 Gamma Distribution

The PDF of the gamma distribution is given by

$$f(x|\alpha,\beta) = \frac{1}{\beta^{\alpha}}\left(\frac{1}{\Gamma(\alpha)}\right)(x^{\alpha-1}) \exp\left(\frac{x}{\beta}\right), \tag{11}$$

with α and β as the shape and scale parameters, respectively.

2.4 Maximum Likelihood Parameter Estimation

Maximum likelihood (ML) estimation is mostly used for fitting probability distributions to data. For an assumed distribution, it produces asymptotically efficient and unbiased estimates [22].
Let $X_1, X_2, X_3, \ldots, X_n$ be a joint density denoted by

$$f_{\theta}(x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots, x_n|\theta). \tag{12}$$

Given observed value $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, the likelihood of θ is the function

$$L(\theta|X) = f(x_1, x_2, \ldots, x_n|\theta), \tag{13}$$

which is considered as a function of θ.

The ML estimate of θ is that value of θ that maximises $lik(\theta)$. If the X_i's are iid, then the likelihood simplifies to

$$likelihood = L(\theta|X) = \prod_{i=1}^{n} f(x_i|\theta), \tag{14}$$

where $L(\theta|X)$ is the likelihood of the set of parameter θ given the observation X and $\prod_{i=1}^{n} f(x_i|\theta)$ is the probability density function of the probability model.

Maximising the log likelihood, we get

$$loglikelihood = \ln[L(\theta|X)] = \sum_{i=1}^{n} \ln[f(x_i|\theta)]. \tag{15}$$

2.5 Model Selection Criterion

The goodness-of-fit of the distributions is assessed using Akaike's information criterion (AIC) and Bayesian information criterion (BIC).

© 2021 NSP
Natural Sciences Publishing Cor.

2.5.1 Akaike's Information Criterion

The models are selected according to the values of AIC. The model or distribution with the lowest value of AIC is chosen to be the best.

The AIC is defined as:

$$AIC = 2K - 2\log(L), \tag{16}$$

where
K is the number of parameters in the statistical model,
L is the maximum value of the likelihood function for the estimated models.

2.5.2 Bayesian Information Criterion

The BIC assesses goodness-of-fit of a distribution or model, but avoids overfitting by penalising an additional degree of freedom [23]. The model with the lowest BIC value is chosen as the best.

$$BIC = \log(n)k - 2\log(\widehat{L}), \tag{17}$$

where
$\widehat{L}$ is the maximised value of likelihood function of the model,
n is the number of data points,
k is the number of free parameters to be estimated.

2.6 Generalised Extreme Value Distribution Model

The generalised extreme value (GEV) distribution consists of parametric distributions, namely the: Gumbel, Fréchet, and Weibull distributions. The GEV cumulative distribution function (CDF) is given by

$$G(\mu,\sigma,\xi;x) = \begin{cases} \exp\left(-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right), & \text{for } 1+\xi\left(\frac{x-\mu}{\sigma}\right) > 0,\ \xi \neq 0, \\ \exp\left(-\exp(-\frac{x-\mu}{\sigma})\right), & x \in \mathbb{R},\ \ \xi = 0, \end{cases} \tag{18}$$

where μ, σ and ξ are location, scale and shape parameters, respectively [16,24]. The shape parameter, ξ, is such that for $\xi > 0$, we have the Fréchet family of distributions, $\xi < 0$ gives the Weibull family of distributions, while $\xi = 0$ is the Gumbel family of distributions. Under usual conditions, the GEV distribution is applied to block maxima data set. However, in the present study, the GEV distribution is applied to the complete data set without blocking since it is used as a validation technique or goodness-of-fit test to establish whether the Weibull class indeed is a good fit to the data. This is a unique and unusual application of the GEV distribution to help establish consistency of the Weibull class in the parent distributions fitted. This is one of the gaps we intend to close in candidate parent distributions fitting in this study.

2.7 Trend Analysis

Mann-Kendall Test

The non-parametric Mann-Kendall (M-K) test is commonly used to detect monotonic trends in series of environmental, climate and hydrological data [25,26]. The advantage of M-K test is that it is a non-parametric test and does not require the data to be normally distributed. The null hypothesis, H_0, is that there is no trend or serial correlation from a population while the alternative hypothesis, H_1, states that there is monotonic increasing or decreasing trend [27,28].

Let $x_1, x_2, \ldots, x_n$ be a sequence of random variables, then the M-K statistic can be calculated using the equation

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} sign(x_j - x_k), \tag{19}$$

where x_j denotes the ordered data values, n is the length of observation and S is the M-K statistic. The sign$(x_j - x_k)$ expression in (19) is given by

$$sign(x_j - x_k) = \begin{cases} 1 & \text{if } x_j - x_k > 0, \\ 0 & \text{if } x_j - x_k = 0, \\ -1 & \text{if } x_j - x_k < 0. \end{cases} \tag{20}$$

For $n \geq 10$, the statistic S is approximately normally distributed with mean zero $(E(S) = 0)$ and variance as follows:

$$v(s) = \frac{n(n-1)(2n+5) - \sum_{k=1}^{nk} t_k(k)(k-1)(2k+5)}{18}, \tag{21}$$

where t_k denotes the number of duplicates to extend k. Equation (21) is used in case of tie values in time series, where nk is the total number of ties in dataset. In case of having $n \geq 10$, the standardised test statistic for M-K can be written as

$$Z_s = \tau = \begin{cases} \frac{s-1}{\sqrt{v(s)}}, & \text{if } s > 0, \\ 0, & \text{if } s = 0, \\ \frac{s+1}{\sqrt{v(s)}}, & \text{if } s < 0. \end{cases} \tag{22}$$

The test statistic Z_s, also referred to as Kendall's τ, is used to measure the significance of trend. A positive Z_s or τ value indicates an upward trend, whereas a negative value indicates a downward trend [29]

3 Results and Discussion

This section presents results of the study in the form of tables and figures, as well as a comprehensive discussion of the results.

3.1 Normality Test Results

The normality test results are presented in Table 1 and Table 2 for the average maximum and minimum temperatures, respectively. The significance level, α, was taken as 0.05. The p-values for both maximum and minimum temperature for all stations are less than the significance level (p$<$ 0.05), suggesting that the null hypothesis, H_0, is rejected and concluding that the temperature series for all the stations are not normally distributed.

Table 1: Shapiro-Wilk and Jarque-Bera test results for the maximum temperature

Station name	Test name	Test statistics	p-value
Mara	Shapiro-Wilk	0.981	<0.001
	Jarque-Bera	25.554	<0.001
Messina	Shapiro-Wilk	0.970	<0.001
	Jarque-Bera	110.69	<0.001
Polokwane	Shapiro-Wilk	0.967	< 0.001
	Jarque-Bera	27.43	<0.001
Thabazimbi	Shapiro-Wilk	0.966	<0.001
	Jarque-Bera	11.771	0.003

3.2 Model Selection and Diagnostic Statistics Test Results

The parent distributions model selection information criteria, AIC and BIC, test results are presented in Table 3 and Table 4 for the average maximum and minimum temperatures, respectively. The results reveal that the best fitting parent distribution for both maximum and minimum temperatures is in the Weibull domain of attraction at all the stations, except for Thabazimbi and Polokwane, where the best fitting parent distributions for the minimum temperature were found to be in the domain of attraction of the normal distribution for both stations. However, given that the results from the SW and JB

© 2021 NSP
Natural Sciences Publishing Cor.

Table 2: Shapiro-Wilk and Jarque-Bera test results for the minimum temperature

Station name	Test name	Test statistics	p-value
Mara	Shapiro-Wilk	0.915	<0.001
	Jarque-Bera	75.121	<0.001
Messina	Shapiro-Wilk	0.921	<0.001
	Jarque-Bera	78.103	<0.001
Polokwane	Shapiro-Wilk	0.930	<0.001
	Jarque-Bera	68.317	<0.001
Thabazimbi	Shapiro-Wilk	0.894	<0.001
	Jarque-Bera	30.388	<0.001

normality tests in Table 1 and Table 2 show that the maximum and minimum temperature series for all the stations are not normally distributed; this, therefore, means that the minimum temperature series for Polokwane and Thabazimbi cannot be in the normal domain of attraction. This necessitated performing diagnostic statistics goodness-of-fit tests such as quantile-quantile (Q-Q) plots, probability-probability (P-P) plots, empirical and theoretical density plots and cumulative distribution functions (CDFs) plots presented in Figs. 1 to 10. The results of the diagnostic plots for Figs. 1, 3, 5 and 8 suggest that the maximum temperature series for all the four stations in the Limpopo province can be suitably modelled by a distribution in the Weibull domain of attraction. These diagnostic statistics findings were further confirmed by the GEV distribution results in Table 5. Regarding the minimum temperature series for all the stations, Figs. 2, 4, 6 and 9 present the results of the diagnostic plots for the Weibull distribution, while Figs. 7 and 10 present the results of the diagnostic plots for the normal distribution. The main thrust for presenting both the Weibull and normal distributions diagnostic plots for some stations was to make a visual comparative analysis of the plots. The visual comparative analysis was necessary to Polokwane and Thabazimbi where the model selection information criteria, AIC and BIC, results contradict those of the SW and JB normality tests. Based on the AIC and BIC results, the Weibull distribution was the second best distribution for both Polokwane and Thabazimbi with values very close to those of the normal distribution, while there was no doubt about the performance of the Weibull class of distributions for both Mara and Messina minimum temperature series. However, Figs. 6 and 7 for the Polokwane minimum temperature series Weibull and normal diagnostic plots, respectively, and Figs. 9 and 10 for the Thabazimbi minimum temperature series Weibull and normal diagnostic plots, respectively, do not exhibit much difference in their goodness-of-fit results for both stations. Combining these findings with those from normality tests and information criteria model selection, we can conclude that minimum temperature series parent distribution for all the stations in this study follow the Weibull domain of attraction. These findings are further confirmed by the GEV distribution results in Table 5.

Table 5 presents the maximum likelihood estimates of the stationary GEV distribution with standard errors on the parentheses and 95% confidence intervals (CIs) for the shape parameter, ξ. The estimates and standard errors were combined to give approximate CIs. In particular, a 95% CI for ξ was obtained as $\xi \pm 1.96(SE)$. The results reveal that both the maximum and minimum temperature data for all the meteorological stations: Mara, Messina, Polokwane and Thabazimbi can be modelled by the Weibull family of distributions, since $\xi < 0$ and CIs are significantly different from zero, respectively, for all the stations.

© 2021 NSP
Natural Sciences Publishing Cor.

Table 3: Summary of parent distributions information criteria model selection tests for the maximum temperature

Station name	Distribution	AIC value	BIC value
Mara	Normal	4231.283	4240.75
	Lognormal	4262.193	4271.66
	Gamma	4249.286	4258.753
	Weibull	**4221.385**	**4230.852**
Messina	Normal	4733.907	4743.512
	Lognormal	4858.128	4867.732
	Gamma	4801.965	4811.57
	Weibull	**4709.401**	**4719.006**
Polokwane	Normal	5305.021	5314.9
	Lognormal	5305.021	5314.9
	Gamma	5348.834	5358.712
	Weibull	**5282.204**	**5292.082**
Thabazimbi	Normal	1663.706	1671.114
	Lognormal	1677.577	1684.984
	Gamma	1671.713	1679.121
	Weibull	**1656.678**	**1664.085**

Table 4: Summary of parent distributions information criteria model selection tests for the minimum temperature

Station name	Distribution	AIC value	BIC value
Mara	Normal	5064.259	5073.726
	Lognormal	5309.453	5318.92
	Gamma	5183.499	5192.966
	Weibull	**5050.361**	**5059.828**
Messina	Normal	5536.302	5542.907
	Lognormal	5740.935	5750.557
	Gamma	5643.047	5652.652
	Weibull	**5500.478**	**5510.083**
Polokwane	Normal	**6247.431**	**6257.333**
	Log-normal	6647.782	6657.684
	Gamma	6419.108	6429.009
	Weibull	**6253.977**	**6263.878**
Thabazimbi	Normal	**1976.296**	**1983.703**
	Log-normal	2156.169	2163.577
	Gamma	2049.755	2057.162
	Weibull	**1999.005**	**2006.412**

Table 5: Maximum likelihood estimates of the GEV distribution parameters with standard errors on the parentheses and the 95% CI for the shape parameter (ξ) of the maximum and minimum temperatures

Station name		$\hat{\mu}$	$\hat{\sigma}$	$\hat{\xi}$	95% for ξ
Mara	Max	26.465(0.116)	3.114(0.084)	-0.377(0.017)	(-0.410,-0.344)
	Min	11.542(0.019)	5.403(0.012)	-0.637(0.000)	(-0.637,-0.637)
Messina	Max	27.949(0.275)	5.376(0.301)	-0.548(0.018)	(-0.583,-0.513)
	Min	14.314(0.189)	5.674(0.114)	-0.584(0.002)	(-0.588,-0.580)
Polokwane	Max	23.527(0.107)	3.246(0.071)	-0.214(0.007)	(-0.228,-0.200)
	Min	9.437(0.159)	4.852(0.110)	-0.241(0.008)	(-0.257,-0.225)
Thabazimbi	Max	28.027(0.244)	3.939(0.172)	-0.319(0.023)	(-0.364,-0.274)
	Min	11.995(0.424)	6.974(0.325)	-0.576(0.025)	(-0.625,-0.527)

© 2021 NSP
Natural Sciences Publishing Cor.

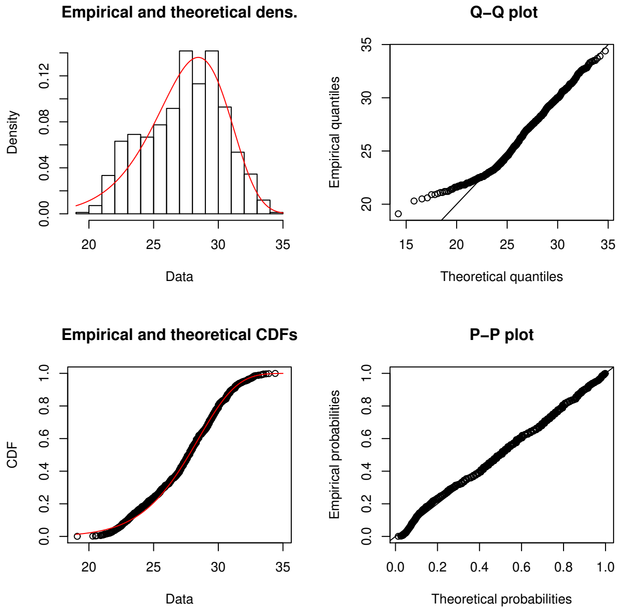

Fig. 1: Diagnostic plots illustrating the fit of the Weibull family of distributions to the average maximum temperature for Mara

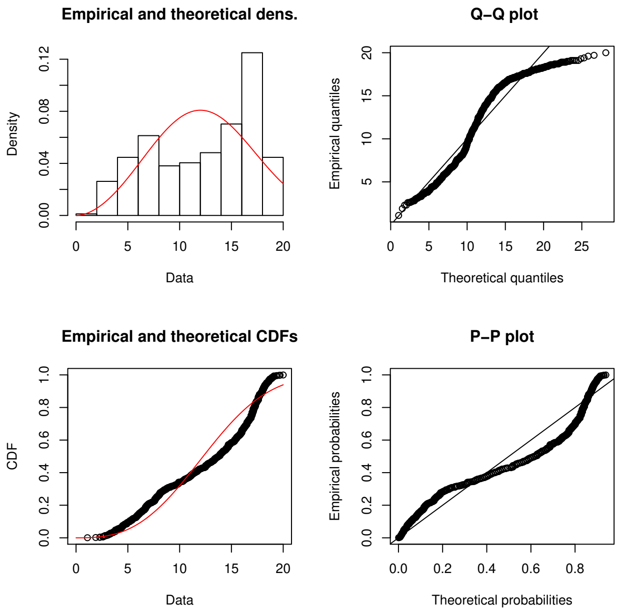

Fig. 2: Diagnostic plots illustrating the fit of the Weibull family of distributions to the average minimum temperature for Mara

3.3 Long-term Trends Analysis Results

The Mann-Kendall test statistic trend analysis results are presented in Table 6. The significance level, α, was taken as 0.05. The p-values for minimum temperature for the following stations: Mara, Messina and Polokwane are less than the

© 2021 NSP
Natural Sciences Publishing Cor.

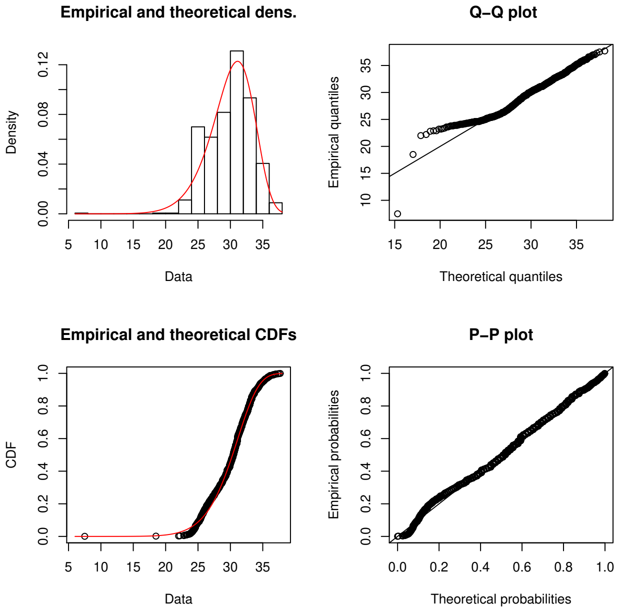

Fig. 3: Diagnostic plots illustrating the fit of the Weibull family of distributions to the average maximum temperature for Messina

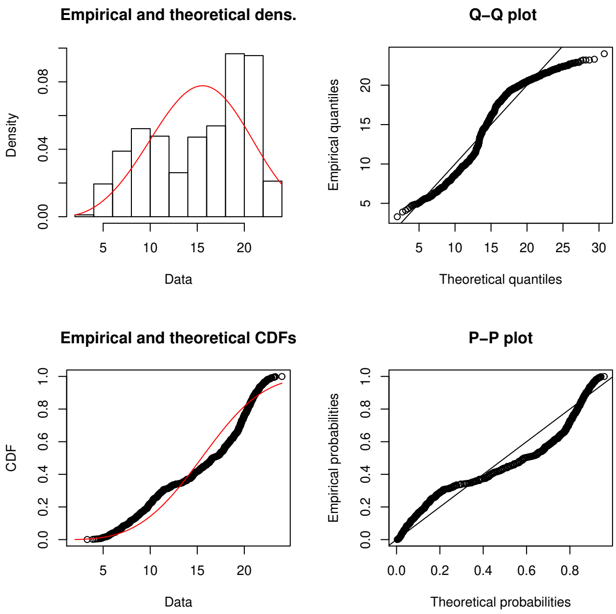

Fig. 4: Diagnostic plots illustrating the fit of the Weibull family of distributions to the average minimum temperature for Messina

significance level ($p < 0.05$), suggesting that there is a significant trend in the minimum temperature data for these stations, while for the maximum temperature, the results suggest that there is no significant trend in the data for all the stations ($p > 0.05$). For both the maximum and minimum temperatures in Thabazimbi, the results reveal that there is no significant trend in the data, since the p-values are greater than the significance level ($p > 0.05$). The M-K test statistic

© 2021 NSP
Natural Sciences Publishing Cor.

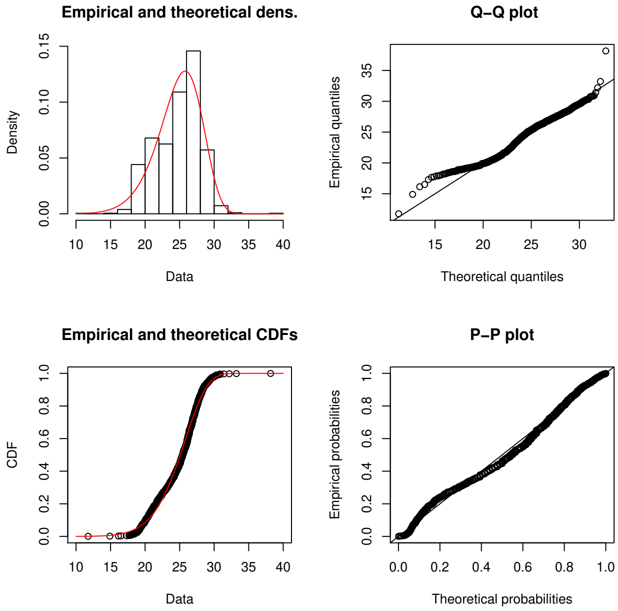

Fig. 5: Diagnostic plots illustrating the fit of the Weibull family of distributions to the average maximum temperature for Polokwane

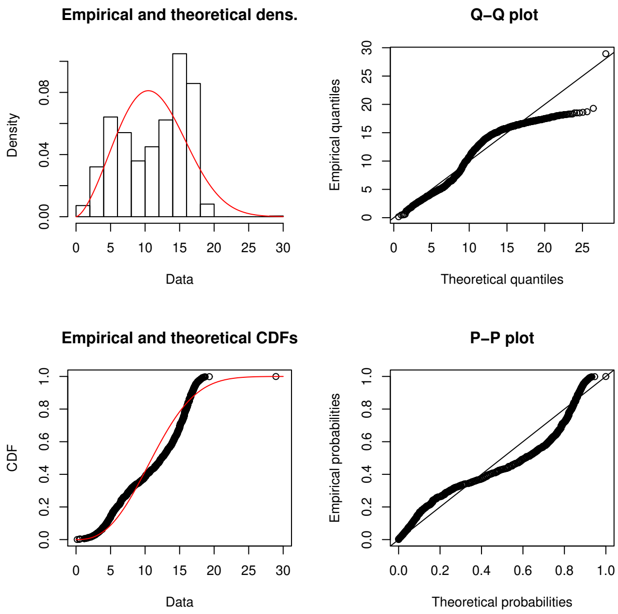

Fig. 6: Diagnostic plots illustrating the fit of the Weibull family of distributions to the average minimum temperature for Polokwane

and Kendall's τ values are all negative for the minimum temperature suggesting that there is a monotonic decreasing trend in the minimum temperature for all the stations. Therefore, there is a significant monotonic decreasing trend in the minimum temperature for the stations Mara, Messina and Polokwane ($\tau < 0$ and $p < 0.05$), while for Thabazimbi, the monotonic decreasing trend in minimum temperature is insignificant ($\tau < 0$ and $p > 0.05$). As for maximum

© 2021 NSP
Natural Sciences Publishing Cor.

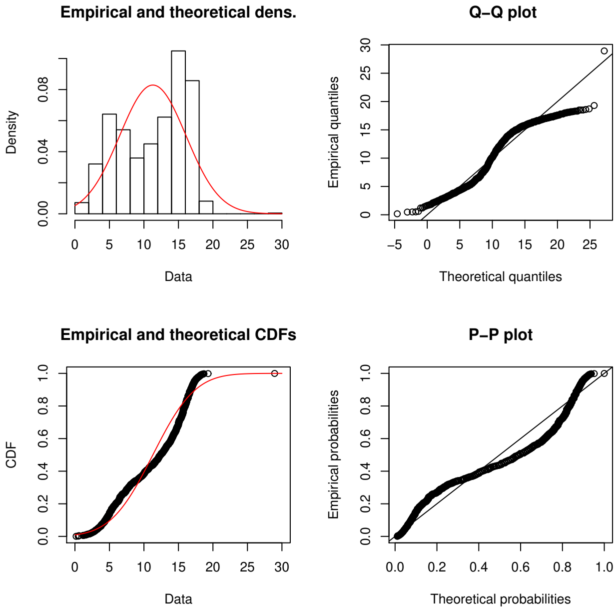

Fig. 7: Diagnostic plots illustrating the fit of the normal distribution to the average minimum temperature for Polokwane

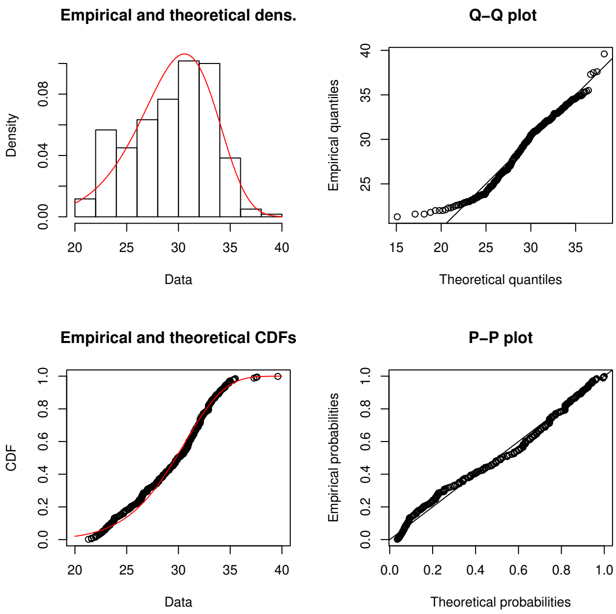

Fig. 8: Diagnostic plots illustrating the fit of the Weibull family of distributions to the average maximum temperature for Thabazimbi

temperature, the M-K test statistic and Kendall's τ results in Table 6 reveal that there is an insignificant monotonic increasing trend in the maximum temperature for Mara, Messina and Thabazimbi ($\tau > 0$ and $p > 0.05$), while for Polokwane, there is an insignificant monotonic decreasing trend in maximum temperature ($\tau < 0$ and $p > 0.05$). These findings are also supported by the time series plot results in Figs. 11 to 14. Moreover, the use of M-K test statistic helped to uncover the long-term trends which were not discernable in the time series plots.

© 2021 NSP
Natural Sciences Publishing Cor.

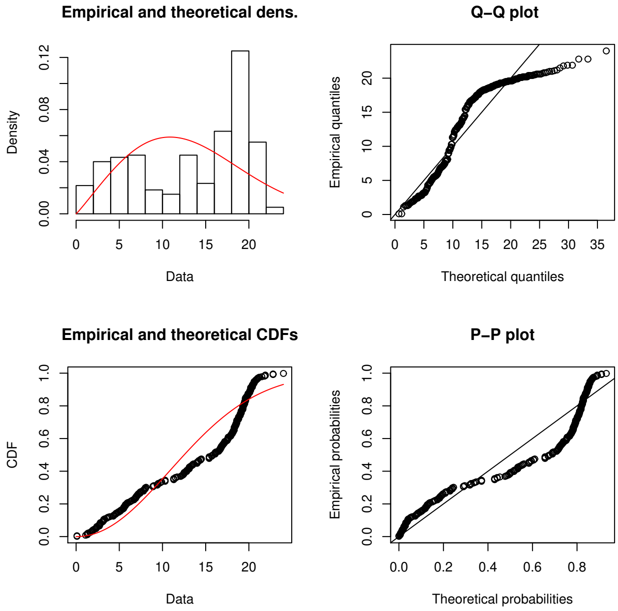

Fig. 9: Diagnostic plots illustrating the fit of the Weibull family of distributions to the average minimum temperature for Thabazimbi

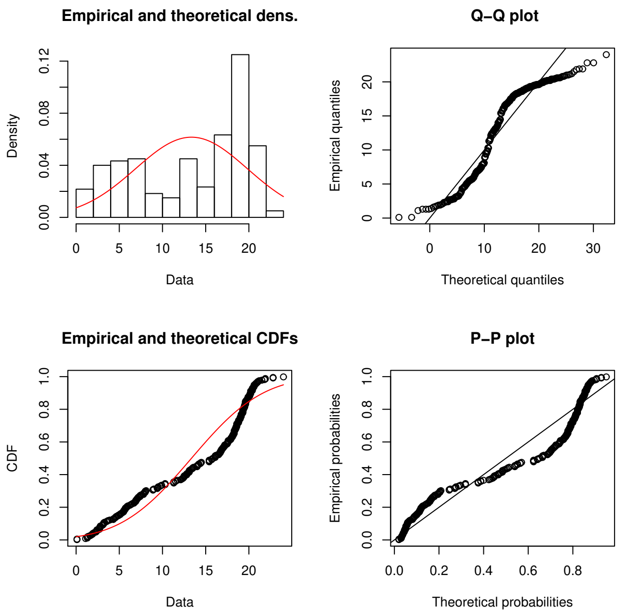

Fig. 10: Diagnostic plots illustrating the fit of the normal distribution to the average minimum temperature for Thabazimbi

Overall, the results in Table 6 reveal that there is a significant long-term downward trend of minimum temperature and long-term upward trend of maximum temperature in the Limpopo province. These findings are in agreement with those for [13] which indicated an increase in warm extremes and a decrease in cold extremes for South Africa. Polokwane was the only station that had an insignificant downward trend for both maximum and minimum temperatures. There are no

© 2021 NSP
Natural Sciences Publishing Cor.

clear reasons attributed to this anomalous behaviour of a downward trend in maximum temperature for Polokwane, but one clear distinction is that Polokwane station is in the capital city of the province as opposed to the other three stations.

Table 6: Summary of Mann-Kendall trend analysis

Station name		M-K test statistic(S)	Kendall's τ	Var(S)	p-value
Mara	Max	71.987	0.138	7233.832	0.659
	Min	-155.681	-1.099	7234.027	0.0004
Messina	Max	26.031	0.0160	8896.256	0.886
	Min	-120.334	-0.741	8896.724	0.014
Polokwane	Max	-119.951	-0.153	3645.197	0.589
	Min	-157.297	-0.718	3773.997	0.0105
Thabazimbi	Max	60.096	0.907	1216.053	0.085
	Min	-24.605	-0.371	1216.122	0.481

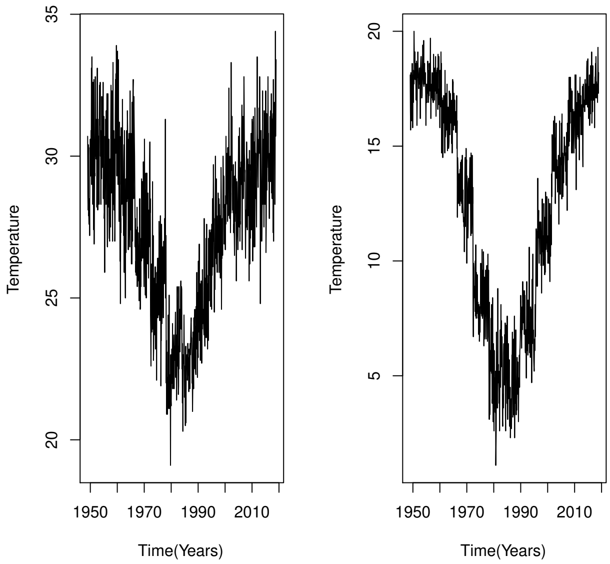

Fig. 11: Time series plot of Mara maximum (left panel) and minimum (right panel) temperatures from 1949 to 2018

© 2021 NSP
Natural Sciences Publishing Cor.

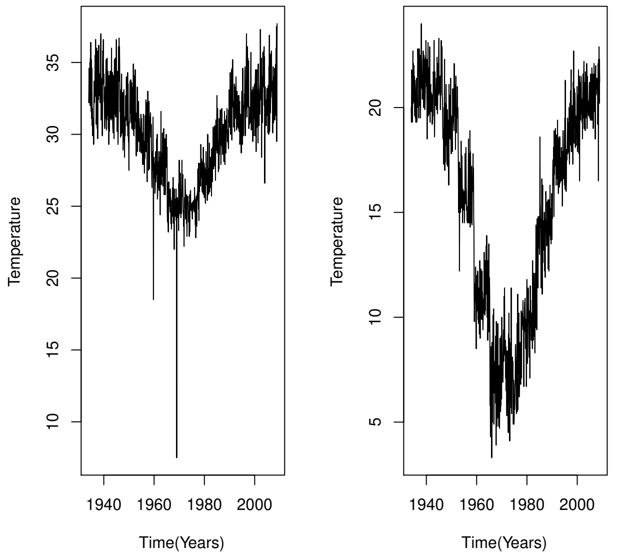

Fig. 12: Time series plot of Messina maximum (left panel) and minimum (right panel) temperatures from 1934 to 2009

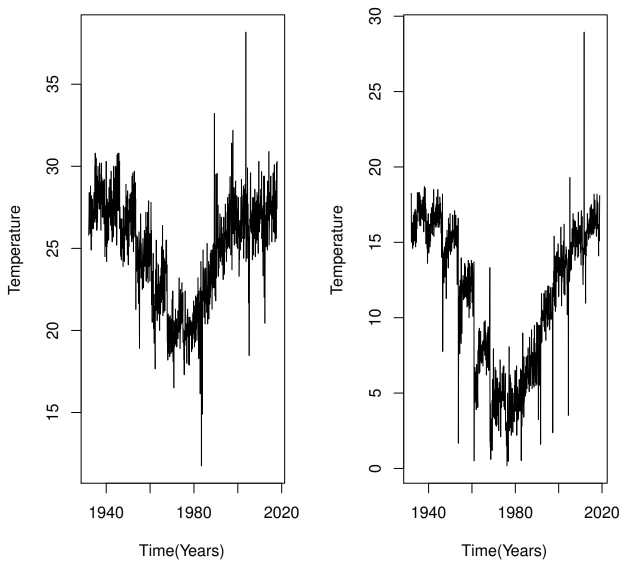

Fig. 13: Time series plot of Polokwane maximum (left panel) and minimum (right panel) temperatures from 1932 to 2018

© 2021 NSP
Natural Sciences Publishing Cor.

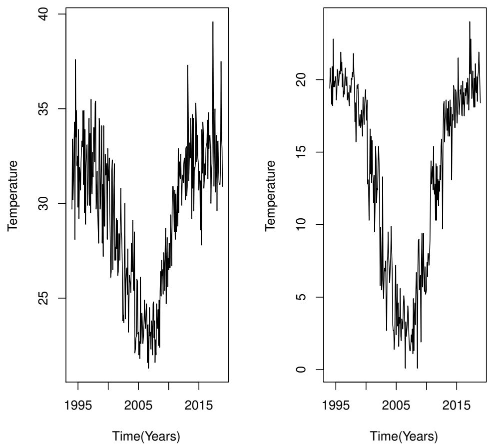

Fig. 14: Time series plot of Thabazimbi maximum (left panel) and minimum (right panel) temperatures from 1994 to 2018

4 Conclusion

The four parent distributions, namely: normal, lognormal, gamma and Weibull were fitted to the maximum and minimum temperature data. The findings reveal that the Weibull domain of attraction (which is a family of short-tailed distributions that includes the Weibull distribution) is the best fitting parent distribution for both the maximum and minimum temperatures at all the stations of the Limpopo province studied. The long-term trend was investigated using Mann-Kendall (M-K) test and time series plots. The findings reveal both increasing and decreasing long-term trends of maximum and minimum temperature extremes, respectively. The time series plots could not exhibit discernable long-term trends in both maximum and minimum temperatures and this necessitated the application of the non-parametric M-K test to detect hidden long-term trends in the temperature data. The M-K test findings reveal that there is a downward long-term trend in the minimum temperature for all the stations considered in this study, while the maximum temperature exhibits an upward long-term trend for all the stations, with the exception of Polokwane station which reveals a downward long-term trend in both maximum and minimum temperatures. The findings from the M-K test further reveal that the downward long-term trends in minimum temperature for Mara, Messina and Polokwane are very significant. These climate change findings could be attributed mainly to the effects of global warming and natural modes of interdecadal variability such as El Niño and La Niña phenomenon [16,30].

The findings in this study lead to the conclusion that the parent distribution for the maximum temperature in the Limpopo province lies in the Weibull domain of attraction, while for the minimum temperature, the common parent distributions also lie in the Weibull domain of attraction based on the diagnostic tests conducted. Furthermore, there is a gradual decrease in the minimum temperature and a gradual increase in the maximum temperature in the Limpopo province.

Future studies will look into the possibility of applying both univariate and multivariate extreme value theory techniques to investigate further whether these climatic changes in mean monthly temperature extremes can indeed be attributed to global warming and other natural modes of interdecadal variability such as El Niño and La Niña phenomenon.

Acknowledgement

We thank South African Weather Service (SAWS) for providing the data that we used for the analysis in this paper.

Conflict of Interest

The authors declare that they have no conflict of interest.

© 2021 NSP
Natural Sciences Publishing Cor.

References

[1] CSIR, Climate information and early warning system to support disaster risk reduction and management under future climate conditions in South Africa, CSIR Report 2, (2014).
[2] G. Worku, E. Teferi, A. Bantider and Y. T. Dile, Observed changes in extremes of daily rainfall and temperature in Jemma Sub-Basin, Upper Blue Nile Basin, *Theoretical and Applied Climatology*, **135**(3-4), 839-854 (2019).
[3] D. R. Tanusree and D. K. Kishore, Time series analysis of Dibrugarh air temperature, *Journal of Atmospheric and Earth Environment*, **1**(1) 30-34 (2012).
[4] DEA, Department of Envirinmental Affairs (DEA), South Africa's third climate change, DEA Report, Pretoria, (2017).
[5] B. Bayeh and M. Alemayehu, Scientific community debates on causes and consequences of global warming, *Journal of Agriculture and Environmental Sciences*, **4**(1), 1-14 (2019).
[6] O. O. Ochanda, Time series analysis and forecasting of monthly air temperature changes in Nairobi, Kenya, MSc dissertation, University of Nairobi, (2016).
[7] L. Nhamo, G. Matchaya, T. Mabhaudhi, S. Nhlengethwa, C. Nhemachena and S. Mpandeli, Cereal production trends under climate change: Impacts and adaptation strategies in Southern Africa, *Agriculture*, **9**(2), 30 (2019).
[8] E. Andersson and E. Nilsson, An extreme value approach to modelling risk of extreme rainfall in Bangladesh, Bachelor Thesis in Mathematical Statistics, Lund University, (2018).
[9] C. Y. Wright, R. M. Garland, M.Norval and C. Vogel, Human health impacts in a changing South African climate, *South African Medical Journal*, **104**(8), 579-582 (2014).
[10] C. L. Reddy and K. Vincent, Climate risk and vulnerability, A Handbook for Southern Africa, Second edition, CSIR, Pretoria, South Afica, (2017).
[11] A. C. Kruger and S. Shongwe, Temperature trends in South Africa:1960-2003, *International Journal of Climatology: A Journal of the Royal Meteorology Society*, **24**(15), 1929-1945 (2004).
[12] M. F. Tshiala, J. M. Olwoch and F. A. Engelbrecht, Analysis of temperature trends over Limpopo province, South Africa, *Journal of Geography and Geology*, **3**(1), 13 (2011).
[13] A. C. Kruger and S. S. Sekele, Trends in extreme temperature indices in South Africa: 1962 – 2009, *International Journal of Climatology*, **33**(3), 661-676 (2013).
[14] M. M. Nemukula, Heat waves in South Africa: Observed variability structure and trends, MSc dissertation, University of Witwatersrand, (2018).
[15] M. M. Nemukula, C. Sigauke and D. Maposa, Bivariate threshold excess models with application to extreme high temperatures in Limpopo province of South Africa, Proceedings of the 60th Annual Conference of SASA, South Africa, 33-40, (2018).
[16] D. Maposa, Statistics of extremes with applications to extreme flood heights in the lower Limpopo River basin of Mozambique, PhD thesis, University of Limpopo, (2016).
[17] R. Chifurira, Modelling mean annual rainfall for Zimbabwe, PhD Thesis, University of the Free State, (2018).
[18] S. Y. Kim, B. Sapotta, G. Jang, Y. H. Kang and H. G. Kim, Prefeasibility study of photovoltaic power potential based on a skew-normal distribution, *Energies*, **13**(3), 676 (2020).
[19] Y. R. Gel and J. L. Gastwirth, A robust modification of the Jarque-Bera test of normality, *Economic Letters*, **99**(1), 30-32 (2008).
[20] D. Chikobvu and R. Chifurira, Modelling of extreme minimum rainfall using generalised extreme value distribution for Zimbabwe, *South African Journal of Science*, **111**(9/10), (2015).
[21] R. H. Al-Suhili and R. Khanbilvard, Frequency analysis of the monthly rainfall data at Sulaimania region,Iraq, *American Journal Of Engineering Research*, **3**(5), 212-222 (2014).
[22] W. G. Strupczewski, V. P. Singh and W. Feluch, Non-stationary approach to at-site flood frequency modelling I. Maximum likelihood estimation, *Journal of Hydrology*, **248**, 123-142 (2001).
[23] J. J. Dziak, D. L. Coffman, S. T. Lanza, R. Li and L. S. Jermiin, Sensitivity and specificity of information criteria, *BioRxiv*, 1-41 (2019).
[24] A. K. Dey, A. Edwards and K. P. Das, Determinants of high crude oil price: A non-stationary extreme value approach, *Journal of Statistical Theory and Practice*, **14**(1), 4 (2020).
[25] N. Karmeshu, Trend detection in annual temperature and precipitation using the Mann Kendall test: A case study to assess climate change on select states in the Northeastern United states, MSc dissertation, University of Pennysylvania, 2012.
[26] V. Ongoma and H. Chen, Temporal and spatial variability of temperature and precipitation over East Africa from 1951 to 2010, *Meteorology and Atmospheric Physics*, **129**(2), 131-144 (2017).
[27] M. S. Ndlovu and M. Demlie, Statistical analysis of groundwater level variability across KwaZulu-Natal province, South Africa, *Environmental Earth Science*, **21**(77), 739 (2018).
[28] T. Pohlert, Non-parametric trend tests and change-point detection, Creative Commons Licence, 2020.
[29] G. Meng, L. Jing, H. Hongshi, X. Jiawei and J. Yinghua, Detecting global vegetation changes using Mann-Kendal (MK) trend test for 1982-2015 time period, *Chinese Geographical Science*, **28**(6), 907-919 (2018)
[30] R. W. Katz, Statistics of extremes in climate change, *Climate Change*, **100**(1), 71-76 (2010).

© 2021 NSP
Natural Sciences Publishing Cor.