International Journal of Environment and Geoinformatics (IJEGEO) is an international, multidisciplinary, peer reviewed, open access journal.

Evaluation of Rainfall Extreme Characteristics in Dodoma Urban, a Central Part of Tanzania

Ombeni J. MDEE

Research Article

Evaluation of Rainfall Extreme Characteristics in Dodoma Urban, a Central Part of Tanzania

Ombeni J. Mdee (iD)

College of Earth Sciences and Engineering, University of Dodoma, P.O Box 259, Dodoma, Tanzania.

Abstract

The occurrence of low rainfall in semi-arid areas including Dodoma urban leads to a shortage of water. This paper addresses the evaluation of rainfall extreme characteristics by analyzing the variability indices, probability distribution and return periods. The daily rainfall index shown the magnitude of rainfall varied unpredictably within annual wetted days. The study area experienced a period of one to two months per year with extreme seasonality as evaluated using the rainfall seasonality index. The standardized anomaly index indicated the equivalent of 60% of 20 years experienced the dried years with unpredicted occurrence. The Weibull distribution was selected out of Fifteen probability functions when ranked using Kolmogorov–Smirnov and Anderson-Darling tests. The return periods of rainfall had an average rainfall of 576 mm and were predicted using seventeen plotting position methods and Weibull distribution. Therefore, the magnitude of rainfall in the semi-arid areas would not easily be estimated but using more than one technique would improve the evaluation of rainfall characteristics.

Keywords: Rainfall variability, probability distribution, return periods, plotting position methods, rainfall indices

Introduction

The widespread of droughts in some parts of Africa is caused by uneven distribution or absence of rainfall. The drought occurrence is due to the uneven distribution of rainfall and temperature over a given time and space (Abdalla et al., 2021; USGS, 2021; Twisa and Buchroithne, 2019; Markovic, et al, 2012). The drought is the situation when crops experienced moisture deficiency, a prolonged period of rainfall shortage, deficiency of water supply or decreased the stream flow (USGS, 2021). The understanding of the rainfall variability is very important during designs of hydraulic structures, agricultural planning, river controls, and water supply facilities. However, the increase of extreme events of high rainfall intensity caused excessive flooding (Sharma and Singh, 2019; Menteş et al., 2019; Moazzam et al., 2018). The flooding destructs the buildings, roads, and planted crops resulted to land erosion.

The seasonal variation of rainfall affects the socio-economic activities in urban areas, industrialization, and agricultural activities (Vieira, et al, 2018). The rainfall intensity is highly varied in space and time, especially in semi-arid areas. Critchley et al (Critchley, et al, 1991) indicated the rainfall in semi-arid areas was characterized with short duration, limited areal extent, and relatively high intensity. De Paola et al (2014) indicated irregular rainfall intensity with the increased frequency that occurred in the semi-arid areas. Different studies used several daily rainfall indices like principal component analysis (Joshi, et al, 2014), standardized

precipitation index (WMO, 2012); 95th percentile (Dunkerley, 2019); 99th and 99.9th percentile (Alexander, et al., 2019) to evaluate the rainfall variability. The trend increased or decreased of rainfall was analyzed using the Mann-Kendall test with the sen's slope estimate i.e., the sen's slope of greater than 1 indicates an upward trend in a time series; otherwise, the data presents a downward trend (Twisa and Buchroithne, 2019).

The rainfall varied from time to time when rained and analyzed daily, monthly, annually, and seasonally based on the different probability functions (Mafuru and Guirong, 2018). There are many probability functions used to analyze the rainfall distribution (Sharma and Singh, 2010). Zhan et al (2018) used the modified Weibull distribution called Impulse Weibull distribution function to determine the extreme events and annual precipitation. Kumar and Bhardwaj (2015) used three probability distribution functions such as Log-normal, Gumbel, and Log-Pearson type-III. Using the Chi-square technique helped to select the best fit probability distribution function. Alonge and Afullo (2012) used rain drop-size distribution with the method of moments in Weibull distribution to analyze the rainfall frequency distribution.

The Generalized Extreme Value, Burr, and Weibull distribution functions provide the best fit to both annual and seasonal maximum precipitation than exponential and generalized Pareto 2 distributions (Li, et al, 2015). The mixed exponential, Gamma, Weibull, Log-normal distribution functions were compared using Akaike

Information Criterion (AIC) (Suhaila, et al, 2011). Sasireka et al (2019) used AIC and Bayesian Information Criteria to select the best-fit distribution functions between Gumbel and Gamma distribution functions. But, Vieira et al (2018) indicated Gamma and Weibull distribution functions were most suitable for probabilistic fitting compared with log-normal distribution function using the AIC test.

The Gamma and Weibull distribution functions were more suitable to analyze monsoon rainfall inter-monsoon rainfall, respectively (Syafrina, et al, 2018). Husak et al (2006) indicated the Gamma distribution function is suitable to analyze the rainfall distribution compared with the Weibull distribution function using Kolmogorov–Smirnov (KS) test. Sunusi (2017) used the maximum likelihood method to estimate the Weibull parameters of the Weibull distribution function and the power-law function indicated the same general characteristics. The mixed exponential distribution function was the best fit compared with the Gamma, Weibull, Log-normal, and exponential distribution functions using the Chi-square test and Anderson-Darling (AD) test (Kist and Filho, 2015). Marques et al (2014) used the method of moments, maximum likelihood, and L-moments to estimate the Weibull parameters for the Weibull distribution function and compared with GEV, Gumbel, and Gamma distribution functions to study the rainfall characteristics. Al-Suhili and Khanbilvardi (2014) used three tests such as t-test, F-test, and KS-test to measure the performance of Gamma, Weibull, and exponential distribution functions on the rainfall data, indicated the three distribution functions produce the same rainfall frequency.

Furthermore, a return period is an average time for extreme events of rainfall to come again, i.e., quantifies the frequency of rainfall events to occur with the same magnitude. In other means, the return period defines the likelihood of rainfall events to occur in the future. The return period (T) of any rainfall event (r) is the reciprocal of the probability of exceedance (Critchley, et al, 1991; Nwaogazie and Sam, 2019; Arvind, et al, 2017). The probability of exceedance was estimated using plotting position methods presented in Table 1. Unfortunately, the existing plotting position formulas are originated from Equation (1) and do not provide outstanding performance in any condition (Erto and Lepore, 2014; Yahaya, et al, 2012; Portela and Delgado, 2009). Therefore, this paper aims to assess the shortage of rainfall variability, propose suitable probability distribution functions, and conducting an assessment of return periods.

Study area and data sources

The study is conducted in the Dodoma region located in the central part of Tanzania characterized by a semi-arid climate, warm to hot temperature with an average elevation of 1130 m above sea level. Figure 1 shows the location of the study area, rainfall station, and rapid construction areas with the location of latitude of 6.170^{o} S and longitude of 35.753^{o} E. The rainfall intensity and frequency vary widely in this region like other semi-arid areas. The area receives a single rainfall season between October and May ranging between 289 to 1117 mm/year with the highest in December and January.

Table 1. Plotting position methods for estimating probabilities of exceedance adopted from different literature sources

S/N	Probability of exceedance *100%	Name of investigator	Sources
1	$P(x_r) = \dfrac{r}{N}$	California State Department	
2	$P(x_r) = (r - 0.5)/N$	Hazen	
3	$P(x_r) = r/(N + 1)$	Weibull	(Dirk, 2013)
4	$P(x_r) = (r - 0.375)/(N + 0.25)$	Sevruk and Geiger	
5	$P(x_r) = (r - 0.44)/(N + 0.12)$	Gringorten	
6	$P(x_r) = (r - 0.4)/(N + 0.2)$	Cunnane	(Saul, 1996)
7	$P(x_r) = \begin{cases} \dfrac{r - 0.3175}{N + 0.365}, & r \leq N - 1 \\ 0.5^{\wedge(1/N)}, & r = N \end{cases}$	Filliben	(Olivera and Heard, 2018)
8	$P(x_r) = (r - 0.3)/(N + 0.4)$	Chegodayev	(Kumar and Gaddada, 2015)
9	$P(x_r) = (r - 0.25)/(N + 0.5)$	Adamowski	
10	$P(x_r) = (3r - 1)/(3N + 1)$	Tukey	(Arvind, et al, 2017)
11	$P(x_r) = (r - 0.31)/(N + 0.38)$	Beard	
12	$P(x_r) = (r + 0.5)/(N + 1)$	Hirsch	(Yahaya, et al, 2012)
13	$P(x_r) = (r - 0.5)/(N + 0.25)$	IEC56	
14	$P(x_r) = (r - 0.35)/N$	Landwehr	
15	$P(x_r) = (r + 1)/(N + 2)$	Laplace	
16	$P(x_r) = (r - 0.4)/N$	McClung	
17	$P(x_r) = (r - 0.35)/(N + 0.25)$	Blom	

$$P(x_r) = \frac{r - A}{N + 1 - 2A} \qquad (1)$$

Where *A* is a constant between 0 and 1 and *N* is a sample size of rainfall data.

The average temperature of Dodoma urban is 24.7° C with the hot season started from October to December and the cool season started from June to August. The population growth has rapidly increased due to the development of buildings, roads, railways, and agriculture activities. The growth of development was the effect of shifting national capital from Dar es Salaam

to Dodoma region since the year 2016. The daily rainfall data of two decades was collected at the Tanzania Meteorological Agency (TMA), central office-Dodoma region, Tanzania from the year 2000 to 2020 to evaluate the daily variability and probability distribution. Also,

monthly scale rainfall data were collected from Hamisi (2013) between the years 1961 and 2000. Then, return periods of rainfall were evaluated for six decades between the years 1961 and 2020.
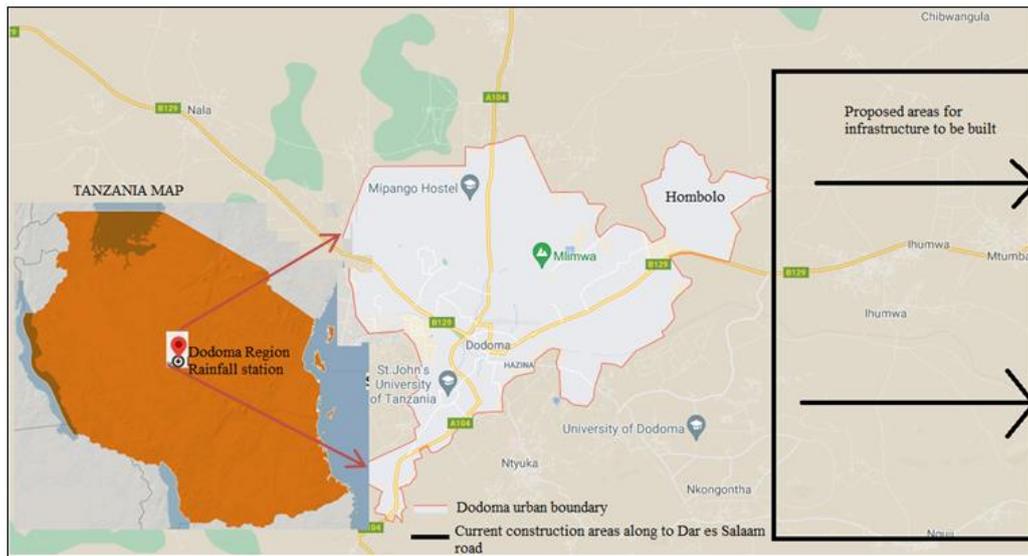

Fig. 1. Map of study area showing the Dodoma urban area, rainfall station, and current construction areas (photo taken from Google Earth exemplified with Tanzania map)

Material and methods
Methods used to evaluate rainfall variability

The rainfall variability was assessed using six statistical methods including mean, standard deviation, standardized anomaly index, daily rainfall, seasonality, and cumulative rainfall departure indices. The mean value shows the central tendency and location of the peak, calculated as expressed in Equation (2). Standard deviation (σ) measures the variability of rainfall data from the average value as expressed in Equation (3).

$$\mu = \frac{1}{N} \sum_{i=1}^{N} R_i \qquad (2)$$

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (R-\mu)^2} \qquad (3)$$

Where N denotes the count of rainfall data recorded from $i = 1, 2,3....N$. The ratio of standard deviation and mean value is called the coefficient of variation. The normal distribution function is characterized by average and standard deviation. The normal distribution is very perfect when plotted symmetrically on both sides of the mean value in the normal line. The Z-score or standardized anomaly index (SAI) tested the normality of annual rainfall distribution as expressed in Equation (4). When SAI = 0, means the annual rainfall is equal to the mean value for all years, SAI > 0 means above normal distribution and wetted years, and SAI < 0 means below normal rainfall distribution and dried years (Kisaka, et al, 2015-2018; Nouaceur and Mursrescu, 2016).

$$SAI = \frac{R-\mu}{\sigma} \qquad (4)$$

The magnitude and frequency of extreme rainfall events for daily rainfall were calculated using the daily rainfall index (DRI) as expressed in Equation (5) (Zengeni, et al, 2016).

$$DRI = \frac{\overline{R}}{T_w} \qquad (5)$$

Where T_w is the total number of wet days in each year, and $\overline{R}$ is the annual rainfall in the same year. Rainfall seasonality refers to patterns of increased rainfall at a given location. The rainfall seasonality index (RSI) was estimated using Equation (6) for certain months of each year (Elzopy, et al, 2020). When RSI = 0, indicates rainfall spread throughout the year, RSI between 0.6 and 0.79, indicates seasonal and RSI > 1.2 rainfall occurs in 1-2 months, extreme seasonality (Sharma and Singh, 2019).

$$RSI = \frac{1}{\overline{R}} \sum_{i=0}^{12} \left| R_i - \frac{\overline{R}}{12} \right| \qquad (6)$$

Methods used to rank probability distribution functions

Several scholars have been utilized many probability distribution functions as described in Sharma and Singh (2010). About fifteen (15) probability distribution functions were selected to analyze daily rainfall data from November to April for 20 years. This study used the interpolated input data of continuous shape parameter, continuous scale parameter, continuous location parameter, continuous inverse scale parameter, and degree of freedom from the *EasyFit 5.0* software to

evaluate the distribution functions. Also, the distribution functions were ranked using the Kolmogorov-Smirnov (Moccia et al., 20212021) and Anderson-Darling (AD) tests (Marques, et al, 2014). The *EasyFit 5.0* software consists of more than 55 coded probability distribution functions and helping to reduce the time for manual analysis by about 70 to 95% (MathWave Technologies, 2021).

Method used to estimate return period of rainfall

The return period was estimated as the reciprocal of the plotting position methods. The plotting position methods were ranked using R-squared, AD test, and Root Mean Square Error (RMSE). Then, the selected plotting position methods were taken to estimate the return periods. The Weibull distribution function is normally used to analyze the rainfall to the areas of rain scarcity, high extreme events, and temporal heterogeneity (Olivera and Heard, 2018). Considering the cumulative Weibull function as expressed in Equation (7), and equal to the probability of exceedance (Table 1); then introduced natural logarithm to deduce the Weibull scale and shape parameters, α and β respectively as expressed in Equation (8). The α and β *were* used to determine the average Weibull rainfall based on the plotting position methods as expressed in Equation 9.

$$\text{CDF}(x) = P_{(x_r)} = 1 - e^{-\left(\frac{x}{\beta}\right)^{\alpha}} \quad (7)$$

$$\ln\left[-\ln\left(1 - \text{CDF}(x)\right)\right] = \alpha \ln(x) - \alpha \ln(\beta) \quad (8)$$

The Weibull two-parameter is determined when plotting $ln\left[-ln\left(1 - CDF(x)\right)\right]$ against $ln(x)$ and estimated slope of the trendline as equal to the α parameter and y-intercept is equal to the term, $-\alpha ln(\beta)$, i.e., From straight-line form, $y = ax + b$; $a = \alpha$ and $\beta = e^{\left(-(b/\alpha)\right)}$ (Azad, et al, 2014). The rainfall data were ranked in descending order showing the lowest value last and the highest value first.

$$\mu_w = \alpha\Gamma\left(1 + {}^1/_\beta\right) \quad (9)$$

Where μ_w is the average Weibull rainfall and Γ is the gamma function.
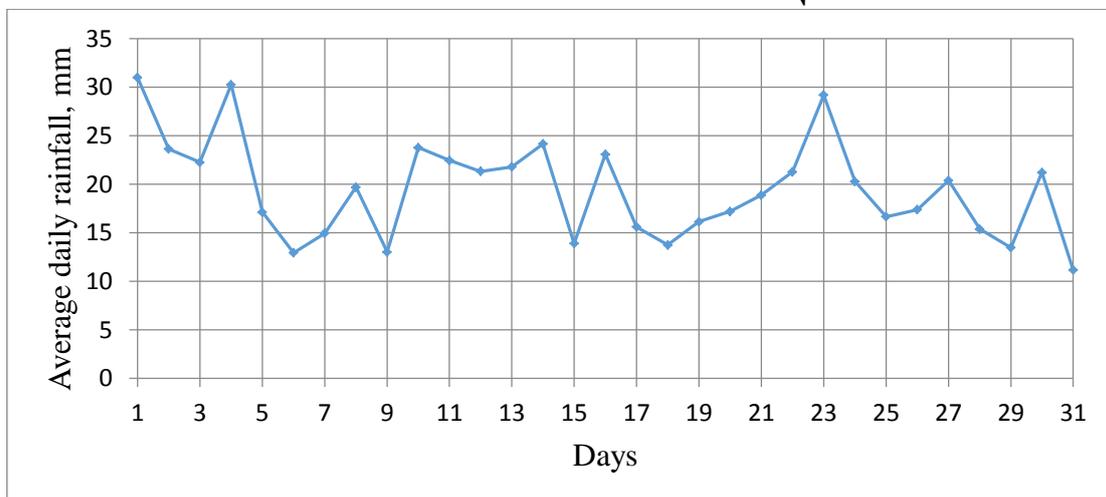
Evaluation of rainfall distribution and plotting position methods

Fifteen probability distribution functions and seventeen plotting position methods were evaluated using correlation technique, best-fit model or R-squared, and RMSE. The correlation is the comparison of two variables and estimating the relationship strength based on the coefficients that ranged between -1 and +1 (Liu, et al, 2017). The coefficient of correlation that is closer to +1 indicates a strong relationship and trended in the same direction or closer to -1 indicates the strong relationship and trended in the opposite direction (Shaban, et al, 2020). The correlation coefficient (*r*) was calculated using Equation (10).

$$r = \frac{n\sum_{i,j=1}^n m_i \cdot m_j - \sum_{i=1}^n m_i \cdot \sum_{j=1}^n m_j}{\sqrt{[n\sum_{i=1}^n m_i{}^2 - (\sum_{i=1}^n m_i)^2]\left[n\sum_{j=1}^n m_j{}^2 - \left(\sum_{j=1}^n m_j\right)^2\right]}} \quad (10)$$

Where i and j denote the measured rainfall data, (m) from two different probability distribution functions or plotting position methods. The R-squared is used to measure the error of two variables derived from measured (m) and predicted (p) from the probability distribution functions or plotting position methods. The R-squared is expressed as shown in Equation (11). The R-squared coefficient is ranging between 0 and 1, closer to 1 indicates less error of fitness between two variables (Kang, et al, 2018).

$$R^2 = \frac{\sum_{i=1}^n (m_i - \mu)^2 - \sum_{i=1}^n (m_i - p_i)^2}{\sum_{i=1}^n (m_i - \mu)^2} \quad (11)$$

The RMSE estimates the associated error that is independent of the expressed specific units. Generally, the lower the RMSE, the more accurate prediction between two different plotting position methods or probability distribution functions, and vice-versa (Kidmo, et al, 2019). The RMSE was expressed in Equation (12).

$$\text{RMSE} = \sqrt{\frac{1}{n} * \sum_{ij=1}^n (m_i - m_j)^2} \quad (12)$$

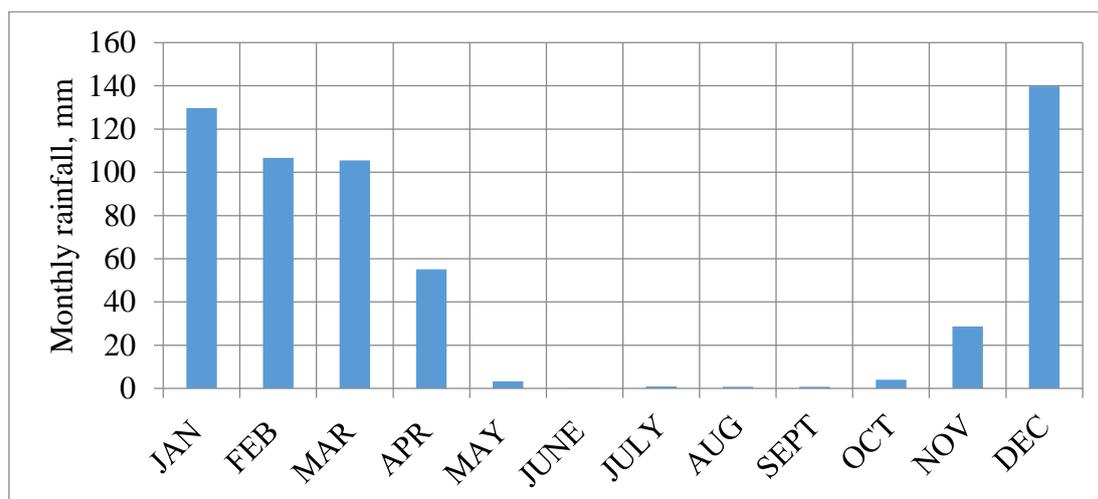Fig. 2. Average daily rainfall from the year 2000 to 2020

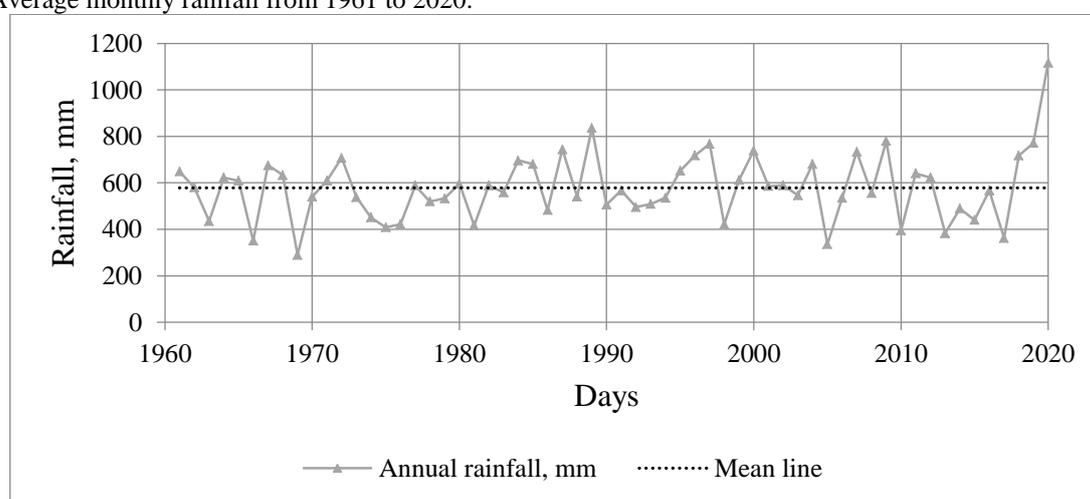Fig. 3. Average monthly rainfall from 1961 to 2020.

Fig. 4. Annual and mean rainfall from the year 1961 to 2020.

Results and discussions

The daily rainfall data was presented in Figure 2 with the highest wetted day shown on the first day-of-month followed by the fourth day for 20 years. Figure 3 shows the average monthly rainfall with a mean of 47.95 mm and a coefficient of variation of 117.9%. Figure 4 shows the annual rainfall with an average of 578 mm and coefficient of variation of 24.6%. The annual minimum and maximum rainfalls were 289 mm in the year 1969 and 1117 mm in the year 2020, respectively.

Shortages of rainfall over two decades

The shortage of rainfall was evaluated as shown in Figure 5. Considering wetted days as the summation of days rains in a given month within twenty years. T_{wd} is the total wetted days divided by the actual number of days in each year over twenty years. T_v is the average wetted monthly per twenty years that had similar trends with different magnitude when compared with the T_{wd} values. The number of rainfall shortages during the wet season concerning actual days within twenty years indicated about 57.9% for January i.e., equal to 100%

minus maximum value of T_{wd} (42.1%) followed by December, February and March with the ratio of 65%. While the long dry period starts early May to late October every year.

Assessment of rainfall variability

Three techniques were used to assess the rainfall variability including daily rainfall index (DRI), standardized anomaly index (SAI), and rainfall seasonality index (RSI) between 2000 and 2020 years, i.e., for two current decades (Figure 6). Ideally, when the DRI is equal to 1, means the total annual rainfall is equal to the number of wetted days. The maximum wetted days of 82 seen in the year 2020 with the annual rainfall of 1117 mm (DRI = 13.6) and minimum wetted days of 33 seen in the year 2005 with the annual rainfall of 336.1 mm (DRI = 10.2). However, the highest DRI of 15.6 was seen in the year 2009 with the wetted days of 50 and annual rainfall of 780 mm; Similarly, a minimum DRI of 8.1 was seen in the year 2006 and annual rainfall of 536 mm with the wetted days of 66. These values of DRI indicated the possibility of less annual rainfall with a less occurrence of rainfall in the same year.

When SAI is closer to zero, annual rainfall is closer to the mean value. The maximum and minimum values of SAI were 2.9 in the year 2020 and -1.5 in the year 2005, respectively. In these two decades, twelve years i.e equivalent to 60% of 20 years had SAI less than zero

which showed the dried years trended from the year 2001 to 2003; 2005 and 2006; 2008, 2010, and 2013 to 2017. Again, when RSI is closer to zero, rainfall spread throughout the year. The average value of RSI was 1.2 with the coefficient of variation of 11.1%; maximum and minimum values were 1.4 and 0.9, respectively.
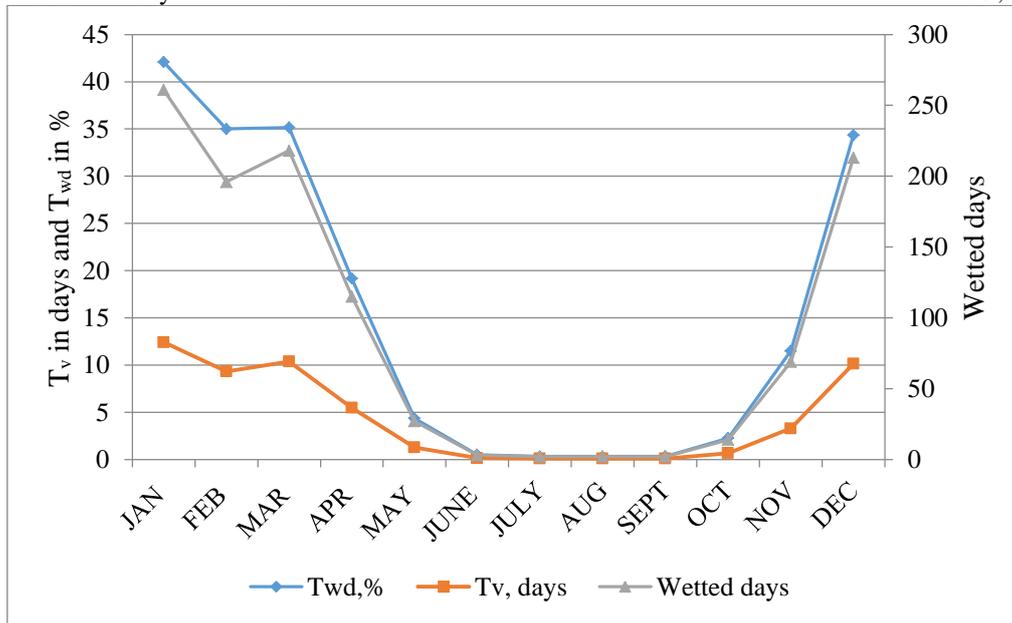

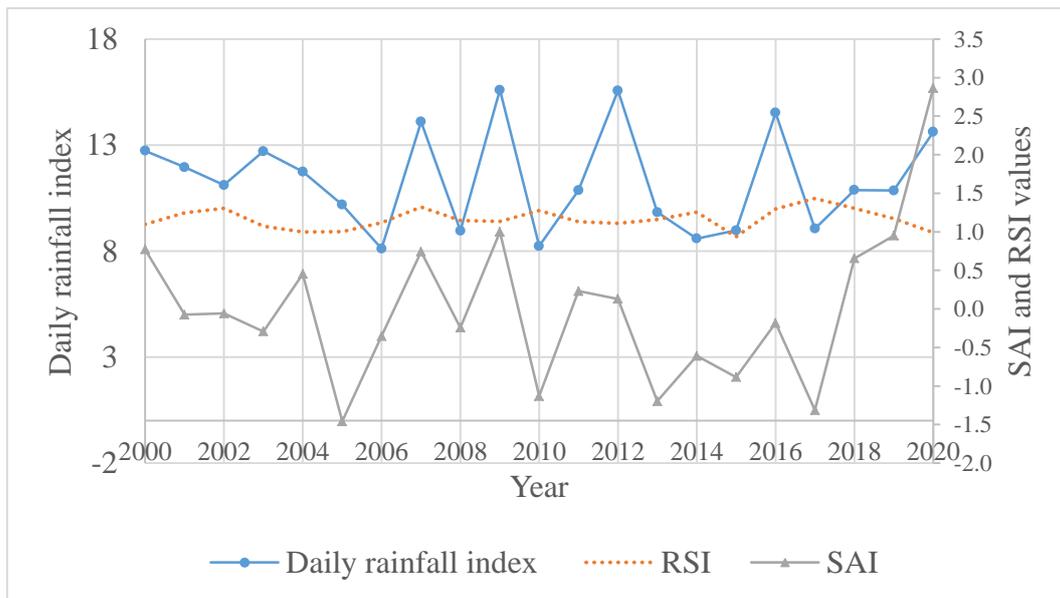Fig. 5. Variation of wetted days for twelve months over 20 years

Fig. 6. Standardized anomaly index for annual rainfall from 1961 to 2020.

The values of RSI indicated the extreme seasonality to the study area and only received rainfall within one to two months. However, three indices were not correlated as shown in Table 2 except for the DRI and SAI of r = 0.5975, but each index provided useful information about the rainfall variability.

Table 2. Correlation of rainfall indices

	DRI	*RSI*	*SAI*
DRI	1		
RSI	-0.0787	1	
SAI	0.5975	-0.2036	1

Selection of probability distribution functions

This study analyses the best fit of the probability distribution functions of rainfall events during November to April for the years 2000 to 2020. Using input parameters in section 2.2, fifteen probability distribution functions were ranked based on the low values of KS and AD tests. Figure 7 shows the variation of ranked probability distribution functions from November to April. All probability distribution functions had the critical value of greater than 0.05 with the ranging between 0.05374 and 0.05467 that demonstrated the normality fit to the given daily rainfall data at a significance level of 5%. All the probability distribution

functions had statistic values greater than the critical value of 2.5018, which helped to reject the null hypothesis of rainfall data originated from a specified probability distribution function at a significance level of 5%. The probability distribution function ranked first for November was Normal and Gamma functions with total days of 637, December was Normal and Weibull functions with total days of 6624; January was Normal and Weibull functions with total days of 620; February was Beta and Gamma functions with total days of 625; March was Normal and Gamma functions with total days of 629; April was Normal and Gamma functions with total days of 617 for KS and AD tests, respectively.
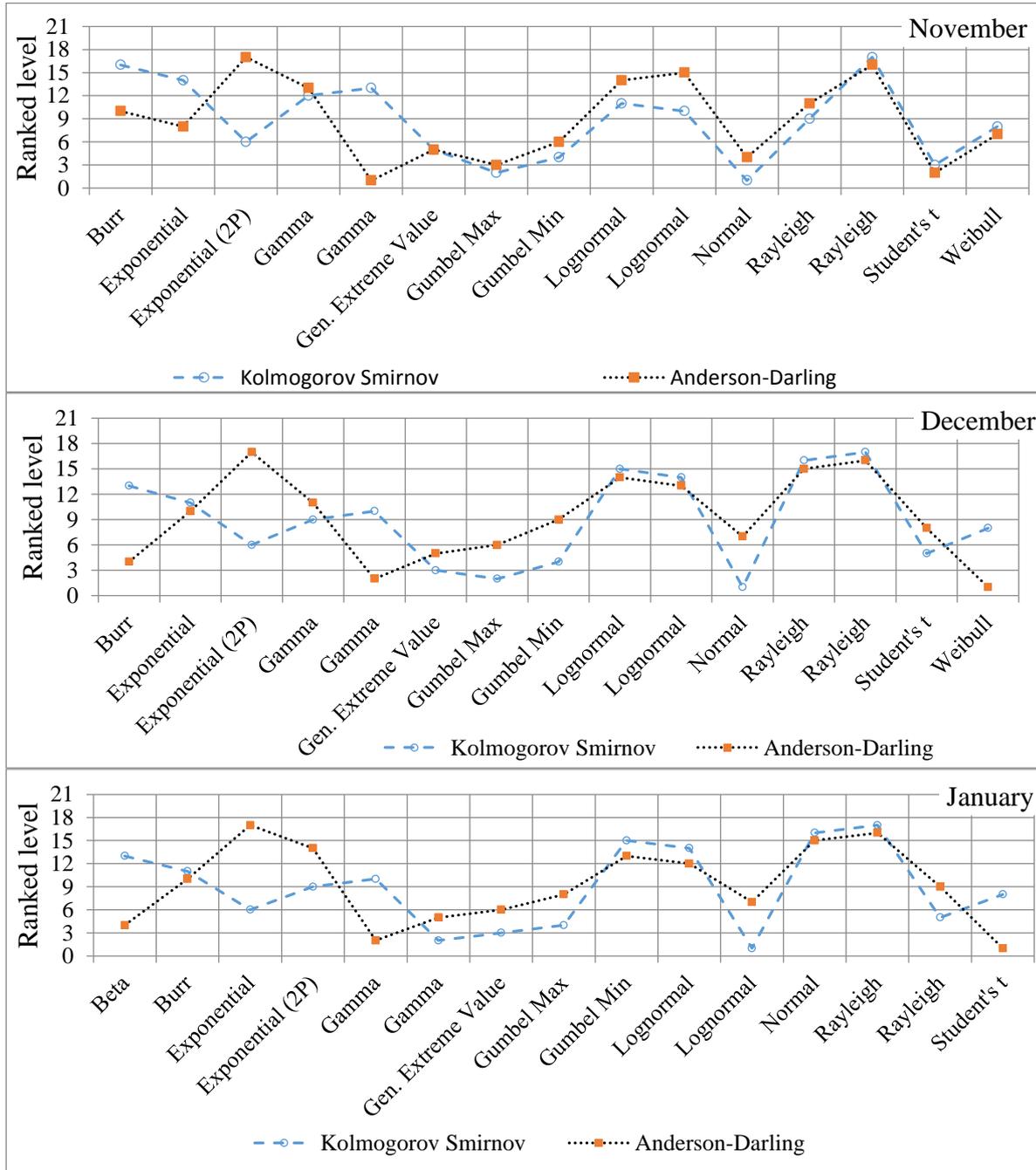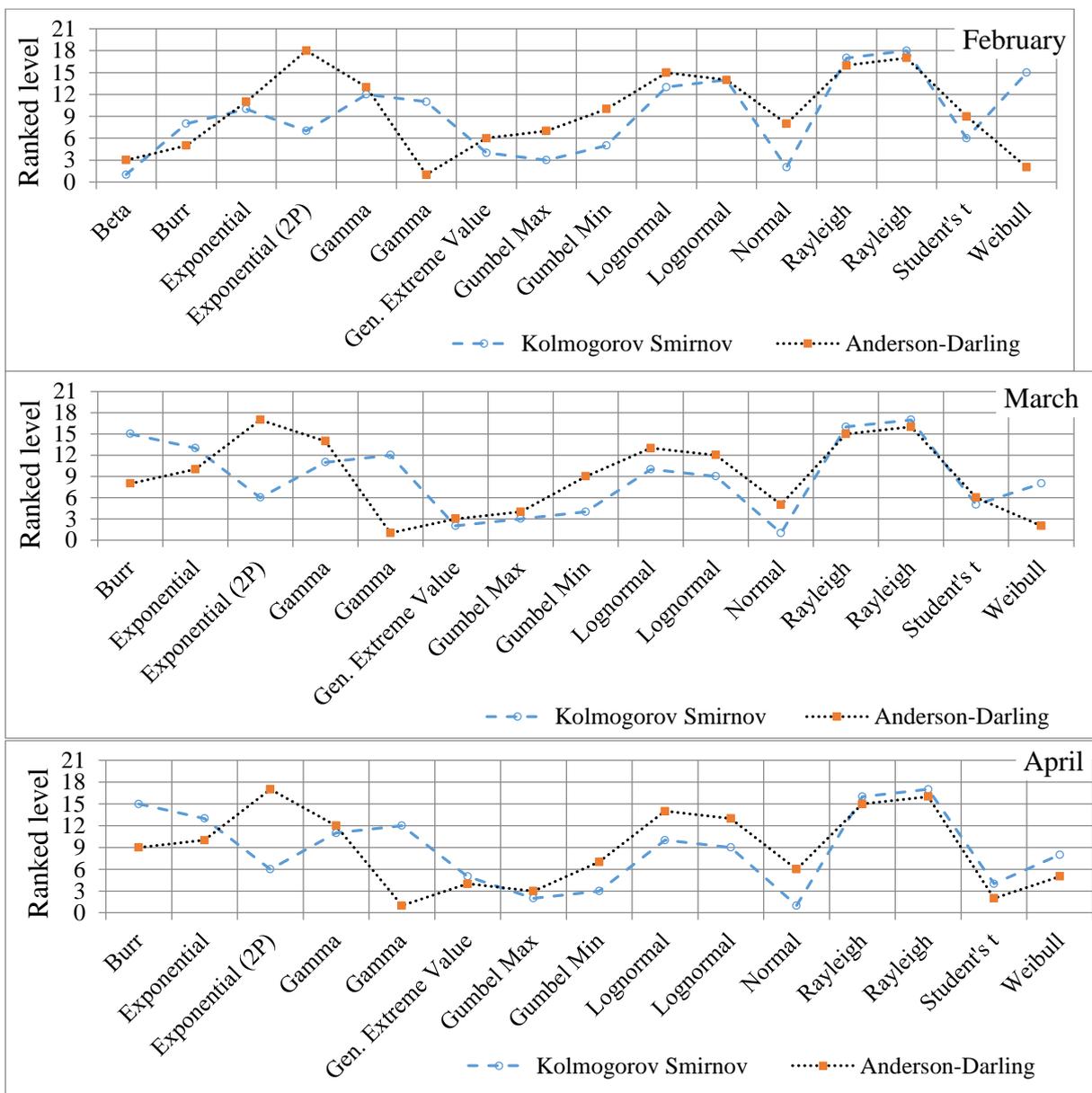
Fig 7. Probability distribution functions ranked with two statistic tests using monthly rainfall for twenty years from Nov-Apr

Table 3. Correlation analysis of the selected distribution functions with the highest-ranked using KS and AD tests

Months	Months	November		December		January	
	Distribution functions	*Normal*	*Gamma*	*Normal*	*Weibull*	*Normal*	*Weibull*
November	Normal	1					
	Gamma	0.5504	1				
December	Normal	0.8380	0.3867	1			
	Weibull	0.9190	0.7782	0.8151	1		
January	Normal	0.7703	0.3548	0.9886	0.7777	1	
	Weibull	0.8942	0.8262	0.7671	0.9961	0.7277	1
February	Beta	0.5586	0.9999	0.3997	0.7862	0.3687	0.8331
	Gamma	0.8947	0.8044	0.8112	0.9974	0.7804	0.9958
March	Normal	0.9037	0.4264	0.9880	0.8525	0.9552	0.8084
	Gamma	0.5993	0.9976	0.4313	0.8172	0.3972	0.8616
April	Normal	0.9201	0.4451	0.9821	0.8677	0.9457	0.8253
	Gamma	0.5691	0.9997	0.4054	0.7938	0.3730	0.8403

Months	Months	February		March		April
	Distribution functions	*Beta*	*Gamma*	*Normal*	*Gamma*	*Normal*
February	Beta	1				
	Gamma	0.8125	1			
March	Normal	0.4383	0.8420	1		
	Gamma	0.9982	0.8401	0.4730	1	
April	Normal	0.4567	0.8560	0.9992	0.4923	1
	Gamma	0.9998	0.8189	0.4454	0.9991	0.4643

Table 4. Plotting position methods with Weibull two parameters, R-squared for line fits and AD test

S/N	Plotting positions	Parameters		R-squared for line fits	AD Test	Weibull mean, mm
		α	β			
1	Hirsch	4.9146	614.2	0.9947	0.0807	563.34
2	CSD	4.8606	623.3	0.9828	0.0778	571.35
3	Laplace	4.3786	628.0	0.9746	0.1287	572.15
4	Landwehr	4.9431	627.5	0.9697	0.0224	575.75
5	Weibull	4.7009	631.0	0.9689	0.0431	577.28
6	McClung	4.9623	628.0	0.9677	0.0198	576.38
7	Adamowski	4.8381	630.1	0.9672	0.0249	577.46
8	Chegodayev	4.8687	630.0	0.9666	0.0225	577.52
9	Beard	4.8750	629.9	0.9665	0.0220	577.49
10	Filliben	4.8791	629.8	0.9663	0.0218	577.47
11	Tukey	4.8899	629.8	0.9662	0.0211	577.47
12	Sevruk and Geiger	4.9173	629.5	0.9656	0.0197	577.47
13	Cunnane	4.9343	629.5	0.9653	0.0190	577.52
14	Blom	4.9281	629.9	0.9647	0.0198	577.83
15	Gringorten	4.9173	629.3	0.9646	0.0182	577.28
16	Hazen	5.0071	629.2	0.9634	0.0177	577.72
17	IEC56	4.9759	630.7	0.9607	0.0224	578.95
	Average	4.8701	628.2	0.9691	0.0354	575.91

Table 3 shows the correlation analysis for selected distribution functions with the highest-ranked using KS and AD tests. The Weibull distribution function had the highest coefficient of correlation (> 0.72) for four probability distribution functions ranked from December to April

Evaluation of plotting position methods and return periods

The return period was evaluated using the plotting position methods and Weibull distribution function. Before estimating the return periods, the plotting position methods were assessed using R-squared coefficients and AD testing, arranged to determine the first and last over 60 years (Table 4). Using R-squared techniques, the Hirsch and IEC56 were selected as the highest and lowest R-squared coefficients of 0.9947 and 0.9691, respectively, i.e., the percentage change of 3%. Hazen and Laplace were selected as having the lowest

and highest values for the AD test of 0.01774 and 0.1287, respectively i.e., the percentage change of 11%. The cumulative probability functions of the four plotting position methods were presented in Figure 8. All plotting position methods had AD values less than the AD critical value, i.e., no significant difference of using one or more plotting methods. Again, RMSE was used to determine the statistical error between the lowest and highest of R-squared coefficients and AD test for the plotting position methods as shown in Table 5. The RMSE varied from 0.24% to 1.88% for six combinations from four selected plotting position methods.

Furthermore, the Weibull parameters were estimated as described in section 2.3. The average annual rainfall based on the Weibull distribution function was 575.9 mm using seventeen plotting position methods with an average shape factor of 4.8701 and scale factor of 628.2. Figure 9 shows the distribution of annual total rainfall against return periods for four selected plotting position

methods. The average rainfall data predicted using the Weibull distribution function is likely to return after 2 years.

The probability distribution functions based on the estimated input parameters from the plotting position methods and measured rainfall data were evaluated.

Table 6 shows the correlation analysis of Weibull density function based on the plotting position methods compared with normal, lognormal, and gamma density functions. The correlation of eight probability density functions indicated a strong relationship with a correlation coefficient greater than 0.8.
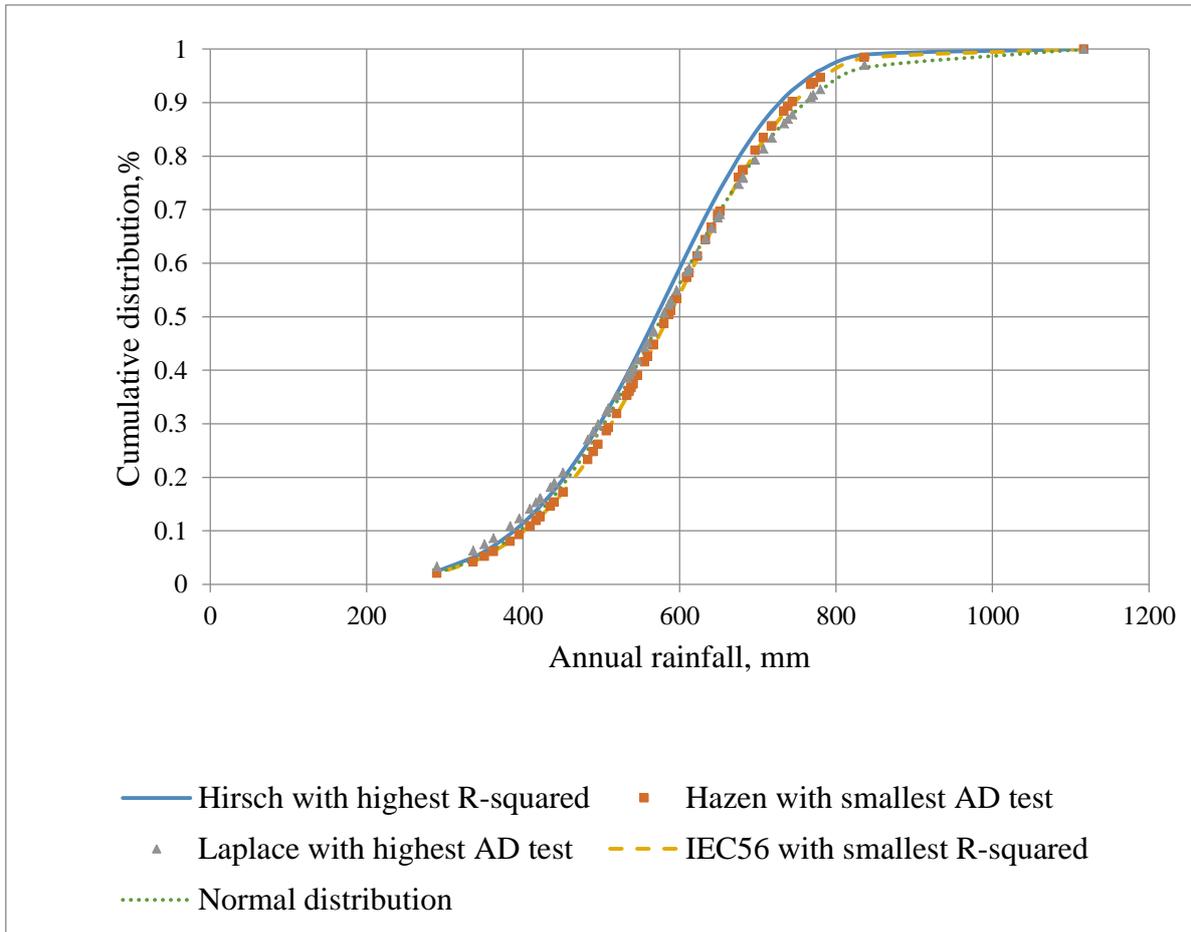

Fig.8. Selected four plotting position methods with the highest and smallest values of R-squared and AD test
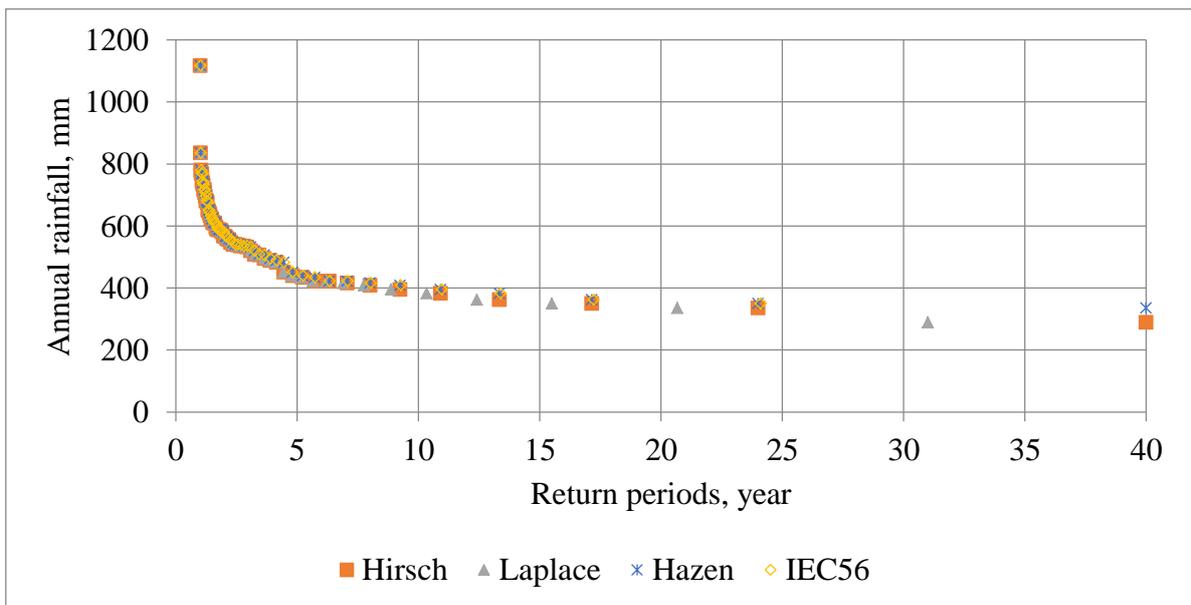

Fig. 9. Annual total rainfall against return periods for four selected plotting position methods

174

Table 5. Statistical errors of selected four plotting position methods using R-squared and AD test

Plotting position methods	RMSE, %		
	Hazen	Laplace	Hirsch
Hazen	1		
Laplace	1.2318	1	
Hirsch	1.6667	1.2676	1
IEC56	0.2396	1.2986	1.8780

Table 6. Correlation analysis of Weibull density function based on the plotting position methods compared with normal, lognormal, and gamma density functions

	Weibull-Hazen	*Log Normal*	*Weibull-Laplace*	*Weibull*	*Weibull-Hirsch*	*Normal*	*Weibull-IEC56*	*Gamma*
Weibull-Hazen	1							
Log Normal	0.8438	1						
Weibull-Laplace	0.9935	0.8878	1					
Weibull	0.9996	0.8570	0.9962	1				
Weibull-Hirsch	0.9843	0.9152	0.9910	0.9878	1			
Normal	0.9845	0.9234	0.9946	0.9886	0.9959	1		
Weibull-IEC56	0.9999	0.8377	0.9928	0.9993	0.9816	0.9826	1	
Gamma	0.9054	0.9915	0.9381	0.9157	0.9586	0.9651	0.9005	1
Weibull-IEC56	0.9843	0.9152	0.9910	0.9878	1.0000	0.9959	0.9816	0.9586

Conclusions

Understanding rainfall variability is very important during the designs of hydraulic structures and agricultural planning to avoid rainfall shortages. The rainfall variability and distribution of the central part of Tanzania has changed during the last 20 years. The study area experienced an unpredicted magnitude of rainfall and drier conditions during the wet season. Four probability distribution functions out of seventeen were ranked first using KS and AD test, respectively for the month ranged between November and April.

The Normal distribution appeared five times followed by Gamma distribution three times, Weibull distribution two times and Beta distribution appeared once. But, the correlation analysis indicated the Weibull distribution had the highest coefficient of correlation (> 0.72) for four distribution functions ranked from November to April.

The return period of rainfall was estimated using seventeen plotting position methods. Two plotting position methods were taken based on the lowest and highest values of the AD test calculated and indicated the percentage changes of 11%. Again, two plotting position methods were taken based on the lowest and highest values of R-squared and indicated the percentage changes of 3%. Taking any of the plotting position methods would provide the return period of rainfall with the RMSE between 0.24% and 1.88%. As a follow to the present study, the rainfall periods and magnitude during wet seasons are expected to be identified for estimating the extreme events.

References

Abdalla, Y., Makokha, M., Maalim, M. K. (2021). Effect of Rainfall Variability on Spring Discharge in the Masingini Catchment, Zanzibar-Tanzania, *International Journal of Environment and Geoinformatics*, 8(1), 39-48, doi. 10.30897/ijegeo. 784000

Alexander, LV. et al., (2019). On the use of indices to study extreme precipitation on sub-daily and daily timescales, *Environmental Research Letters*, 14(13): 125008.

Alonge, A., Afullo, T. (2012). Rainfall drop-size estimators for Weibull probability distribution using method of moments technique, *South African Institute of Electrical Engineers*, 103(2): 83-93,

Al-Suhili, RA., R Khanbilvardi, (2014). Frequency Analysis of the Monthly Rainfall Data at Sulaimania Region, Iraq, *American Journal of Engineering Research*, 3(5): 212-222.

Arvind, G., P A Kumar, S G Karthi, C R Suribabu, (2017). Statistical analysis of 30 years rainfall data: a case study, *In IOP Conference Series: Earth and Environmental Science. IOP Publishing*, 80(1): 012067.

Azad, AK., M G Rasul, T Yusuf, (2014). Statistical Diagnosis of the Best Weibull Methods for Wind power Assessment for Agricultural Applications, *Energies*, 7, 3056-3085.

De Paola, F., M Giugni, M E Topa, E Bucchignani, (2014). Intensity-Duration-Frequency (IDF) rainfall curves, for data series and climate projection in African cities, *SpringerPlus*, 3(1): 1-18

Dirk, R., (2013). Frequency analysis of rainfall data, Katholieke Universiteit Leuven, College on Soil

Physics – 30th Anniversary (1983-2013), Department of Earth and Environmental Sciences, Belgium.

EasyFit 5.0. (2021) MathWave Technologies. Available Online. https://easyfit

Elzopy, KA., A K Chaturvedi, K M Chandran, G Gopinath, U Surendran, Elzopy, K. A., Chaturvedi, A. K., Chandran, K. M., Gopinath, G., Surendran, U. (2020). Trend analysis of long-term rainfall and temperature data for Ethiopia, *South African Geographical Journal*, 1-14, 2020.

Erto, P., A Lepore, (2014). Plotting positions close to the exact unbiased solution: application to the Pozzuoli's bradyseism earthquake data, arXiv preprint arXiv:1412.5663., Naples, Italy.

G J Husak, J Michaelsen, and C Funk, 2006. Use of the gamma distribution to represent monthly rainfall in Africa for drought monitoring applications, *International Journal of Climatology*, no. 10.1002/joc.1441.

Hamisi, J., (2013). Study of rainfall trends and variability over Tanzania, Postgraduate Diploma in Meteorology, University of Nairobi, Nairobi, Kenya.

Joshi, S., K Kumar, V Joshi, B Pande, (2014). Rainfall variability and indices of extreme rainfall-analysis and perception study for two stations over Central Himalaya, India, *Natural hazards*, 72(2): 361-374,

Kang, D., K Ko, J Huh, (2018). Comparative study of different methods for estimating Weibull parameters: A case study on Jeju Island, South Korea, *energies*, 11(356): 1-19.

Kidmo, DK., R Danwe, S Y Doka, N Djongyang, (2015). Statistical analysis of wind speed distribution based on six Weibull methods for wind power evaluation in Garoua, Cameroon, *Revue des Energies Renouvelables*, 8(1): 105-125.

Kisaka, M. (2018). Spatial Variation of Ground water Quality Parameters and its Suitability for Drinking at Makutopora Aquifer, Dodoma Municipality, Tanzania. *International Journal of Environment and Geoinformatics*, 5(3), 337-352, doi. 10.30897/ijegeo.462691

Kisaka, MO., et al., (2015). Rainfall variability, drought characterization, and efficacy of rainfall data reconstruction: case of Eastern Kenya, *Advances in Meteorology*, 2015: 16.

Kist, A., J S.D Filho, (2015). Probabilistic analysis of the distribution of daily rainfall in the State of Paraná, *Ambiente & Água - An Interdisciplinary Journal of Applied Science*, 10(1): 172-182.

Kumar, KSP., S Gaddada, (2015). Statistical scrutiny of Weibull parameters for wind energy potential appraisal in the area of Northern Ethiopia, *Renewables: Wind, Water and Solar*, 2(14): 1-15,.

Kumar, R., A Bhardwaj, (2015). Probability analysis of return period of daily maximum rainfall in annual data set of Ludhiana, Punja, *Indian Journal of Agricultural Research*, 49(2): 160-164.

Kutty, SS., M.G M Khan, M R Ahmed, (2019). Estimation of different wind characteristics parameters and accurate wind resource assessment for Kadavu, Fiji, *AIMS Energy*, 7(6):. 760–791.

Li, Z., Z Li, W Zhao, Y Wang, (2015). Probability modeling of precipitation extremes over two river basins in northwest of China, *Advances in Meteorology*, 2015: 13.

Liu, S., G LI, H Xie, X Wang, (2017). Correlation Characteristic Analysis for Wind Speed in Different Geographical Hierarchies, *Energies*, 10(237): 1-20, Shiyu Liu, Gengfeng Li *, Haipeng Xie and Xifan Wang.

Mafuru, KB., T Guirong, (2018). Assessing prone areas to heavy rainfall and the impaction of the upper warm temperature anomaly during March–May rainfall season in Tanzania, *Advances in Meteorology*.

Markovic, S., N Cerekovic, N Kljajic, N Rudan, (2012). Rainfall analyses and water deficit during growing season in Banja Luka region, *International Congress of Ecologists, Ecological Spectrum*, 1167-1176,.

Marques, R F.P., C R de Mello, A M da Silva, C S Franco, A S de Oliveira, (2014). Performance of the probability distribution models applied to heavy rainfall daily events, *Ciênc. Agrotec., Lavras*, 38(4): 335-342.

Menteş, E. N., Kaya, Ş., Tanık, A., Gazioğlu, C. (2019). Calculation of Flood Risk Index for Yesilirmak Basin-Turkey. *International Journal of Environment and Geoinformatics*, 6(3), 288-299, doi. 10.30897/ijegeo.661533

Moazzam, M. F. U., Vansarochana, A., Rahman, A. U. (2018). Analysis of flood susceptibility and zonation for risk management using frequency ratio model in District Charsadda, Pakistan. *International Journal of Environment and Geoinformatics*, 5(2), 140-153, doi. 10.30897/ijegeo.407260

Moccia, B., C Mineo, E Ridolfi, F Russo, and F Napolitano, (2021). Probability distributions of daily rainfall extremes in Lazio and Sicily, Italy, and design rainfall inferences, *Journal of Hydrology: Regional Studies*, 33,100771.

Nouaceur, Z., O Mursrescu, (2016). Rainfall variability and trend analysis of annual rainfall in North Africa, *International Journal of Atmospheric Sciences*, 2016:16.

Nwaogazie, IL., M G Sam, (2019). Probability and non-probability rainfall intensity-duration-frequency modeling for Port-Harcourt metropolis, Nigeria, *International Journal of Hydrology*, 3(1): 66-75.

Olivera, S., C Heard, (2018). Increases in the extreme rainfall events: Using the Weibull distribution, *Environmetrics*, 1-9,doi.org/10.1002/env.2532

Portela, MM., J M Delgado, (2009). A new plotting position concept to evaluate peak flood discharges based on short samples, *Water Resources Management. WIT Transactions on Ecology and the Environment*, 125, 415-427.

Sasireka, K., C R Suribabu, T R Neelakantan, (2019). Extreme rainfall return periods using Gumble and Gamma distribution, *International Journal of Recent Technology and Engineering*, 8, 4S2.

Saul, AJ. (1996). Hydraulic Assessment. Sewers: Repair and Renovation.": Elsevier, p. 149.

Shaban, AH., A K Resen, N Bassil (2020). Weibull Parameters Evaluation by Different Methods for Windmills Farms, *Tmrees, EURACA, 04 to 06 September 2019*, vol. 6, 188-199.

Sharma, MA., J B Singh, (2010). Use of probability distribution in rainfall analysis, *New York Science Journal*, 3(9): 40-49,.

Sharma, S., P K Singh, (2019). Spatial trends in rainfall seasonality: a case study in Jharkhand, India, *Weather*, 74(1): 31-39.

Suhaila, J.,K Ching-Yee, Y Fadhilah, F Hui-Mea, (2011(. Introducing the Mixed Distribution in Fitting Rainfall Data, *Open Journal of Modern Hydrology*, 1: 11-22.

Sunusi, N. (2017). A comparison between weibull and power law model in estimating conditional intensity function for rainfall data, *International Journal of Management and Applied Science*, 3(2): 16-18.

Syafrina, AH., A Norzaida, A Kartini, K Badron (2018). Comparison of gamma and weibull distributions in simulating hourly rainfall in peninsular Malaysia, *Journal of Fundamental and Applied Sciences*, 10(3S), 331-337.

USGS, United States Geological Survey. (2021). Droughts: Things to Know. www.usgs.gov

Vieira, FMC., J M.C Machado, E de Souza Vismara, J C Possenti, (2018). Probability distributions of frequency analysis of rainfall at the southwest region of Paraná State, Brazil, *Revista de Ciencias Agroveterinarias*, 17(2): 260-266,

W Critchley, K Siegert, and C Chapman. (1991). Rainfall-runoff analysis. www.fao.org/

Yahaya, AS., et al., (2012). Determination of the probability plotting position for type I extreme value distribution, *Journal of Applied Sciences*, 12(14): 1501-1506.

Zengeni, R., V Kakembo, N Nkongolo, (2016). Historical rainfall variability in selected rainfall stations in Eastern Cape, South Africa, *South African Geographical Journal*, 98(1): 118-137.

Zhan, C., W Cao, J Fan, and C K Tse, (2018). Impulse Weibull distribution for daily precipitation and climate change in China during 1961–2011, *Physica A*, doi.org/10.1016/j.physa.2018.07.033.