atmosphere

MDPI

Article

Regionalization of Rainfall Characteristics in Semiarid Botswana Using Gridded Data and L-Moments

Godiraone A. Nkoni [1,*], Kgakgamatso M. Mphale [1,*], Nicholas C. Mbangiwa [1] and Sydney. H. Samuel [2]

[1] Physics Department, University of Botswana, Gaborone Private Bag 0022, Botswana; mbangiwan@ub.ac.bw
[2] Department of Earth and Environmental Sciences, Botswana International University of Science and Technology, Palapye Private Bag 16, Botswana; michaellsydney@gmail.com
* Correspondence: nkoniga@ub.ac.bw (G.A.N.); mphalekm@ub.ac.bw (K.M.M.)

Abstract

The monthly CHIRPS ver. 2 gridded rainfall dataset from 1981 to 2016 was employed to analyze distinct precipitation variability patterns and regimes in semi-arid Botswana. An S-mode eigen analysis was performed on the correlation matrix of the rainfall data to extract principal components. The principal component scores (pc-scores) were further rotated using the Varimax eigen analysis method to yield unique precipitation patterns. The rotated pc-scores indicated three separate sub-regions displaying varying precipitation patterns over time. The application of non-hierarchal clustering (K-means) on the pc-scores identified four distinct zones characterized by unique rainfall patterns. A regional frequency study of rainfall in the sub-regions was performed using L-moments. Probabilistic analysis was utilized to model annual rainfall using six common regional frequency analysis probability distribution functions (pdfs): Pearson Type 3 (PE III); three-parameter Weibull; generalized; extreme value (GEV), normal (GNO), logistic (GLO), and Pareto (GPA). The pdfs that demonstrated the optimal correspondence were determined by the goodness-of-fit test, utilizing the Z-statistic. Each cluster displayed unique pdfs and goodness-of-fit pdfs, with the GLO, GEV, GNO, and Weibull offering the most precise representations.

Keywords: regional rainfall frequency analysis; principal components; L-moments; gridded data; homogeneous zones; S-mode; K-means

1. Introduction

Precipitation is an essential resource that regulates hydrological processes critical for the management and planning of agropastoral, ecological and water resources. Its fluctuation has implications for the lives of sub-Saharan African communities. The communities mostly depend on rain-fed agropastoralism for their sustenance [1]. Consequently, understanding the spatial–temporal variability in rainfall quantity, intensity, frequency, and distribution is essential for the communities. This may involve evaluating the likelihood of extreme precipitation events, a crucial element of water resource management.

The evaluation and modeling of rainfall patterns are frequently conducted using quantiles. Rainfall quantiles are traditionally estimated using at-site frequency analysis. The challenge with the approach lies in the fact that numerous gauge stations in sub-Saharan Africa have brief, fragmentary, and inconsistent historical rainfall records. This may result in erroneous and less robust estimations of precipitation extremes [2]. According to [3], the most effective way to address the issue is to combine data from different locations with comparable precipitation parameters at a target site through a regionalization technique.

Check for updates

Academic Editor: Tomeu Rigo

Received: 30 March 2026
Revised: 17 June 2026
Accepted: 18 June 2026
Published: 23 July 2026

Several interrelated variables can characterize precipitation regimes. Reducing the number of variables before regionalization is crucial because managing their analysis is frequently time-consuming and complex, particularly when dealing with multiple climatic factors that can obscure the underlying patterns in precipitation regimes. Multivariable analysis procedures use dimensionality reduction to discern similarities and differences across variables within a dataset. They examine the interconnections among climatic variables, which are recognized to be closely connected. A predominant method employed for this objective is principal component analysis. The approach reduces dimensionality by creating non-correlated components, which are arranged in decreasing order. The linear statistical technique has been extensively utilized for this purpose in several studies, such as those by [4–6].

After dimensional reduction, clustering techniques, including hierarchical clustering [5], L-moments [7], fuzzy C-means [8], self-organizing maps [9], and K-means [10]), can be used to classify a variety of precipitation gauge stations into homogeneous sub-regions. The pdf of gauge stations within the specified precipitation sub-regions will be assumed to be similar. The distribution functions are determined using mathematical statistics.

In terms of elucidating the variability of rainfall within homogeneous regions, each regionalization method has its own unique advantages and limitations. Eslamian and Feizi [11] observe that L-moments-based regional frequency analysis (Rfa-Lmom) generates predictions and results that are exceedingly precise when implemented in conjunction with distribution parameters and goodness-of-fit evaluations. The effectiveness of L-moments is attributed to its low susceptibility to sample variability effects. It is also recognized for its robustness to outliers and unbiased nature in small samples [12].

Several hydro-meteorological investigations worldwide have implemented the Rfa-Lmom methodology for regionalization purposes because of its theoretical advantages. For example, after considering the heterogeneity in certain ISMR zones identified by the India Meteorological Department (IMD), Satyanarayana and Srinivas [13] proposed using large-scale atmospheric variables obtained from a high-resolution IMD gridded dataset as regionalization vectors to create new homogeneous sub-regions. Using the L-moment goodness-of-fit test along with K-means clustering, Satyanarayana and Srinivas were able to redefine seventeen (17) homogeneous sub-regions, where the rainfall probability distribution function is mostly described by the three-parameter log-normal (LN3) distribution.

In their paper, Nam et al. [14] consider the rainfall regime in South Korea. To investigate regional differences in the precipitation patterns, the authors used fuzzy C-means clustering and L-moment regional frequency analysis techniques. Using fuzzy C-means clustering, the sixty-seven (67) gauge stations across the country were clustered into five (5) homogeneous precipitation zones described by GLO, GEV, GNO, and GUM probability distributions. A similar study was performed by Hazarika and Sarma [15], who considered the rainfall data in Northern Brahmaputra Valley of Assam. Applying the same technique of Fuzzy C-means clustering, the authors managed to classify eighty-three (83) gauging stations located in this valley into three (3) homogeneous precipitation zones, considering L-moments to assess zone homogeneity.

Another similar example is presented by the findings of Dikbaş et al. [16] in Turkey, who created six (6) homogeneous zones based on hydrology of the area and statistically verified them for homogeneity applying the L-moment heterogeneity test. In a more recent work by Gocic et al. [17], annual precipitation data from twenty-eight (28) stations in Serbia were analysed. Using L-moment diagrams and goodness-of-fit criteria (relative error and correlation coefficient), the authors found that the GEV probability distribution function better describes the precipitation regime in most regions than GLO and GPA distributions. Furthermore, Guttmann [18] used L-moment diagrams to subdivide the continent of America into one hundred four (104) homogeneous zones.

A limited number of studies were undertaken to investigate regionalization and spatial–temporal patterns of precipitation in Botswana. The investigations include [6,12] and [19–21]. Most of the studies were limited by the duration of the precipitation time series, and the allowable number of precipitation gauge stations, as can be seen in [19]. Jamshidi and Samani [22] assert that the distribution of gauge stations and their quantity, as well as the duration of rainfall data, affect regionalization. Despite the longer rainfall data records and a larger number of gauge stations analyzed by Chipanshi and Maphanyane [21] and Nkoni et al. [6], the spatial distribution of these gauge stations was uneven.

According to Sahu et al. [23], the uneven gauge station distribution leads to regional bias and inadequate cluster delineation when the regionalization process is carried out. In addition, management of agricultural, natural, and water resources is contingent upon the regional variability of climatic conditions. Even so, climate factors observed at a given gauge station are particular to the site. Station observations in sub-Saharan Africa frequently exhibit inadequate spatial coverage, biases linked to gauge-measured processes, and issues related to missing data. This can be mitigated through the utilization of high-resolution gridded databases [24,25].

Disagreement exists in the number of clusters determined in the afore-mentioned investigations, varying from one to eight, which complicates the interpretation of the regionalization results. All the investigations exclusively employed gauge station data. A complementary study that uses high temporal–spatial resolution data and Rfa-Lmom could validate the results of the prior research. In this regard, satellite precipitation estimation products offer an extensive and consistent time series of precipitation data with comprehensive spatial coverage. The precipitation estimation products address the deficiencies caused by sparse gauge station networks (see [26,27]).

Several freely accessible satellite estimate databases, such as the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS; Funk et al. [28]) and the Tropical Rainfall Measuring Mission (TRMM; Kummerow et al. [29]), can function as proxies for precipitation measurements in sparse gauge network situations. CHIRPS (ver. 2.0), utilizing integrated satellite infrared and gauge station data, has exhibited enhanced performance relative to other satellite precipitation estimates, especially in semi-arid regions (e.g., in [30,31]). The assertion is substantiated by the finding that the correlation between CHIRPS (ver. 2.0) and gauge station datasets in South Africa and Eswatini is robust at the monthly scale, with averages of 0.77 and 0.73 in correlation coefficients, respectively [32,33]. However, the efficacy of CHIRPS (ver. 2.0) depends on regional climatic variances. Therefore, the dataset requires validation prior to its application.

In arid and semi-arid regions, the amount of rainfall received annually is very little, highly variable, highly uneven, inconsistent, and seasonally focused. Even though dry years are prevalent in arid and semi-arid regions, the years where there is rainfall tend to be very wet years, implying a heavy-tailed distribution. The main reason for the occurrence of rainfall in these regions is due to the occurrence of convectional storms, which are rare and unpredictable. Tropical storms, cutoff lows, and meridional moisture transport all play a role in causing high amounts of rainfall in the regions [34].

Heavy-tail distributions that are sufficiently flexible to account for positive skewness may be used to model the disproportionately large occurrences due to the cutoff low-pressure systems and rare tropical invasions in semi-arid regions. Seven (7) of the thirteen (13) pdfs in the statistical toolbox of the R software (ver. 4.5.2) package L-mom [7] are three-parameter distributions. GEV, GLO, LN3, GNO, WAK, PE III, and GPA are the three-parameter distributions. The skewed, heavy-tailed, and physically constrained rainfall data typical of arid and semi-arid regions may be successfully replicated by the three-parameter distributions. While retaining a single regional shape parameter, the three-parameter distributions provide enough flexibility to capture a greater variety of site-specific behaviors within a homogeneous region. An essential analytical tool for climate regionalization, the L-moment ratio diagram, provides a distinct curve for each three-parameter distribution.

This study is aimed at evaluating the effectiveness of CHIRPS (ver. 2.0) as a substitute for gauge station data and to use the high-resolution gridded dataset to identify homogeneous precipitation zones in semi-arid Botswana through the K-means clustering method. The study employed Rfa-Lmom to determine the most suitable three-parameter pdfs for precipitation in these zones. Assessing rainfall quantity, duration, and frequency can help identify regions vulnerable to extreme precipitation events and enhance accuracy in rainfall modeling. The evaluation is predicated on the optimal three-parameter probability density function.

The manuscript is organized as outlined below: Section 2 gives a short account of the location of the study and the research data. The subsequent section summarizes the methodology used in multivariate analysis, Rfa-Lmom regionalization, and probabilistic statistics for fitting distribution functions. Section 4 discusses the results from the study. Section 5 of the manuscript contains an overview of the findings in the investigation.

2. Study Area

Situated in the longitudes of 20° E and 29° E and latitudes of 18° S and 27° S, Botswana is in the center of the extensive southern African plateau and has an area that covers 582,000 square kilometers (Figure 1). The country is separated into two major ecoregions, *Sandveldt* and *Hardveldt*, and has a semi-arid to dry climate. Less than 5% of the land area is suitable for farming that relies on rainfall. The latter is in the *Hardveldt* ecoregion. A large percentage of the country's landmass, mainly in the *Sandvelt* region, has surface water scarcity.

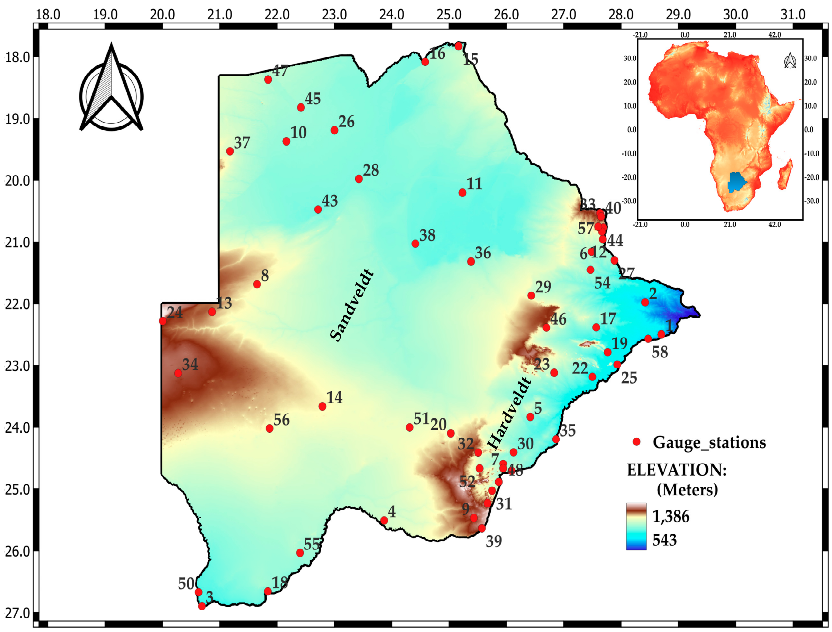


Figure 1. Location and topography of gauge stations in the study area.

The Kalahari Basin, a lowland depression within the plateau, constitutes more than 70% of the country's landmass. The basin has a gently sloped plane that is around 1100 m above sea level on average. The Kalahari Basin is influenced by several major systems that bring rainfall, with the Intertropical Convergence Zone being the most notable. The Intertropical Convergence Zone (ITCZ) shifts southward throughout the summer in the Southern Hemisphere. The southward displacement of the ITCZ results in the division of the high-pressure ridge over the subcontinent, leading to persistent aridity in winter. This separation permits the low-pressure system termed the "easterly wave" in the Indian Ocean to infiltrate the area of summer precipitation, culminating in significant convective rainfall in northern Botswana.

In late summer, the Kalahari heat forms low in the southwestern region of the country [35]. This system is characterized by a weak upward airflow in the lower atmosphere and a downward airflow in the upper atmosphere. The thermal low induces the movement of moisture-rich air from the subtropical southwestern Indian Ocean to the subcontinent. The Drakensberg Mountains serve as a barrier, obstructing the ingress of humid easterly air into the Kalahari Desert. This deviation results in winds arriving at the Kalahari from the northeast, a somewhat longer and more turbulent path. The considerable distance air must traverse to reach the Kalahari results in significant moisture loss through evaporation, accounting for the low annual precipitation observed in the southwestern part of the country.

The distribution of occurrences is severely skewed, with >60% occurring during the summer months (December–February; see Figure 2). The amount of rainfall differs throughout the country, with the southwestern region receiving an average of less than 300 mm per year, while the northern part receives up to 640 mm. The country cannot possess a parent distribution function to characterize precipitation due to varying rainfall-generating mechanisms, including extensive atmospheric circulation systems and geographical diversity, which lead to significant differences in rainfall patterns across different regions.

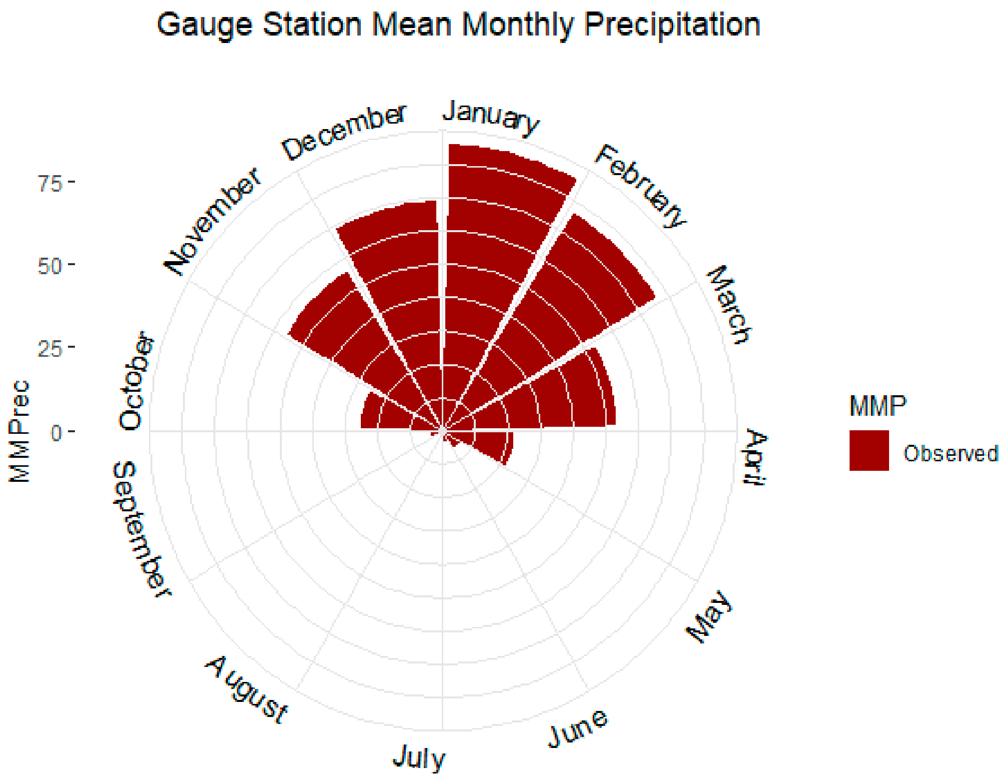


Figure 2. A polar plot of station average monthly rainfall in Botswana.

3. Data and Methodology

3.1. Gauge Station Data

The Department of Meteorological Services for Botswana provided monthly rainfall data from fifty-eight (58) stations for this investigation. The data were gathered during a span of 36 years (1981–2016). The gauge stations considered are dispersed throughout the area of study, but their density is low and uneven (see Figure 1). The data were used in a previous publication, Nkoni et al. [6]. In the article, the data were tested for homogeneity and were found to be homogeneous at the 95% significance level.

According to Nashwana et al. [36], when evaluating the spatial variability of rainfall, it is not advisable to rely on gauge station data that are irregularly distributed. The data from the gauge stations were employed in conjunction with gridded CHIRPS monthly data to validate the performance of the database for use as a substitute source of precipitation data.

3.2. Gridded Data

3.2.1. Precipitation Data

CHIRPS (ver. 2.0) gridded monthly precipitation data estimates were used in the study. The data can be extracted from the University of California's website (see [33]). The CHIRPS dataset was created by integrating the TMPA 3B42 v7 products with data from various meteorological agency stations worldwide. It offers spatial resolutions of 0.05° and 0.25°, covering the globe from 50° S to 50° N. The updated second edition of CHIRPS delivers improved daily rainfall time series from 1981 to the present. The considerable development of CHIRPS and its use for drought monitoring in Africa are explained in detail by Funk et al. [37].

The investigation utilized a 0.25° resolution CHIRPS dataset. The spatial resolution reduces noise and improves alignment with the available ground data in areas with fewer rain gauges. The application scope and computing efficiency also influenced the spatial resolution selection, as higher resolutions can provide more detailed information but require significantly more computational resources and time to process.

3.2.2. Temperature Data

The daily maximum and minimum air temperature data from CHIRTS (ver. 1.0) were amalgamated to calculate the mean monthly values. The spatial extent of the dataset is quasi-global, encompassing latitudes from 60° S to 70° N. The satellite estimations are obtained from infrared satellite observations, European Reanalysis Version 5 (ERA 5), and data gathered from ground stations. The monthly CHIRTS dataset was then aggregated to a resolution of 25 km to align with the CHIRPS data. Keoagile et al. [38] indicated that the CHIRTS dataset had adequate performance when assessed as a substitute for in situ temperature data in Botswana.

3.3. Validation of CHIRPS Dataset

The study used the pairwise method of point-pixel analysis to validate the performance of a high-resolution CHIRPS (ver. 2.0) monthly rainfall dataset. The gauge stations were points, and the CHIRPS (version 2.0) dataset comprised pixels. The monthly data from the gauge stations were juxtaposed with the matching monthly CHIRPS pixel data. The validation employed point-to-pixel analysis due to the potential for uncertainties in the interpolation of the observed or gauge data.

The capability of CHIRPS (ver. 2.0) to substitute gauge station precipitation amounts was analyzed through four metrics: Kling–Gupta efficiency, percentage bias, correlation coefficient, and root mean square error. The indices are described in Table 1. The correlation coefficient indicates the degree of association between gauge station data and satellite data, and the percentage bias assesses the accuracy of CHIRPS (ver. 2.0) in predicting precipitation relative to gauge stations. The ideal values for the correlation coefficient and percentage bias are 1 and 0, respectively.

Table 1. Description of continuous indices for CHIRPS validation (see Bai and Liu [39]).

Index:	Equation:	Perfect Score:
Percentage Bias (*PBIAS*):	$PBIAS = 100 \times \frac{\sum_{k=1}^{n}(S_k - O_k)}{\sum_{k=1}^{n}(O_k)}$	0
Correlation Coefficient (*r*):	$r = \frac{\sum_{k=1}^{n}(O_k - \overline{O}_k)(S_k - \overline{S}_k)}{\sqrt{\sum_{k=1}^{n}(O_k - \overline{O}_k) \times \sum_{k=1}^{n}(S_k - \overline{S}_k)}}$	1
Root Mean Square Error (*RMSE*):	$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(S_i - \overline{S}_i)^2}{N}}$	0
Kling–Gupta Efficiency (*KGE*):	$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_{sim}}{\sigma_{obs}} - 1\right)^2 + \left(\frac{\mu_{sim}}{\mu_{obs}} - 1\right)^2}$	1

3.4. Principal Component Analysis (PCA)

PCA is an orthogonal transformation process that lowers the redundancy caused by variable correlation in a dataset. It can also be used to preprocess data before cluster analysis. The technique involves computing the eigenvalues and eigenvectors of a covariance or correlation matrix from the dataset to be analyzed.

Consider a monthly rainfall data matrix $\mathbf{R}$ of dimension $\mathbf{s} \times \mathbf{m}$. Here, $\mathbf{m}$ is the rainfall data time step, and $\mathbf{s}$ denotes the grid point or gauge station considered in a study. If $\boldsymbol{Q}$ represents the spatiotemporal variation in rainfall in the study area, then $\boldsymbol{Q}$ can be related to spatial and time matrices as follows:

$$\boldsymbol{Q} = \boldsymbol{UV} \tag{5}$$

Here, $\boldsymbol{U}$ and $\boldsymbol{V}$ are time and spatial variation matrices, respectively.

The computation of principal components can be achieved through various modalities, the selection of which is contingent upon the configuration of the feature vector (data). One of the methods is given below:

Let the elements $\boldsymbol{q_{fg}}$ of $\boldsymbol{Q}$ be given by the following relation.

$$\boldsymbol{q_{fg}} = \sum\nolimits_{k=1}^{s} \boldsymbol{f_{fk}} \times \boldsymbol{v_{kg}} \tag{6}$$

Considering $\mathbf{V}$, in Equation (5), to be orthonormal, then

$$\boldsymbol{I} = \boldsymbol{VV'} \tag{7}$$

Here, $\boldsymbol{I}$ and $\mathbf{V'}$ are the identity and transpose of the $\boldsymbol{V}$ matrices, respectively.

To solve for $\boldsymbol{U}$ and $\boldsymbol{V}$, a matrix $\mathbf{Z}$ related to $\boldsymbol{Q}$ could be assumed to be of the following form:

$$\boldsymbol{Z} = \boldsymbol{Q'Q} \tag{8}$$

It then follows that

$$\boldsymbol{V'ZV} = \boldsymbol{UU'} = \boldsymbol{J} \tag{9}$$

Here, $\boldsymbol{J}$ is a diagonal matrix. The columns of $\boldsymbol{V}$ are eigenvectors or principal components of $\boldsymbol{Z}$, and the elements of $\boldsymbol{J}$ are the eigenvalues of $\boldsymbol{Z}$. Each element represents the proportion of variance that the principal components account for.

The eigenvectors provide valuable information about the proportionate contributions to the newly generated components from the original data. In contrast, the eigenvalues quantify the degree to which each newly introduced variable explains the observed variation.

In this study, PCA was used for semi-qualitative regionalization and data preprocessing procedures for K-means. The number of principal components to be kept for use as input in K-means was determined using parallel analysis [40] and the scree plot [41].

3.5. K-Means Clustering

A clustering method that is non-hierarchical, known as K-means, is often employed to identify the inherent correlations among objects. The clustering technique relies on the level of similarity across things. The objects were divided into a predetermined number of $\boldsymbol{J}$ clusters, each defined by centroids. The centroids ($\boldsymbol{C_j}$) were computed as follows:

$$C_j = \frac{1}{n_j}\left\{\sum_{\forall Y_P \in C_j} Y_p\right\} \tag{10}$$

where $\boldsymbol{Yp}$ signifies objects and $\boldsymbol{n_j}$ signifies the number of objects within the cluster.

The clustering method requires the user to provide a predetermined number of clusters, typically between 2 and 10, with cluster centroids being randomly chosen (see [42]). The distances between objects within the same cluster and those in different clusters should be high. To allocate objects to the closest cluster and their centroids, the K-means method uses the Euclidean distance (μ). In reducing the distance between objects within the same cluster, the methodology employs an objective function that is determined as follows:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{m} \left\| S_i^j - C_j \right\|^2 \tag{11}$$

where $\boldsymbol{\mu} = \left\| S_i^j - C_j \right\|^2$ is the Euclidean distance. The flowchart shown in Figure 3 summarizes the steps carried out in the K-means clustering.

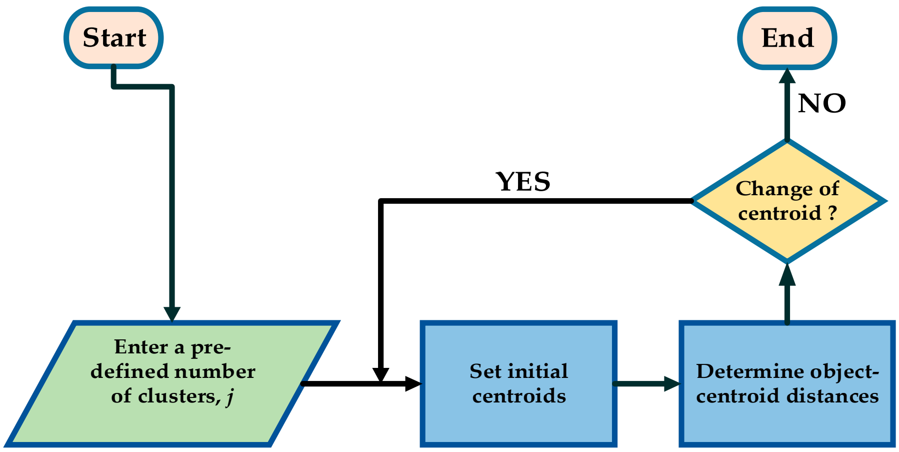


Figure 3. A flowchart depicting the K-means clustering algorithm used in the study.

Upon conclusion of K-means clustering, the silhouette index [43] was employed to assess the quality of the clustering results. Thereafter, the homogeneity and rainfall distribution within the clusters (sub-regions) were assessed utilizing L-moments statistics (Hosking [7]), as described in Section 3.7.

3.6. Quality of Clustering

The silhouette index is a popular metric that analyzes the quality of formed clusters. It calculates the average silhouette width within a cluster. The index is determined using the following relation (Rousseeuw [43]):

$$\boldsymbol{S(k)} = \left\{ \frac{\boldsymbol{b(k) - a(k)}}{\boldsymbol{max(a(k), b(k))}} \right\} \tag{12}$$

where the sample's k-mean distance from other samples in the same cluster is denoted by ***a(k)***, while the minimum average distance between the sample k and samples in any other cluster is denoted by ***b(k)***. When the **S(k)** value is zero, the sample must be relocated to the nearest cluster; conversely, when the value is one, the chosen cluster is the most appropriate for the sample.

Dudek [44] notes that even though the silhouette index has the advantage of low computational complexity and simple interpretation rules, its score can be significantly affected by noisy data. Furthermore, Modarres and Sarhadi [45] assert that relying solely on cluster analysis is insufficient to verify that rainfall sub-regions are homogeneous. It is recommended to assess the homogeneity of the clusters utilizing the H-statistic derived from L-moments statistics.

3.7. The L-Moment Method

3.7.1. Linear Moments

An estimation method for frequency analysis that is *less sensitive to sample variability* and robust in the presence of extreme values is the L-moment method. It is a mathematical statistics approach based on probability-weighted moments (PWMs) combined linearly as described by Hosking and Wallis [46]. According to Novo and Loggia [47], PWMs are described by the following relationship:

$$\boldsymbol{\beta_r = E\left\{a_x[F(a_x)]^r\right\}} \tag{13}$$

where $\boldsymbol{a_x}$ is a random variable with a cumulative distribution function $\boldsymbol{F}$ and **r** = 0, 1, 2 . . . are positive integers.

According to the linear combination of PWMs proposition, the following relation could provide the L-moments:

$$\boldsymbol{\lambda_{r+1} = \sum_{k=0}^{r} p_{r,k}\beta_k} \tag{14}$$

where $\boldsymbol{p_{r,k}} = -1^{\boldsymbol{r-k}}\binom{r}{k}\binom{r+k}{k}$.

The initial four moments derived from Equation (10) are presented as follows:

$$\boldsymbol{\lambda_1 = \beta_0} \tag{15}$$

$$\boldsymbol{\lambda_2 = 2\beta_1 - \beta_0} \tag{16}$$

$$\boldsymbol{\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0} \tag{17}$$

$$\boldsymbol{\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0} \tag{18}$$

The probability distribution's mean, dispersion, skewness, and kurtosis are quantified by the L-moments λ_1, $\boldsymbol{\lambda_2}$, $\boldsymbol{\lambda_3}$, and $\boldsymbol{\lambda_4}$, as determined from Equation (15) to (18). The mean distribution function ($\boldsymbol{\mu}$) = ***E*[*X*]** is represented by $\boldsymbol{\beta_0}$ when ***r*** = **0** in Equation (11). The ratio $\boldsymbol{\tau_2} = \boldsymbol{L_{CV}}= \lambda_2/\lambda_1$ represents the L-coefficient of variation (*dispersion*), whereas the ratios

$\tau_3 = L_{skew} = \lambda_3/\lambda_2$ and $\tau_4 = L_{kurt} = \lambda_4/\lambda_2$ denote the L-coefficients for skewness and kurtosis, respectively.

Four essential procedures are performed during regional frequency analysis. The techniques include (*a*) screening data for regional analysis, (*b*) identifying homogeneous sub-regions, (*c*) determining appropriate frequency distributions for the sub-regions, and (*d*) calculating rainfall quantiles.

3.7.2. Discordance Test

In the first stage, grid points without a shared rainfall distribution function with other sites in the sub-region were excluded from the study or reassigned to different clusters. The grid points will normally be distant from others in an L-moment ratio diagram and discordant.

A statistic that evaluates whether a particular grid point should be omitted from the study is the D_i -statistic. Hosking and Wallis [46] describe the D_i -statistic with the following expression:

$$D_i = \frac{1}{3}(u_i - \overline{u})^T S^{-1}(u_i - \overline{u}) \tag{19}$$

where u_i denote a vector of L-moment ratios for grid point i, whereas $\overline{u}$ and S signify the sample mean and covariance matrix of u, respectively. Grid points with a Di-statistic over 3 were considered grossly discordant in relation to the cluster.

3.7.3. Heterogeneity Test

Hosking and Wallis [46] suggested a test statistic, H, for evaluating the heterogeneity of a cluster. It contrasts the inter-grid point variability in a sample of L-moments for a group of points with the anticipated values for a homogeneous region. The H statistics are derived from the following equation:

$$H = \frac{(V - \mu_v)}{\sigma_v} \tag{20}$$

where $V = \frac{\sum n_i \left(L_{cv}^{(i)} - \overline{L}_{cv}\right)^2}{\sum n_i}$ while μ_v and σ_v represent the sample mean and standard deviation of $L_{cv.}$ The criteria for evaluating heterogeneous regions are as follows:

- $H_1 < 1$ is sufficiently homogeneous.
- $1 < H_1 < 2$ is potentially homogeneous.
- $H_1 \geq 2$ is unequivocally heterogeneous.

3.7.4. Probability Distribution Functions

Upon identifying the homogeneous regions, suitable frequency distributions for precipitation were established. Some of the most prevalent distributions employed to model meteorological occurrences are the following: the three-parameter Weibull, PE III, GLO, GEV, GNO, and GPA. The probability distribution functions were fitted to precipitation data, and the $\mathbf{Z^{dist}}$-index was utilized as a goodness-of-fit metric to determine the ideal function that characterizes the data. The following expression defines the Z-index:

$$Z^{dist} = \frac{\left(\overline{L}_{kur} - L_{kur}^{dist}\right)}{\sigma_{kur}} \tag{21}$$

The fit is adequate at a 10% confidence level when $\left|Z^{dist}\right|$ is less than 1.64. Table 2 provides a description of the distribution functions.

Table 2. Probability distribution functions used in the study. (Note: In the equations, $\boldsymbol{\xi}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\kappa}$ are location, scale, and shape parameters for distributions, respectively. $\boldsymbol{\Gamma}$ is the normalization factor and $\boldsymbol{Y = -\kappa^{-1} log(1 - \kappa(x - \xi)/\alpha)}$. See [48,49].

Distribution:	Probability Distribution Function (Pdf):	Quantile Function:
Pearson Type 3 (PE III)	$f(x) = \frac{1}{\alpha\, \Gamma(\kappa)}\left(\frac{x-\xi}{\alpha}\right) exp\left(-\frac{(x-\xi)}{\alpha}\right)$	$X(F) = \xi + \frac{\alpha}{\kappa}\{1 - (-lnF)^{\kappa}\}$
Generalized Pareto (GPA)	$f(x) = \alpha^{-1} exp\{-(1-\kappa)Y\}$	$X(F) = \xi + \frac{\alpha}{\kappa}\{1 - (1-F)^{\kappa}\}$
Generalized Extreme Value (GEV)	$f(x) = \alpha^{-1} exp\{-(1-\kappa)Y - exp(-Y)\}$	$X(F) = \xi + \frac{\alpha}{\kappa}\{1 - (-lnF)^{\kappa}\}$
Generalized Logistic (GLO)	$f(x) = \frac{\alpha^{-1} exp\{-(1-\kappa)Y\}}{[1+exp(-Y)]^2}$	$X(F) = \xi + \frac{\alpha}{\kappa}\left\{1 - \left(\frac{1-F}{F}\right)^{\kappa}\right\}$
Generalized Normal (GNO)	$f(x) = \frac{1}{\alpha\sqrt{2\pi}} exp\left\{\kappa Y - \frac{Y^2}{2}\right\}$	$X(F) = \xi + \frac{\alpha}{\kappa}\{1 - exp\left(\kappa\theta^{-1}(F)\right)\}$
Three-parameter Weibull	$f(x) = \frac{\kappa}{\alpha}\left(\frac{x-\xi}{\alpha}\right)^{\kappa-1} exp\left\{-\left(\frac{x-\xi}{\alpha}\right)^{\kappa}\right\}$	$X(F) = 1 - exp\left(\frac{X-\xi}{\alpha}\right)^{\kappa}$

3.8. Grid Points Subsampling

The grid points from homogeneous sub-regions, which exhibited super-homogeneity, were intentionally reduced and spatially distributed within a sub-region to mitigate or eradicate the spatial connection between sites that skews the homogeneity test. When grid points are closely spaced, their rainfall records exhibit dependence. The kappa-distribution simulations used to calculate H1 presume independence. Subsampling reduces the network density, ensuring that the remaining points are sufficiently spaced to function almost independently, hence making the H1 statistic interpretable at face value instead of unnaturally negative.

Dense grid points result in several adjacent locations exhibiting virtually similar L-moment statistics. Establishing a minimum separation distance, such as retaining every third or fourth point, eliminates redundant correlated information and provides H1 with a valid basis for comparison against the simulated reference distribution.

The systematic grid thinning of the super-homogeneous zones was executed by recalculating regional L-moment statistics and H1 on the diminished dataset. This is repeated across many random subsamples of same size to evaluate stability. The H1 values from subsampled sets should converge to the range of −1 to +2 if the original negativity was solely a correlation artifact. If they continue to exhibit significant negativity after rigorous thinning, the area indeed has little between-site variability, which strongly supports homogeneity irrespective of the correlation issue.

3.9. Estimation of Rainfall Quantiles

Assuming a uniform probability distribution function over all grid points, differing only in grid-specific scale factors, the following equation can be employed to generate rainfall quantiles at grid points utilizing regional quantiles established from optimal distribution fit.

$$\boldsymbol{P_i(T) = \overline{l_i} p(T)} \tag{22}$$

where $\boldsymbol{P_i(T)}$, $\overline{l_i}$, and $\boldsymbol{p(T)}$ are rainfall quantiles at grid i, mean rainfall at grid i, and regional dimensionless quantile for a given return period (T—years).

3.10. The IDW Method

The interpolation technique of inverse distance weighting (IDW) was employed to map the pairwise CHIRPS (ver.2)–gauge station continuous qualitative metrics across the country. This technique posits that the influence and weight of each measured point diminish with increasing distance between the unsampled points and the rain gauge. The proximity of an observation to the estimated point correlates with its influence, quantified by a weight ($\mathbf{W_o}$), defined as follows [50]:

$$W_O(x) = d(x_u, x_i)^s \tag{23}$$

where $\mathbf{x_u}$ denotes the place at which rainfall data are unavailable, x_i represents a point with available rainfall data, d signifies the distance between the two points, and s serves as an exponent that facilitates various configurations of the weighting function. The greater the value of s, the less the significance attributed to more distant observations.

4. Results

4.1. Validation of CHIRPS

Figure 4 presents a scatter diagram that illustrates the correlation between CHIRPS precipitation estimates and gauge station data for each month spanning the period from 1981 to 2016. The chart indicates that CHIRPS underestimates observed extreme precipitation at higher values, suggesting its limitations for flood monitoring within the country. Under conditions of moderate precipitation, bias values exhibit minimal deviation. Across all stations analyzed, the data from gauge stations and CHIRPS exhibited an extremely strong correlation (correlation coefficient = 0.97, bias = −1.57 mm, mean absolute error = 0.64 mm, and Kling–Gupta efficiency = 0.94).

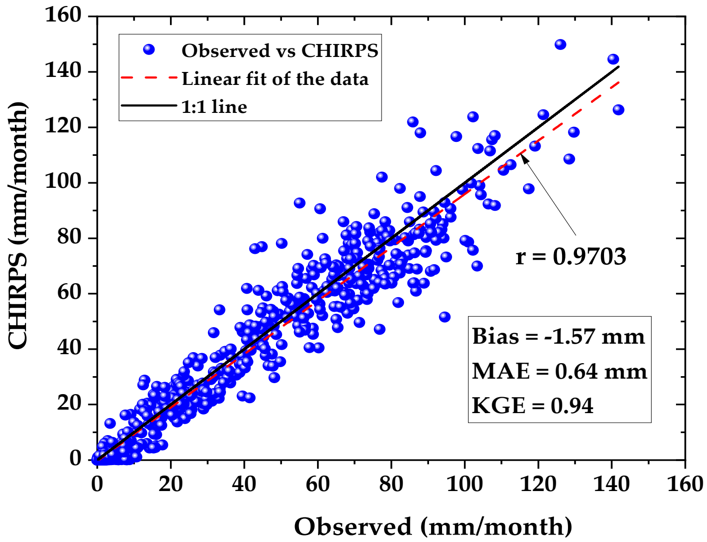


Figure 4. Monthly CHIRPS (ver. 2) precipitation estimates against gauge station data.

Slight positive and negative biases are noted in the reliability evaluation of the dataset as a substitute for precipitation gauge data. Figure 5a illustrates minor negative bias values (−0.90 to −2.10 mm/month) in the arid semi-dry region of the country. The bias values result from occasional, light, and localized precipitation, which is a challenge for satellites to detect. The illustration additionally reveals significant negative biases in the Southern Kalahari, Makgakgadi Pans, and Southeastern Botswana. The bias values are greater than −2.10 mm/month. Bhattachan et al. [51] note that the Makgadikgadi pans and southern Kalahari are significant contributors to atmospheric dust in the southern Hemisphere. The dust, linked to frequent droughts, caused plant loss in central Botswana and dune remobilization in the southern Kalahari. Atmospheric dust leads to misinterpretation of cloud-top temperature signals, resulting in negative biases in the measurement of convective precipitation amounts.

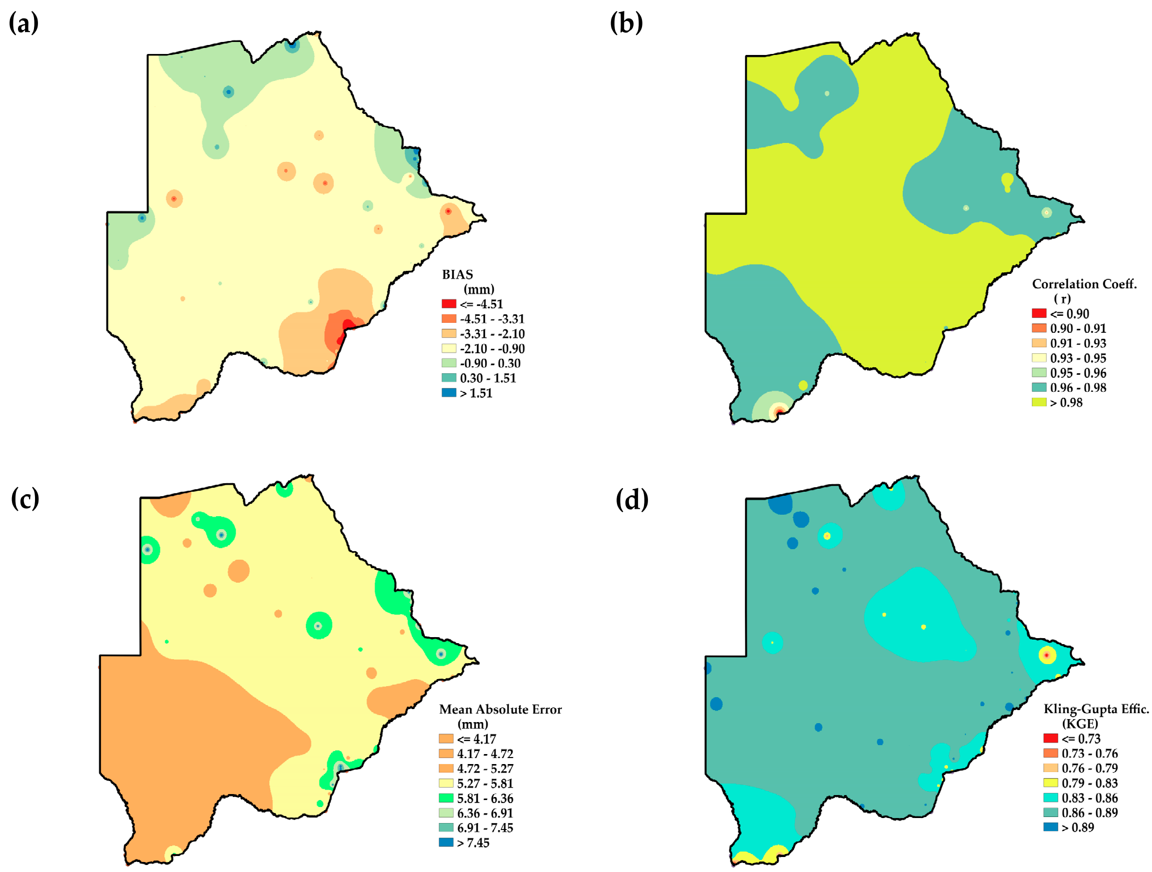


Figure 5. Spatial distribution for station-averaged metrics: (**a**) *bias*, (**b**) *correlation coefficient*, (**c**) *mean absolute error*, and (**d**) *Kling–Gupta efficiency*.

Northern Botswana exhibits slight positive biases (Figure 5a). The region may be categorized as a humid semi-arid zone. During late summer (JFM), the region experiences widespread and extensive rainfall. The intrusion of the Congo Air Mass Boundary (CAB) associated with tropical systems results in extensive heavy rainfall across the region. The tropical system generates widespread cloud cover in the region, which consistently results in precipitation. When misinterpreted by satellite signals, this attribute results in a positive bias. The positive bias values reach up to 5.02 mm/month.

Figure 5b demonstrates an excellent correlation within the semi-arid zone of the study area, exceeding 0.98. The country possesses substantial gauge station coverage in the region, and rainfall is broadly distributed, particularly in eastern Botswana. The coverage enhances CHIRPS' integration of ground station and satellite data, thereby resulting in the observed high correlation values. However, the correlation coefficient is lower (0.96–0.98) in arid and drought-prone regions, which are characterized by vegetation erosion. The eroded areas frequently experience dust storms during the arid seasons, which obscure cloud signals from satellite detection, and dust layers may also be mistaken for high, cold clouds. High surface reflectivity in the eroded areas disrupts the infrared signals transmitted from the satellite. The figure indicates that the correlation coefficient is lower (<0.98) in the wetland zone of the study area. The wetlands exhibit significant vegetation density, complicating the detection of rainfall by infrared photography and resulting in a diminished correlation.

The typical absolute inaccuracy of CHIRPS in reproducing station data is illustrated in Figure 5c. The chart illustrates that the mean absolute error in predicting station data in southeastern, northeastern, and northwestern Botswana is considerable (>5.80 mm/month). These regions have an annual rise in urbanization [52]. The elevated rate of urbanization has resulted in alterations in land use and an increase in land surface reflectance. The satellite algorithm inadequately represents the rise in surface albedo, which impacts infrared satellite images. Reduced mean absolute errors are noted in regions with less vegetation coverage along the northwest–southeast axis. Although gauge station coverage is limited, the negative and positive biases balance each other, resulting in low mean absolute errors.

Figure 5d illustrates low KGE values (<0.86) in the vegetation-eroded regions of the Kalahari, Makgadikgadi pans, and *Bobirwa* region, as well as in urbanized southeastern Botswana. There is a strong spatial concordance between CHIRPS and station data across the whole country. The trend in the figure indicates stronger values in the semi-arid zone of the research area, reaching up to 0.97.

Figure 6 illustrates a radar chart showing station-averaged metrics for each month. The figure further indicates that, across most (10 of 12) of the months, CHIRPS underestimated the data acquired from gauge stations. The underestimation ranges from 0.24 to 5.56 mm/month. The most significant underestimation is observed during the months experiencing high precipitation, specifically January and February. From March to September, the dry season, both negative and positive bias values are minimal. The bias values exhibit a range from −0.24 to −0.64 mm. The present observation corroborates the results obtained by Alsilibe et al. [53].

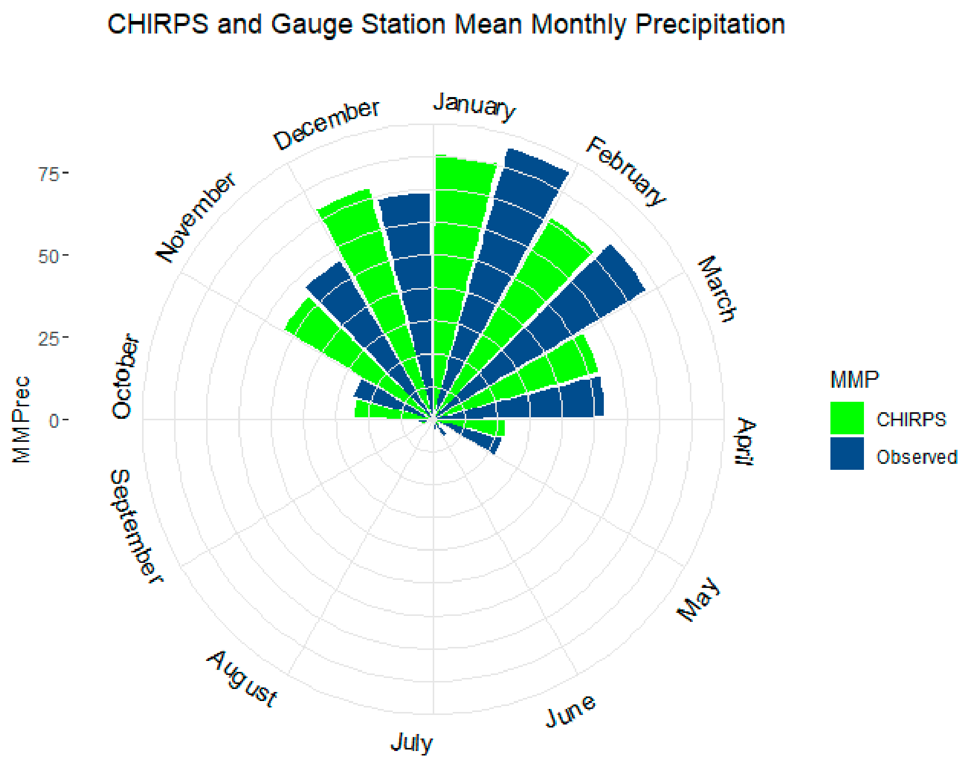


Figure 6. Radar chart showing region-averaged annual variation in observed gauge station data against CHIRPS estimates.

The intra-annual variation in the correlation coefficient is depicted in Figure 7a. The figure illustrates that the metric rises from 0.80 at the beginning of the rainy season (September) to 0.91 at its peak. Subsequently, the coefficient declines to a value of 0.54 upon cessation in March. During the dry or winter season, low coefficient correlation values are observed, with the metric reaching as low as 0.10.

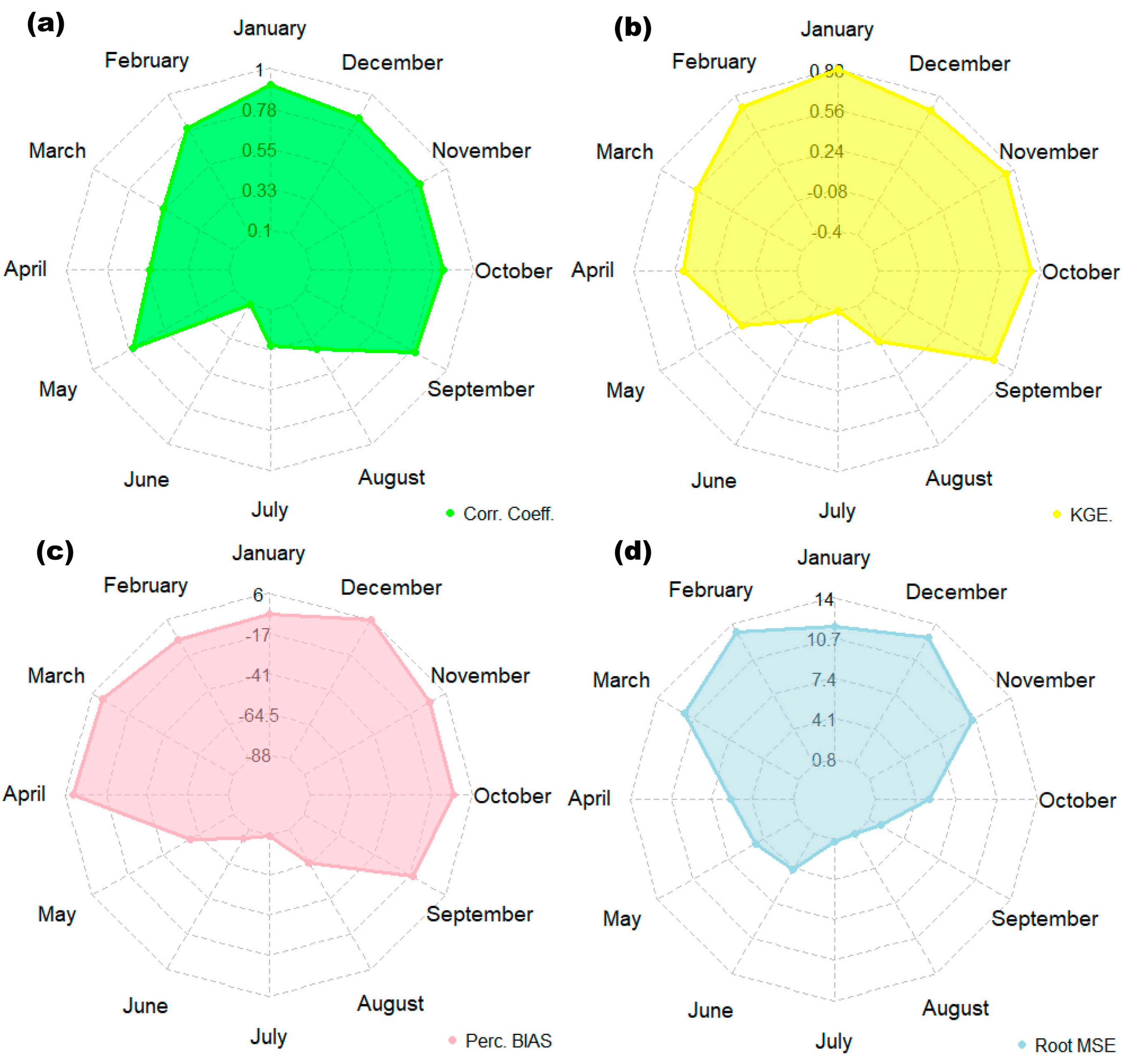


Figure 7. Radar plots showing station-averaged annual variation in validation metrics: (**a**) *cc*, (**b**) *KGE*, (**c**) *perc. bias*, and (**d**) *RMSE*.

Similarly, KGE exhibits elevated values during the rainy season (refer to Figure 7b). The values span from 0.70 to 0.88, with January exhibiting the highest value. The figure further indicates that KGE values are negative during the winter season. This fact is attributable to the minimal precipitation observed throughout the season. However, the values exceed −0.41, which is the established critical mean flow benchmark value [54]. The negative KGE values do not indicate subpar performance for CHIRPS.

Figure 7c depicts extremely high negative percentage bias values, indicative of the dry season (May to September). The values span from −15.61 to −86.97%/month. This variation is attributable to the minimal amount of precipitation that takes place throughout the season. The figure also demonstrates moderate negative percentage bias values that coincide with the rainy season. The range extends from −0.45 to −7.02%/month. The figure additionally indicates positive percentage bias values for December and April.

Figure 7d displays the RMSE's temporal variation. During the wet season (November to February), the measure shows higher values. The range of RMSE values is 10.41 to 13.29 mm/month. Nonetheless, it is observed that the dry season (April to October) has lower metric values. The range of values is 0.84 to 5.91 mm/month.

CHIRPS v2.0 demonstrates a very high ability to delineate the spatial patterns of mean monthly precipitation. It has demonstrated a robust correlation (r > 0.97) with the observed data at typical monthly intervals, accompanied by low MAE values of 0.64 mm/month. It can thus serve as a proxy for gauge station data in the study.

4.2. Principal Components and Spatial Distribution of Scores

A normalized gridded CHIRPS precipitation data matrix was created for the S-mode PCA. The Kaiser–Meyer–Olkin (KMO) statistic, an indicator of sample PCA adequacy, was applied to the data matrix. The KMO test of the variables yielded an overall value of 0.84, which demonstrates the matrix's appropriateness for PCA. The variables in the analysis exhibited a commonality index exceeding 0.7.

A correlation matrix was then generated, from which the eigenvectors and eigenvalues were extracted. Employing both the cumulative variance proportion and the scree plot, we identified *three* principal components for use as input to the k-means clustering process. The retained principal components (PCs) contributed 51.3%, 28.6%, and 7.9% of the total variance in the first, second, and third instances, respectively (see Figure 8). The three principal components represented 87.8% of the total percentage variance.

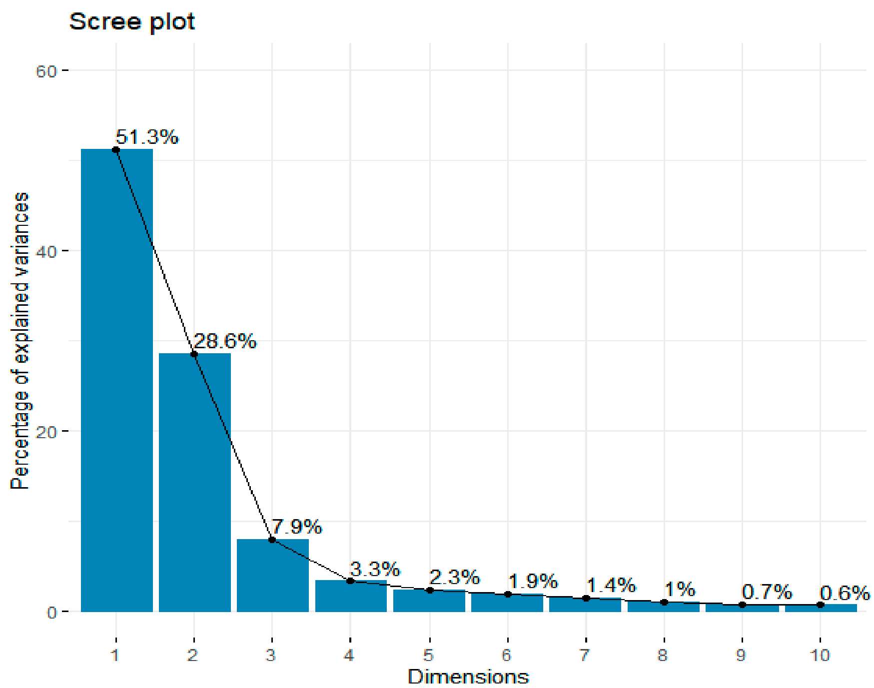


Figure 8. Scree plot of eigenvalues for the first 10 PCs.

Understanding the PCA results necessitates an examination of the PC loading. When the loadings are high, the variables (months) and PCs exhibit a strong correlation with each other. Following the *Varimax* rotation, the correlations between months and PCs are presented in Figure 9. The principal component loadings of the monthly mean precipitation totals mirrored the three fundamental intra-annual fluctuations in precipitation throughout Botswana.

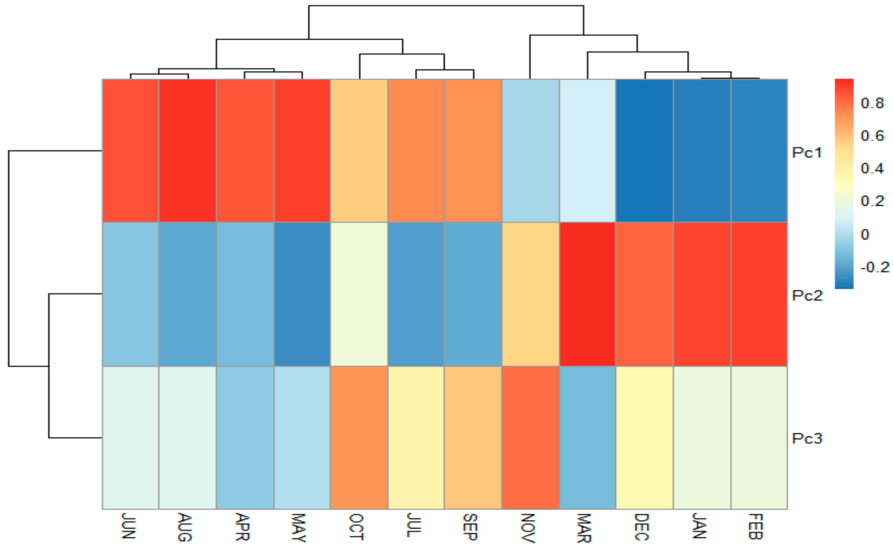


Figure 9. Correlations between variables (months) and PCs.

PC_2 had a high loading from December to March and constituted 28.6% of the overall variation. The northern parts of Botswana receive the most summer rainfall, with the remaining parts receiving less rainfall (Figure 10b). In central Angola's Bié Plateau, a low-pressure system referred to as the Angola low (AL) emerges in early summer [54]. AL serves as the primary tropical origin of the cloud bands that deliver most of the summer precipitation to subtropical southern Africa, with up to 60% occurring in the summer [55].

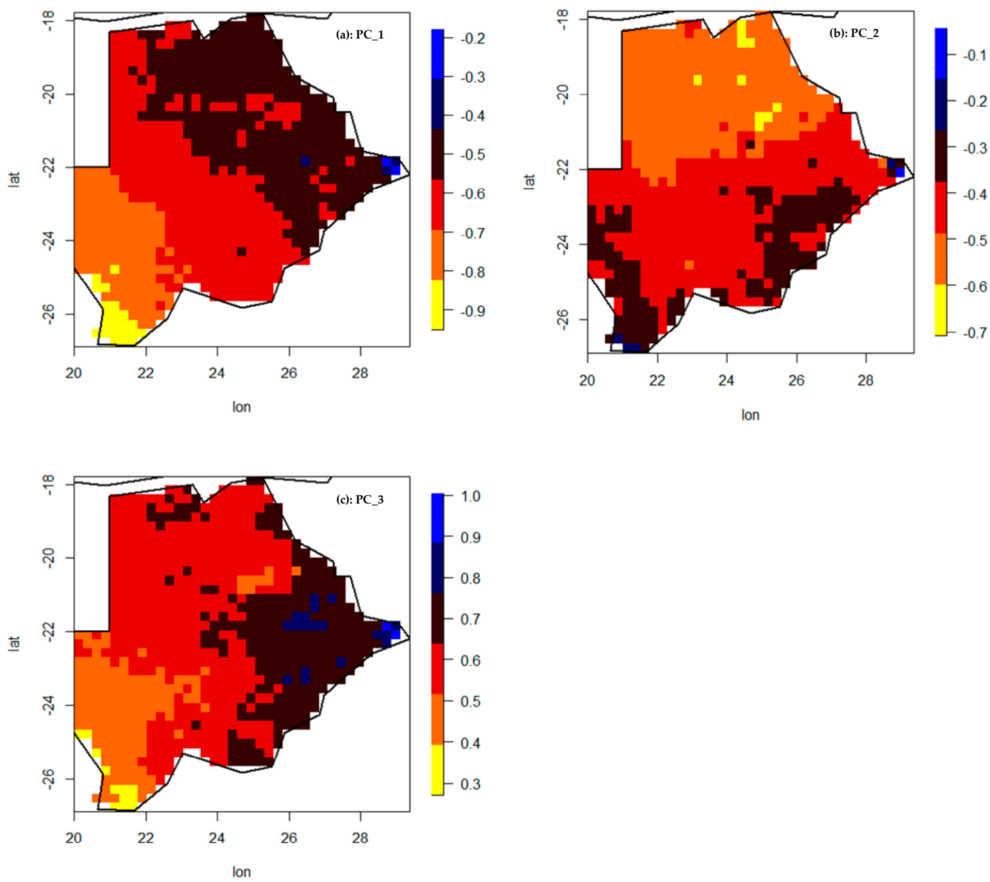


Figure 10. PCA rainfall patterns for (**a**) PC_1, (**b**) PC_2, and (**c**) PC_3 scores.

- **PC_1:** The loadings are high in autumn (*cessation*) and winter months, which are characterized by little or no rainfall (*April* to *September*).
- **PC_2:** Loadings are high in late summer precipitation months (*December* to *March*).
- **PC_3:** Loading is high in rain *onset* months, which are characterized by low precipitation amounts (*October* and *November*).

The PC_1 scores depict the period between the end of the rainy season and the start of early showers in the spring season, thus the transition period. During seasonal transitions, the westerly trough dominates Botswana's southern region (Figure 10a). This phenomenon generates a significant amount of rainfall in the sub-region. On the other hand, the rest of the country experiences strong subsidence and little rainfall due to the dominant continental subtropical high-pressure system.

A more robust eastward migration of moisture from the tropical Southeast Atlantic is linked to an exceptionally powerful atmospheric low-pressure system. This phenomenon causes the moisture flow to significantly converge at lower elevations throughout the season in the northern parts of the country.

PC_3 had the highest loading in both October and November, contributing to 7.9% of the variance. The eastern, southeastern, and northeastern areas of Botswana exhibited elevated PC_3 scores, as shown in Figure 10c. A clear east–west gradient in the score values was observed due to the synoptic effects. This may have occurred because waves in the easterly airflow carrying moisture from the warm currents on the east coast caused the surface to converge. Consequently, an augmentation in rainfall is observed in the eastern region of Botswana.

4.3. Optimum Number of Clusters

The average silhouette width was calculated for several clusters ranging from two to ten using K-means as the clustering algorithm. The average width slightly increased from a value of 0.68 when two clusters were assumed to a maximum of 0.69 when the number of clusters was four. Thereafter, the width decreased gradually to a value of 0.32 when the number of clusters was eight (see Figure 11). The lowest silhouette width was calculated to be 0.16 when the number of clusters was presumed to be 10. Consequently, the number of clusters that yielded the best regionalization was chosen to be *four*. This is consistent with an earlier observation by Nkoni et al. [6].

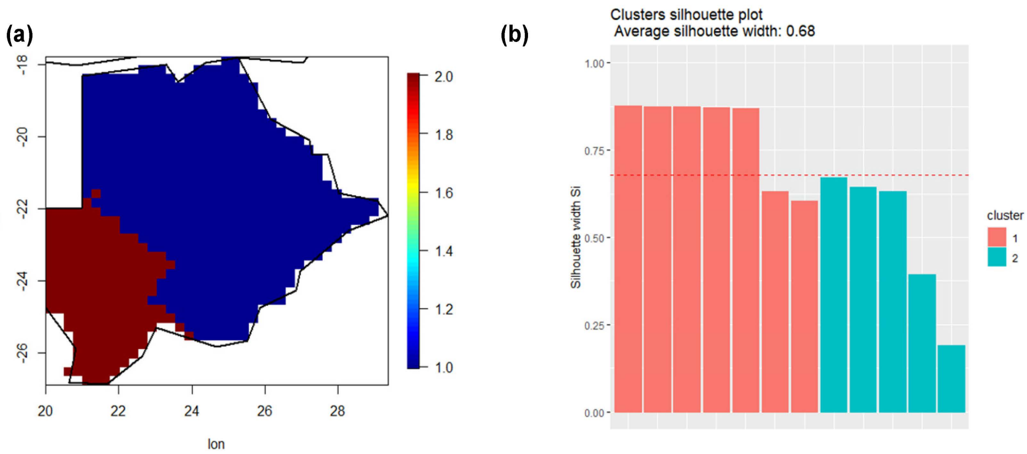


Figure 11. *Cont.*

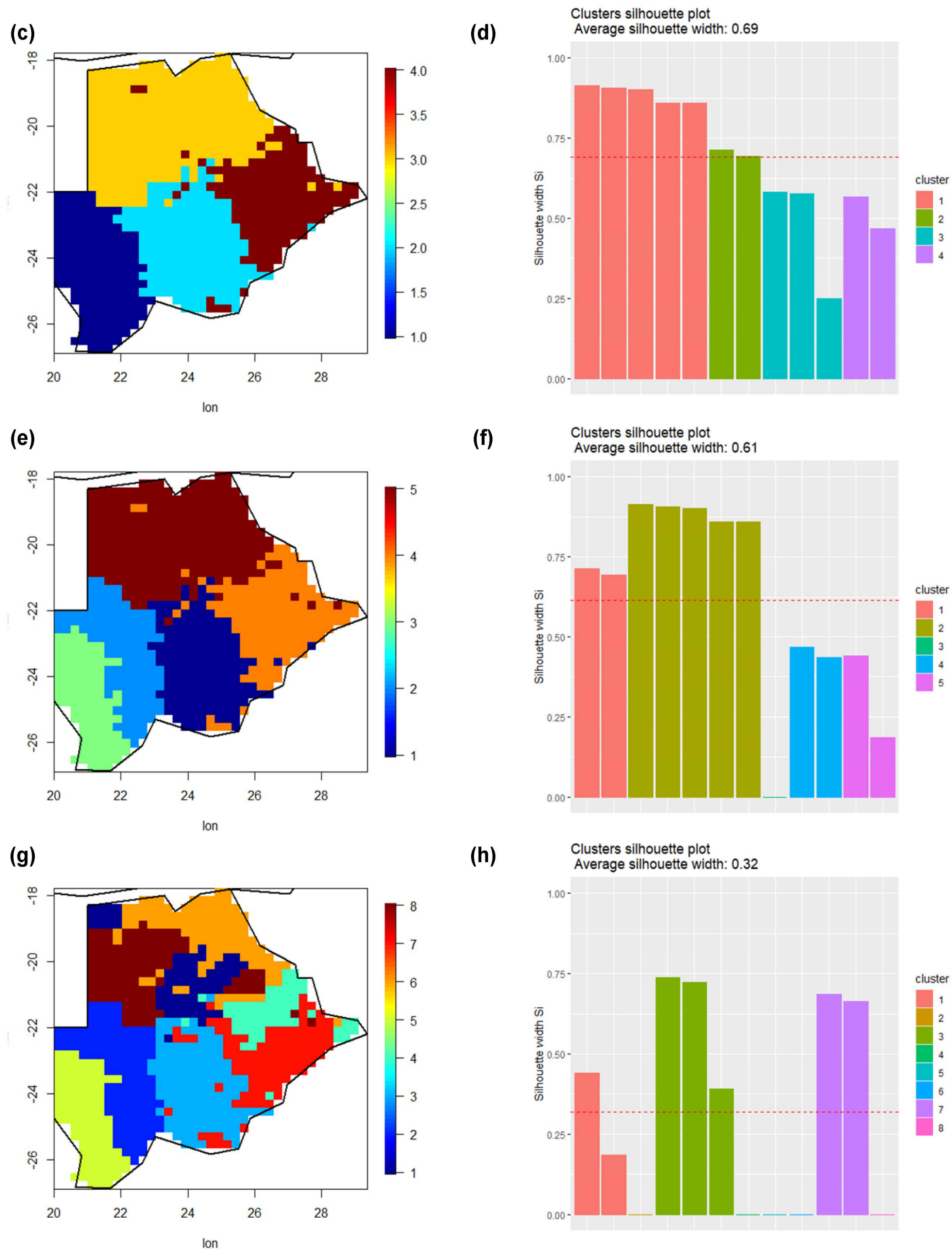


Figure 11. Cluster partitions and average silhouette width plots for (**a**,**b**) two clusters; (**c**,**d**) four clusters; (**e**,**f**) five clusters; and (**g**,**h**) eight clusters (The red dashed line denotes the average silhouette width).

4.4. K-Means Clustering of the PC Scores

We utilized a non-hierarchical cluster analysis with the three principal component scores as inputs to delineate the homogenous precipitation zones. Classification was accomplished with the k-means clustering technique.

The K-means method was chosen because, in contrast to hierarchical clustering techniques, it permits the reassignment of items that may have been inaccurately categorized initially [56]. According to the silhouette width index, the country was categorized into four (4) sub-regions as illustrated in Figure 12.

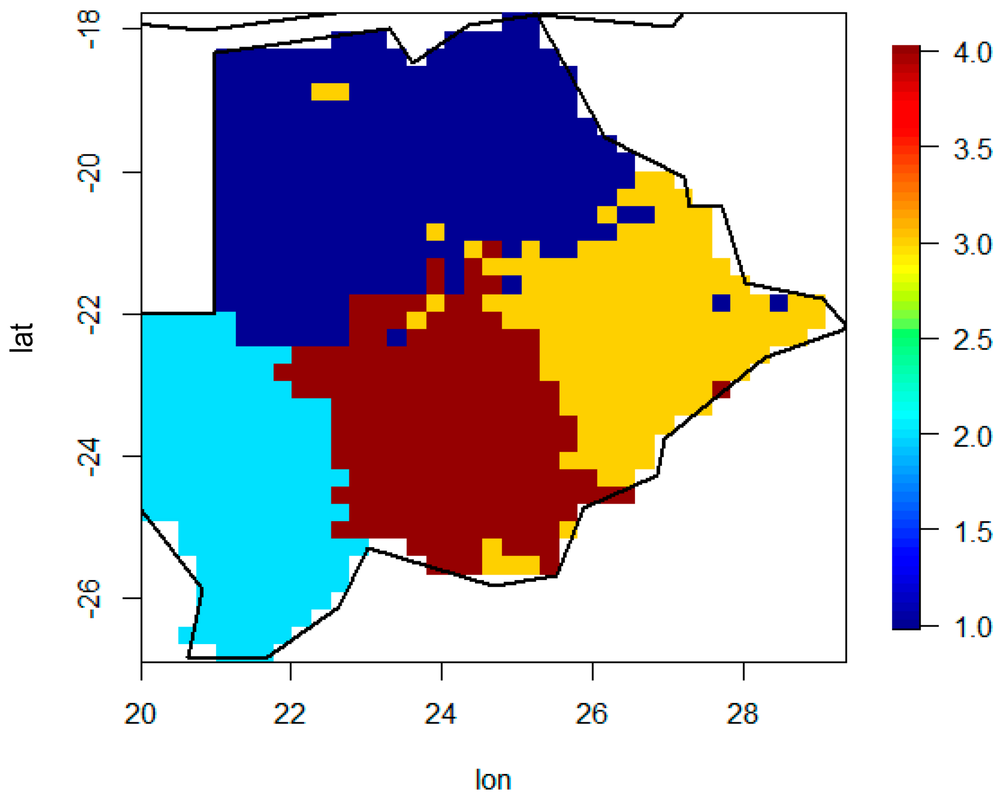


Figure 12. Spatial distribution of sub-regions for precipitation data using k-means.

4.5. Discordancy Measure

The discordancy index (Di) evaluates the L-moment vector of a grid point in relation to the regional average. A high Di value signifies that the grid point is discordant. Multiple discordant grid points from the CHIRPS (ver. 2.0) dataset were identified using the discordancy statistic [49]. Specific grid points displayed potential data anomalies and were incongruent with other places within the designated sub-regions. The following points exhibit a discordancy statistic (Di) over 3.0. The counts of discordant grid points are 51, 68, 41, and 43 for the SW-B, NT-B, SE-B, and EAS-B clusters, respectively.

The grid points were examined for climatic abnormalities resulting from a distinctive microclimate. Nevertheless, we were unable to elucidate the reason for the discordance. As a result, the points could not be preserved and were eliminated from regional pooling to ensure uniformity. Following the elimination of the outlier data points, regional statistics were recalibrated to verify the enhanced sub-regions.

4.6. Annual Rainfall L-Moments

Table 3 presents the values of LCv, L-skewness, and L-kurtosis ratios for the homogenous regions. L-Cv quantifies the relative variability of precipitation in relation to the regional average. Table 3 indicates that NT-B has the lowest L-Cv (0.137), signifying that its rainfall stations have the most internal consistency in variability. This is unsurprising for a more humid, climatically consistent northern region where moisture from the Inter-tropical Convergence Zone prevails more consistently. Table 3 indicates that EAS-B (0.179) and SE-B (0.176) are closely aligned and exhibit modest variability, indicating transitional semi-arid environments. SW-B (0.165) exhibits authentic geographic consistency in aridity rather than varied precipitation patterns. The all-Botswana L-Cv of 0.209 is the highest, indicating that aggregating climatically diverse areas exaggerates perceived variability, hence validating the sub-regional categorization.

Table 3. L-moments of the homogeneous regions determined from the study.

Region Name:	No. of Grids	MAP (mm)	L-Cv	L-Skewness	L-Kurtosis	H_1	$H_{1\ subs}$
SW-B	161	269.49	0.165	0.172	0.192	−7.277 *	**−0.61**
SE-B	184	315.10	0.176	0.200	0.152	−5.420 *	**−0.84**
NT-B	293	520.44	0.137	0.041	0.108	−1.912	
EAS-B	161	396.28	0.179	0.183	0.110	1.526	
Botswana	799	401.44	0.209	0.117	0.089	5.874	

"*" shows super homogeneity values, bolded values are H1 values after subsampling.

SE-B exhibits the highest L-skewness (0.200), indicating the most right-skewed distribution. This is consistent with the unpredictable, episodic rainfall observed in the southern borders. SW-B followed closely (0.172), suggesting that arid conditions are present with intermittent high-magnitude eruptions. EAS-B (0.183) is equally biased, suggesting that the eastern corridor experiences analogous rainfall abnormalities despite having a larger MAP than SW-B. NT-B is significantly distinct, with a measurement of 0.041 and an almost symmetrical shape. This is the most apparent skewness result in the table. It suggests that the rainfall in northern Botswana, which is heavier and more frequently produced by seasonal tropical systems, does not possess the same level of extreme event dominance as that in the south. The average L-skewness of Botswana is 0.117 (Table 3).

The rainfall extremes of SW-B have the most distributional weight, as evidenced by its highest L-kurtosis (0.192). Based on the L-kurtosis value, a distribution with denser tails, such as the generalized extreme value or generalized logistic, is more likely to be suitable in this region than a distribution with a lighter tail. The subsequent element is SE-B (0.152), which is also elevated (Table 3). The EAS-B and NT-B are nearly identical (0.110 and 0.108, respectively) and significantly lower, indicating a more moderate tail behavior and potentially a better fit with distributions such as the Pearson Type III or generalized normal distribution. The distributional signal is flattened by aggregation, as evidenced by the lowest all-Botswana L-kurtosis (0.089).

The negative H1 values of three out of the four sub-regions (SW-B, SE-B, NT-B) indicate a greater homogeneity than expected according to the simulation (Table 3). Only EAS-B exhibits moderate discordance (H1 = 1.53, still within the usual threshold of two). Aggregating all data into single region, the all-Botswana data have an H1 of 5.87, which markedly exceeds two. The extreme instances are SW-B (−7.28) and SE-B (−5.42). NT-B is negative (−1.91). A negative H1 signifies that the observed dispersion in at-site L-Cv across the area is much lower than the predictions made by kappa-distribution simulations for a homogeneous region of equivalent size and record duration. The grids in these sub-regions exhibit more similarity to each other than would be expected by random chance, even in a completely homogenous area. If the rainfall data from gauging stations within a sub-region are correlated and physically proximate, then high negative values of H1 are obtained. This may explain the increased negative H1 values found at SW-B and SE-B. Subsampling was applied to specifically probe the spatial correlation problem.

4.7. Spatial Correlation Correction

To alleviate or eliminate the spatial correlation between grid points impacting the homogeneity test, the grid points from both SWB and SEB were intentionally reduced and spatially allocated inside a sub-region. The objective is for H1 to be less than one (adequately homogeneous) after subsampling. Values ranging from one to two, however small, may be warranted. Substantially negative numbers, such as −7.28 and −5.42, need to transition into the range of 0 to +1 with the removal of spatial correlation, maintaining a uniform distribution while eliminating aberrant negative characteristics. Significantly negative

H1 values that fall far below the lower threshold suggest that both sub-regions start in a state of considerable homogeneity: The sub-regions converge inside the [−1, +1] interval. Figure 13a shows the convergence of both sub-regions to values near zero, which satisfies the Hosking & Wallis [49] criteria for acceptable homogeneity (∣H1∣ < 1), and verifies that the subsampling approach has effectively removed super-homogeneity.

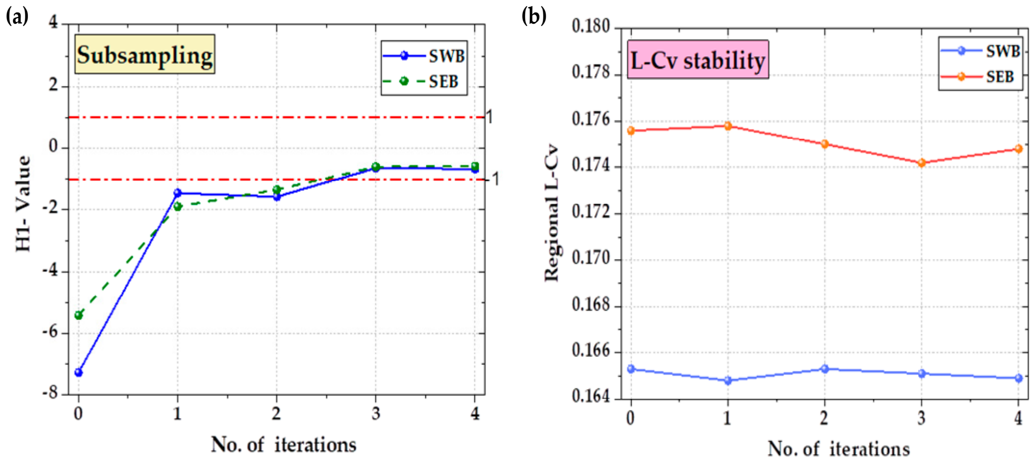


Figure 13. Grid point subsampling for SWB and SEB: (**a**) H1 convergence and (**b**) regional L-Cv stability.

From Figure 13b, the L-Cv exhibits little variation over iterations, ranging from 0.165 to 0.173 for SW-B, and from 0.176 to 0.183 for SE-B. This is the essential diagnosis. The regional L-moment statistics are steady and resilient; the only variable being altered is H1, and this alteration is solely attributable to the elimination of inter-site correlation, rather than any genuine modification in the fundamental frequency characteristics. The regional data is reliable. The permissible H1 at the conclusion must fulfill H1 < 1. Both SW-B (−0.61) and SE-B (−0.84) firmly occupy that range by iteration 4.

Spatial subsampling has been widely applied in regional frequency analysis to avoid correlation among sites due to their common climatic conditions [57,58]. Spatial correlation leads to super-homogeneity of data (H < 0), making the tests for homogeneity, such as the Hosking & Wallis H test, invalid. It may introduce bias into the parameters estimated using the L-moments. With subsampling, only a few stations out of the whole set are used, and thus these problems are circumvented.

However, one major drawback associated with using subsampling as opposed to the benefits accrued from making each observation independent is data utilization. Use of subsampling automatically entails that some amount of data is not considered during the analysis process (Politis et al. [59]). This leads to the fact that the data used are inefficient in determining certain measurements, including L-skewness and L-kurtosis, since they are hard to estimate with a low sample size. Extreme quantiles also face data inefficiency due to their dependency on tails in distributions, which in turn need a greater sample size. Apart from this impact, it also becomes impossible to detect any relevant pattern that must be considered while building a model of the regional process. Moreover, it is important to note that subsampling does not consider any possible spatial dependence between the stations and might not even solve problems of super-homogeneity.

Several approaches can be used to address the problems associated with spatial subsampling. The use of bootstrap sampling and bootstrapping techniques allows one to calculate the variation in the estimate under different subsampling schemes without altering the initial information [60]. In addition, the utilization of regionalization and stratified spatial sampling results in a fairer selection of observation sites. If none of the above-mentioned techniques proves helpful, a different approach must be used as a response to the lack of efficiency in spatial subsampling. The use of copulas and other spatial statistics techniques allows accounting for the correlations between different sites while retaining all data [56]; the application of hierarchical Bayesian models provides error transferability across scale while using all data provided; finally, the use of geostatistics permits assigning appropriate weights to observations made at different stations according to their importance. All the approaches are especially important in cases of low amounts of data, large geographical spread, and reliable quantile estimation of rare events, respectively.

4.8. Cluster Hydro-Meteorological and Morphometric Characteristics

4.8.1. Cluster Precipitation Characteristics

Descriptive statistics, including the mean monthly precipitation and coefficient of variation, were computed for each of the four clusters to illustrate the variability in precipitation throughout the year. In Figure 14, these statistics are illustrated.

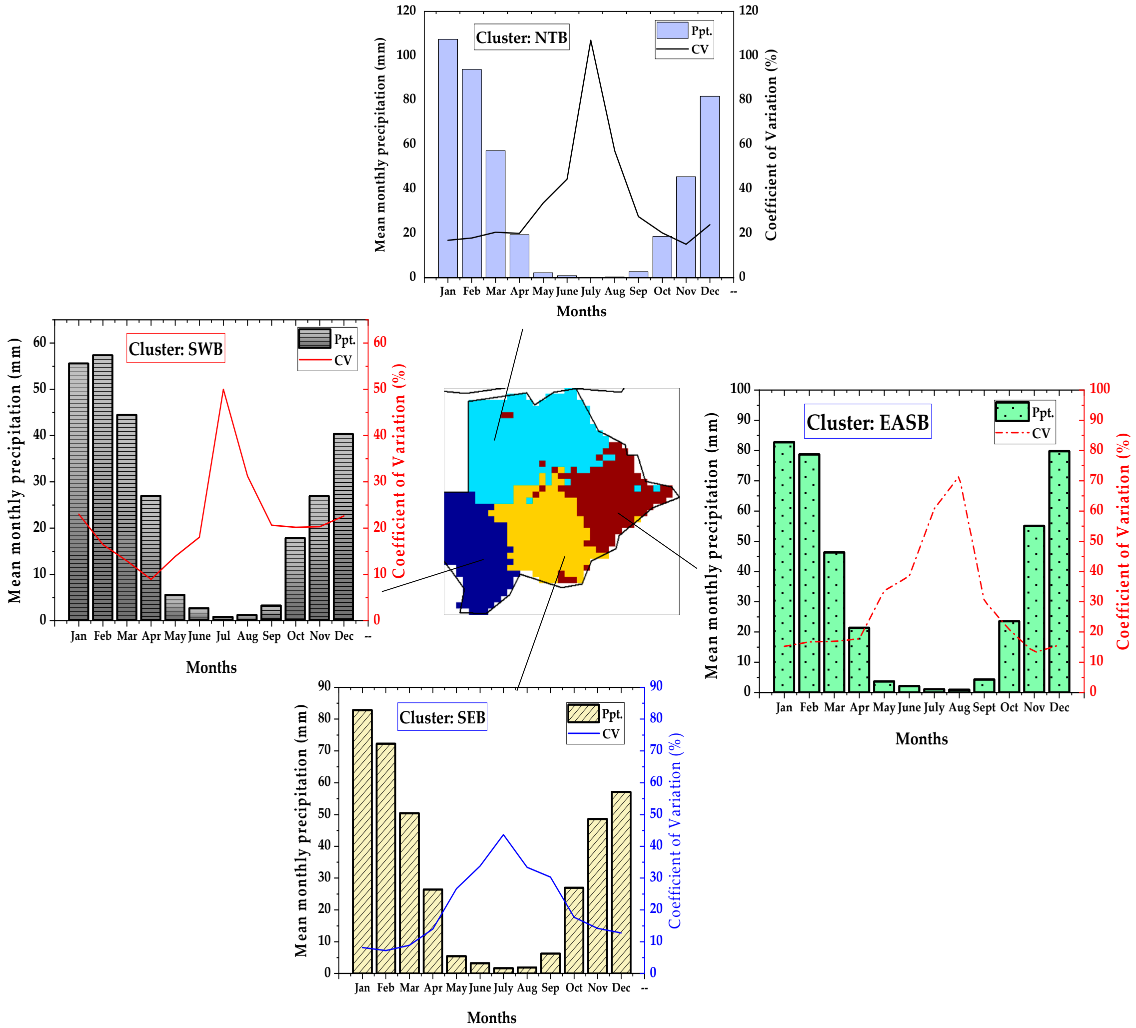


Figure 14. Rainfall characteristics the homogeneous sub-regions.

Botswana's southwestern sub-region, *Cluster 1*, comprises one hundred and one (161) grid points. The average annual precipitation in this sub-region is 242 mm, with an average coefficient of variation of 22%. A considerable proportion of depressions leads to substantial winter precipitation (about 12% of the yearly total), despite the monthly peak occurring in February, significantly influenced by the latitudinal migration of the Inter-tropical Convergence Zone (ITCZ) during summer.

Cluster 1's monthly mean rainfall varies greatly, ranging from 0.67 mm in July to 57 mm in February (Figure 14). Rainfall begins in August and averages 1.1 mm. Meanwhile, the correlation of variation's trend increased steadily from 13% in April to its optimum value of 50% in July and then gradually decreased to 20% in September. The maximal coefficient of variation in July may be attributed to the frequent occurrence of localized thunderstorms and erratic low rainfall in the sub-region.

In Botswana's northern region, *Cluster 2* comprises two hundred and ninety-three (293) grid points. The mean annual precipitation in this cluster is the highest at 430 mm. The monthly precipitation estimates are higher during the DJM months in comparison to those in Cluster 1 (Figure 14). The average precipitation during the months of July and August was less than 1 mm. In January, the monthly mean precipitation reached its highest point, which was 107 mm.

Figure 14 illustrates the rainfall pattern for *Cluster 2*. The monthly mean rainfall quantity for the cluster is depicted in the figure, with a fluctuation from a maximum of 107 mm in January to a minimum of 0.1 mm in August. Rainfall in the sub-region commenced in September, averaging 3 mm. The correlation of variation increased consistently from 20% in March to a peak of 107% in July, thereafter, declining to 15% in September (see Figure 14). The maximum value of the coefficient of variation results from the exceptionally low precipitation in the sub-region.

Cluster 3 is composed of one hundred and eighty-four (184) grid elements. It is in the southeastern region of Botswana. The average annual precipitation in this sub-region is 383 mm, with an average coefficient of variation of 21% (see Figure 14). The sub-region is subject to the significant influence of mid-latitude westerly frontal activities, which contribute to approximately 9% of the annual total, like Cluster 1. January is the month with the highest monthly precipitation, which is approximately 83 mm.

The mean rainfall for Cluster 3 is subject to variation, with a maximum of 83 mm in January and a minimum of 1.7 mm in July (Figure 14). The onset of rainfall is in August, with an average of 1.9 mm. In January, the coefficient of variation is 8%, while it reaches a maximum of 44% in July. In December, the utmost value is also observed to decrease to 13%.

Cluster 4 has the same number of grid points as Cluster 1, which is one hundred sixty-one (161). The sub-region encompasses the eastern portion of Botswana. Figure 14 illustrates the precipitation pattern in *Cluster 4*. The average monthly precipitation in the cluster ranges from a maximum of 83 mm in January to a minimum of 0.9 mm in August. The average rainfall in December is 80 mm. The early summer (OND) season is the time when over one-third of the annual rainfall occurs. Like this, the coefficient of variation reaches its maximum value of 72% in August.

4.8.2. Cluster Topographical and Meteorological Characteristics

According to Table 4, Tmin values vary by 3.2 °C, from 15.0 °C in SE-B to 18.2 °C in NT-B. The temperature range is as follows: SE-B < SW-B < EAS-B < NT-B. Since lower air temperatures are anticipated at greater altitudes, the sub-region SE-B has the lowest Tmin, which corresponds to its maximum altitude of 1109 m. Because NT-B is at the lowest height (978 m), it exhibits a typical lapse-rate association and has the greatest Tmin. The altitude and Tmin ratings of SW-B and EAS-B are intermediate.

Table 4 shows that Tmax values follow a similar pattern to those of Tmin. SE-B has the lowest temperature at 29.4 °C, while NT-B leads once again at 31.4 °C. The variation across sub-regions is 2.0 °C, which is somewhat lower than that of Tmin, indicating that elevation-induced disparities are lessened by midday heating. NT-B has the biggest diurnal temperature range (Tmax-Tmin) at 13.2 °C, whereas SE-B has the smallest (14.4 °C).

Compared to SW-B (13.8 °C) and EAS-B (13.4 °C), SE-B has a greater temperature range (14.4 °C). Compared to higher-altitude locales, NT-B has the smallest diurnal variation (13.2 °C), suggesting that its low elevation results in a warmer surface and a more constrained daily change.

Table 4. Hydro-meteorological and morphometric attributes of the homogeneous sub-regions.

Sub-Region Name	No. of Grids	T_{min} (°C)	T_{max} (°C)	Altitude (m)
SW-B	161	16.0	29.8	1081
SE-B	184	15.0	29.4	1109
NT-B	293	18.2	31.4	978
EAS-B	161	17.3	30.7	999

The altitude varies from 978 m (NT-B) to 1109 m (SE-B), for a variation of 131 m. Table 4 shows the hierarchy of NT-B, EAS-B, SW-B, and SE-B. The 131 m difference, although not dramatic, is climatologically substantial and directly supports the temperature gradient. The inverse relationship between altitude and temperature is clear: SE-B (highest) is the coldest, while NT-B (lowest) is the warmest.

The data in Table 4 show a consistent negative association between altitude and temperature throughout sub-regions. NT-B, the most widespread and lowest sub-region, consistently reports the highest temperatures. SE-B, the highest sub-region, has the coldest temperatures. SW-B and EAS-B are similar in all respects. This geographical pattern is physically coherent, indicating that topographic elevation has a significant impact on temperature variance throughout various sub-regions.

4.8.3. Sub-Regional Minimum Temperature Variability

Figure 15 of Ms_425 8547-R shows the spatial variation in mean annual minimum temperature in the four sub-regions. The figure depicts a distinct north–south gradient, with a 5 °C disparity between the NT-B and SE-B sub-regions. The disparity is substantial enough to define thermally separate regions. Tmin holds ecological and agricultural significance. It is a crucial threshold variable for frost occurrence, frost-free time duration, crop-growing-season definition, and livestock thermal stress. Moreover, minimum temperature is associated with humidity, soil-moisture conditions, and cloud cover. Figure 15 indicates that in the NT-B sub-region, Tmin varies between 17 and 19 °C. The average Tmin value for the sub-region is 18 °C. This sub-region exhibits the highest low temperatures. The elevated average Tmin value indicates the tropical impact at lower latitudes (about 22–24° S), where nocturnal radiative cooling is diminished. The closeness to the Okavango Delta and Chobe wetlands may enhance moisture retention, hence inhibiting overnight cooling through latent heat release.

Figure 15 indicates that the EAS-B sub-region exhibits Tmin values between approximately 15 and 17 °C, with a spatial average of 17 °C. The sub-region lies inside a transitional thermal zone between the temperate north and the cooler south. The area experiences intermittent moisture advection from the Indian Ocean, which helps alleviate nighttime cooling. The eastern escarpment and elevated regions along the Zimbabwe/South Africa border may facilitate mild cooling.

In the SE-B sub-region, Tmin values fluctuate between 13 and 16 °C, with an average of 16 °C. The illustration illustrates localized chilly pockets, perhaps influenced by higher elevation and topographic cooling effects. Increased distance from moisture sources results in arid air and intensified longwave radiative cooling at night, along with cold air drainage into low-lying regions during tranquil, clear nights. The sub-region has the most intra-regional variability, indicating intricate local topography or land surface interactions.

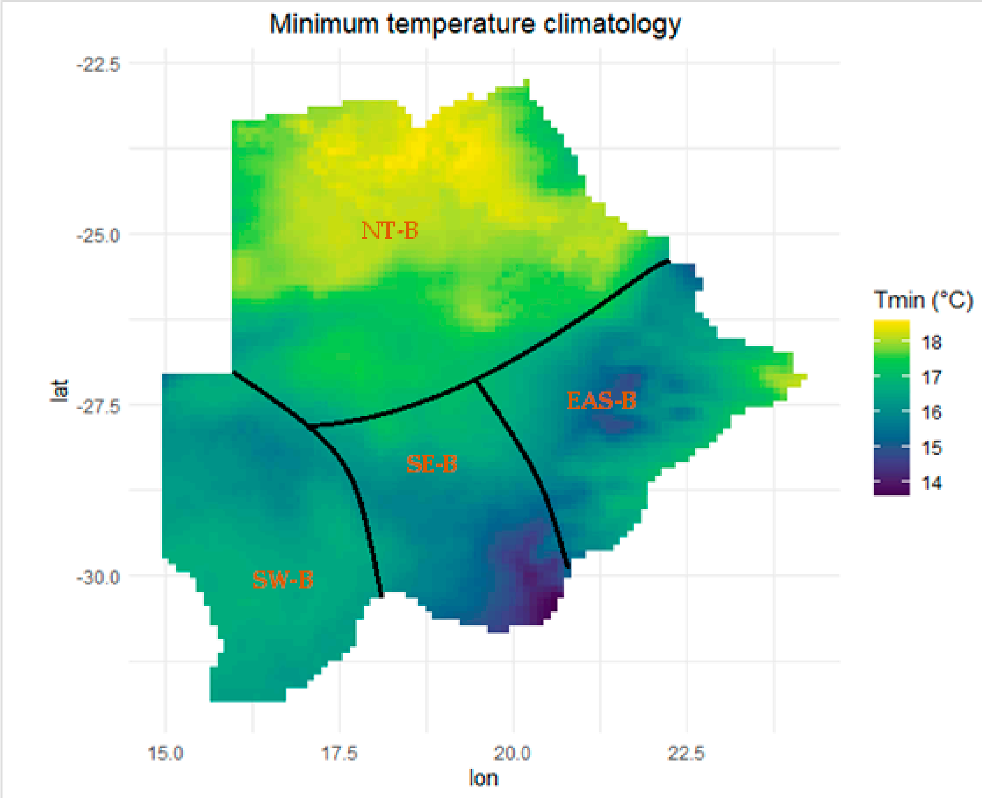


Figure 15. Spatial variation in mean annual minimum temperature in the four sub-regions.

Figure 15 further indicates that the SW-B sub-region experiences a Tmin range of 14 to 16 °C. The CHIRTS dataset indicates a spatial average temperature of 16 °C. The area is encompassed within the Kalahari Desert system, distinguished by semi-arid to desert climates. Although located at comparable latitudes to SE-B, SW-B exhibits distinct cooling dynamics: Sandy Kalahari soils possess poor heat capacity and conductivity, expediting heat release post-sunset. However, the scant vegetation and arid atmosphere facilitate significant radiative cooling. Unclouded skies prevail in this desert region, optimizing nocturnal thermal dissipation.

The minimum temperature has been employed as one of the variables for regionalization in agroclimatic zoning and crop suitability modeling, e.g., Haberle et al. [59]. The profiles in Figure 16 demonstrate statistically distinct thermal regimes: the NT-B/EAS-B cluster is significantly differentiated from the SE-B/SW-B cluster by 4–5 °C in winter low temperatures. The differences in curve shapes (seasonal change rate and trough depth) are sufficiently pronounced to justify the creation of multiple management zones.

NT-B maintains the highest low temperatures year-round. The winter trough (about 8–9 °C) is considerably warmer than SE-B and SW-B (Figure 16). The temperature curve is less prominent and more gradual, indicating a thermally buffered environment, either due to proximity to lower latitudes or coastal influences. The minimum temperature in the warm season consistently surpasses 20 °C.

Figure 16 demonstrates that EAS-B closely parallels NT-B for most of the year, although it displays a slight divergence during the warm season, registering marginally lower minimum temperatures (about 20–21 °C). The winter trough is rather mild, around 8 to 9 degrees Celsius. The curve's shape is nearly parallel to NT-B, suggesting comparable thermal conditions between these two sub-regions; however, EAS-B is somewhat cooler during summer.

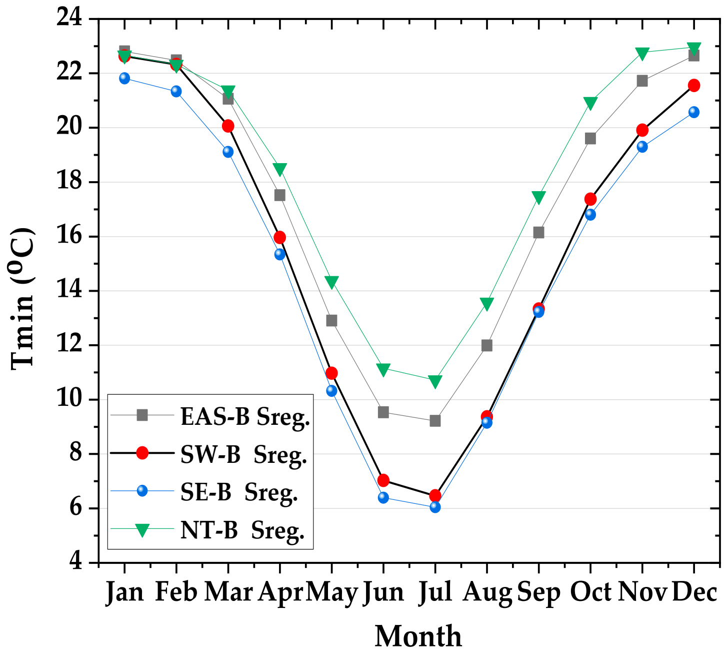


Figure 16. Intra-annual variation in mean monthly minimum temperatures in the four sub-regions.

SW-B exhibits a more pronounced, deeper winter trough, descending to around 4–5 °C in July (Figure 16) The decline from warm-season temperatures is abrupt. This indicates a more continental thermal regime characterized by increased diurnal and seasonal amplitude and diminished buffering ability. The minimum temperatures of the warm season are like those of SE-B, approximately 21 °C.

Figure 16 indicates that SE-B experiences the coldest winter minimum among the four sub-regions, attaining roughly 4–5 °C and potentially rivaling SW-B for the lowest trough. The profile of SE-B closely aligns with that of SW-B throughout the winter months; however, the curves diverge in the shoulder seasons (March–May and September–November), with SE-B generally exhibiting slightly cooler or equivalent temperatures. This sub-region exhibits the greatest thermal amplitude.

SE-B and SW-B experience a low temperature of 4–5 °C in July (Figure 16). This is perilously near or under the chilling injury threshold for tropical and subtropical crops (e.g., maize, beans, cassava, and banana—all often cultivated in southern African agrosystems). These sub-regions are incapable of sustaining warm season crops throughout the year without considerable frost and cold risk mitigation. NT-B and EAS-B, featuring winter floors of 8–9 °C, provide significantly enhanced protection for cold-sensitive crops.

4.9. Heavy-Tail Behavior of Annual Rainfall in the Sub-Regions

In semi-arid environments, precipitation is typically sporadic and convective. The yearly total is not incrementally gathered through twelve approximately equal monthly contributions; rather, it is mostly influenced by a limited number of significant rainfall events concentrated over a brief wet season or even a solitary month. The maximum monthly rainfall is not merely a factor of annual rainfall; it is the principal determinant of it. In numerous semi-arid regions, the single wettest month constitutes a significant proportion of the yearly precipitation total. The following months contribute minimally, with some recording nearly negligible amounts.

Figure 17 illustrates a robust connection (>0.78) between maximum monthly rainfall and total yearly rainfall across the sub-regions. Consequently, the country's precipitation is markedly inconsistent and frequently occurs in a limited number of extreme episodes. A single month of intense precipitation significantly impacts the annual total. The figure also shows that years characterized by relatively low maximum monthly precipitation typically exhibit reduced annual totals, indicating that in the absence of strong rainfall events, the yearly accumulation stays small. The surge in the graphs around 1988 and 2000 plainly demonstrates that one month of extraordinarily strong rainfall significantly increased the annual total. The years coincide with the occurrence of a strong La Niña event and tropical cyclone Eline, respectively. Figure 17 also indicates that Botswana's yearly precipitation is significantly influenced by the severity of its wettest month, suggesting that a few intense storms can determine the water balance for the entire year.

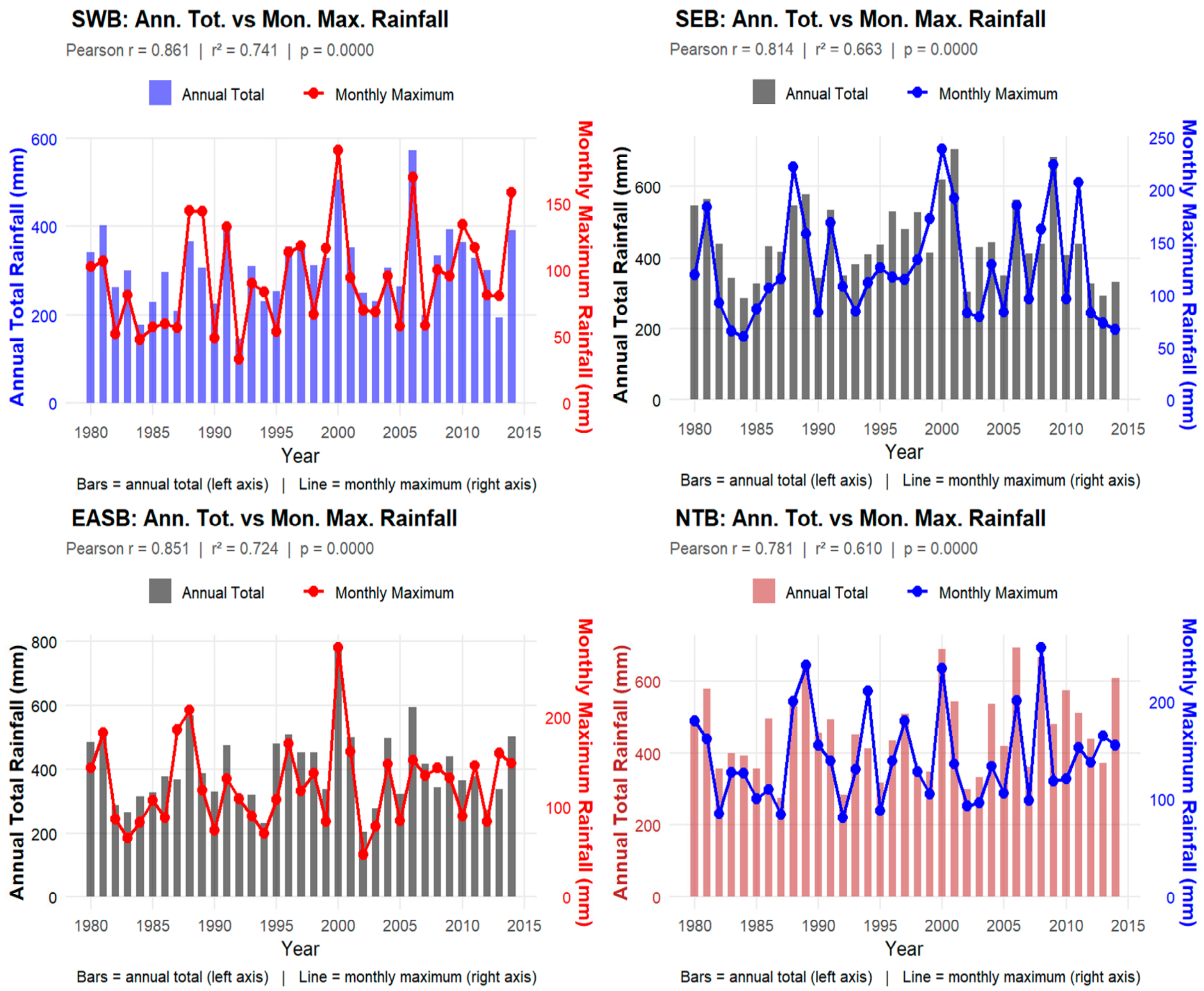


Figure 17. Correlation between total annual precipitation and maximum monthly values.

4.10. Best-Fitting Distribution Functions

4.10.1. Goodness-of-Fit Index

Table 5 presents the Z^{distr} index for the six (6) commonly utilized distribution functions applied to the precipitation data of the four homogenous sub-regions. If the candidate distribution functions possess Z^{distr} index values below 1.64, they are statistically acceptable fits at the 90% confidence level (Ahmad et al., [61]). Distribution functions with an index value exceeding 1.64 are categorically rejected. The Z^{distr} indices of the acceptable distributions are thereafter compared directly. A distribution with Z^{distr} values near zero typically yields a superior fit. Table 5 indicates that the Z^{distr} index values for PE3 and GPA distributions exceed 1.64 across all sub-regions. Consequently, they are dismissed as candidate distributions in the examined sub-regions.

Table 5. Goodness-of-fit measures (Z^{distr}) for the sub-regions.

Sub-Region Name	Distribution Functions					
	GLO	GEV	GNO	PE3	GPA	Weibull
SW-B	−0.77	−5.67	−6.32	−8.01	−16.58	−7.33
NT-B	8.92	0.96	2.51	2.39	−14.07	−1.63
SE-B	4.43	0.92	0.08	−1.65	−7.22	−3.97
EAS-B	17.80	9.60	8.18	−4.91	−8.96	0.67

Table 5 also indicates that the sole distribution with the lowest Z^{distr} index for the SWB sub-region is GLO (−0.77), rendering it a suitable distribution for the area. In the EAS-B sub-region, the sole probability density function with a Z^{distr} index magnitude below 1.64 is Weibull (0.67).

In the NT-B zone, the Z^{distr} statistic is below 1.64 for both the Weibull (−1.63) and GEV (0.96) distributions. The GEV distribution is regarded as the optimal model for the cluster's mean yearly precipitation due to the Z^{distr} index being nearest to zero. Nevertheless, the latter distribution is the ideal model for the cluster's mean annual precipitation, as it produces the lowest Z^{distr} value (0.96). The SEB sub-region also possesses two suitable PDFs: GEV (0.92) and GNO (0.08). Consequently, the optimal model for mean annual rainfall in the zone is GNO, as it is nearest to zero.

L-moment diagrams are another visual inspection technique for determining the best-fitting distribution model [46]. The L-moment ratio diagrams for the four identified homogeneous sub-regions are shown in Figure 18. The figures show a graphical technique for determining how closely the L-kurtosis of a candidate distribution matches the regional L-kurtosis by plotting observed regional L-skewness values against those of L-kurtosis. The best fit is the candidate distribution function that is closest to the regional average L-skewness and L-kurtosis.

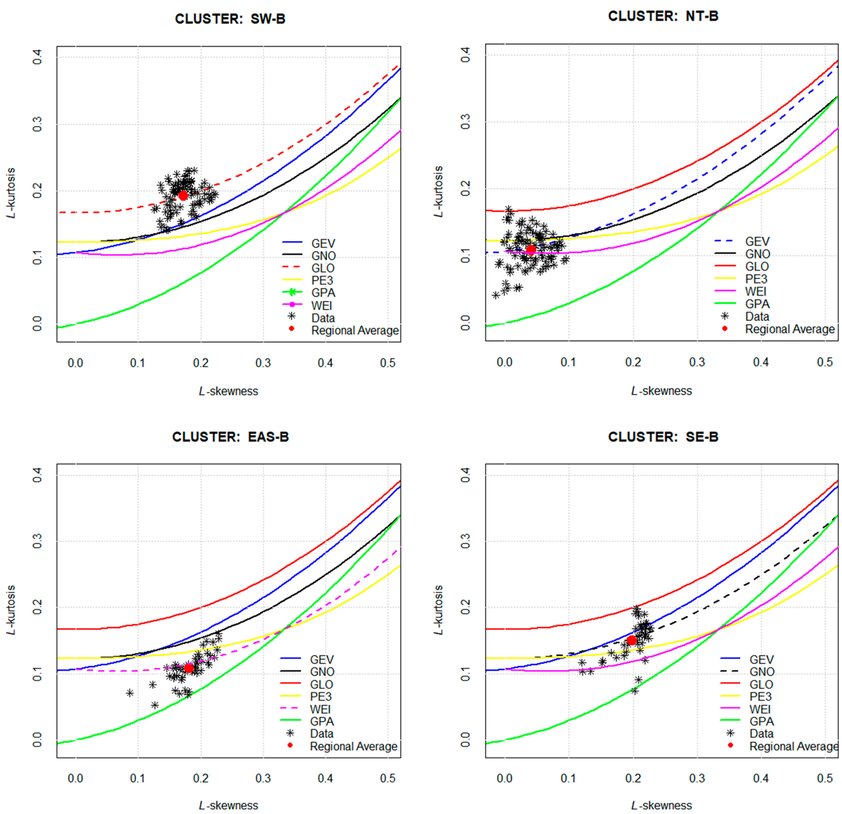


Figure 18. L-moment diagrams for the homogeneous sub-regions.

Plots for the six (6) frequently used distributions and the regionally averaged L-skewness and L-kurtosis points (hereafter referred to as Avgks, displayed in red in Figure 18) are displayed in the L-moment ratio diagram for the SW-B sub-region. GLO (dashed) is the closest distribution to the Avgks. For other pdfs, the L-kurtosis is even lower than zero. Weibull (dashed) is the distribution that is closest to the Avgks point for the EAS-B sub-region. GNO and GEV are the pdfs in the SE-B and NT-B sub-regions that are closest to the Avgks point, respectively. It is clear from the two goodness-of-fit techniques that GEV, GLO, GNO, and Weibull distributions could be used to model annual rainfall in the NT-B, SW-B, SE-B, and EAS-B sub-regions, respectively.

4.10.2. Distribution Parameters

In regional frequency analysis, each distribution is characterized by three parameters: the location parameter, ξ (xi); the scale parameter, α (alpha); and the shape parameter, κ (kappa). The SW-B sub-region is characterized by a generalized logistic (GLO) distribution with parameters indicating a moderate spread (α = 63.48), suggesting that typical event magnitudes exhibit significant variation around the mean and possess a negative skew (Table 6). GLO with negative κ generates one of the heaviest tails among prevalent distributions. A significantly negative κ, such as −0.195, induces an increase in quantile values in the high return periods.

Table 6. Probability distribution function parameters for the homogeneous regions.

Region Name	Probability Distribution	Distribution Parameters		
		ξ	α	κ
SW-B	GLO	336.67	63.48	−0.195
NT-B	GEV	338.71	96.16	−0.018
SE-B	GNO	374.14	101.95	−0.359
EAS-B	Weibull (three-parameter)	−171.03	372.42	0.333

From Table 6, SW-B represents the most severe tail-risk sub-region. The annual precipitation in the NT-B sub-region is characterized by the generalized extreme value (GEV) distribution.

It possesses the most extensive scale parameter among the four regions. It exhibits the greatest dispersion. κ = −0.018 indicates that it is nearly a Gumbel distribution (the generalized extreme value distribution with κ approaching zero converges to Gumbel). The top tail is unbounded yet increases only logarithmically, representing the lightest tail among the four.

The SE-B sub-region possesses the greatest location parameter (ξ = 374.14) and the most negative shape parameter (κ = −0.359) among the four areas. The GNO with this κ generates a significantly skewed distribution characterized by a pronounced upper tail. The EAS-B sub-region is optimally modelled by a three-parameter Weibull distribution, characterized by a positive shape factor (κ) of 0.333 (Table 6). The Weibull distribution with a positive shape parameter κ possesses a finite upper limit.

4.11. Quantile Functions for the Best-Fitting Distributions

The quantile functions (inverse cumulative distribution functions) of the optimal distribution that was fitted to the annual rainfall for two representative sub-region stations are depicted in Figure 19a. Kavimba (a red, continuous line) and Maun (a blue, dashed line) are the representative stations for the NT-B sub-region. In the lower tail (p < 0.25), the two contours converge at 150–300 mm, resulting in an almost indistinguishable appearance. The two stations demonstrated analogous efficacy during low-magnitude rainfall events, with comparable GEV location and scale parameters. Maun consistently outperforms Kavimba by 50–80 mm across all probabilities, as indicated by the mid-range plots ($0.25 < p < 0.75$) (see Figure 19a). This suggests that Maun's annual maxima are more significant due to its elevated scale or location parameter. Kavimba surpasses Maun at approximately $p \approx 0.97$, and the graphs converge and intersect in the upper tail ($p > 0.90$). Kavimba reaches approximately 840 mm, while Maun reaches approximately 750 mm as p approaches 1.00.

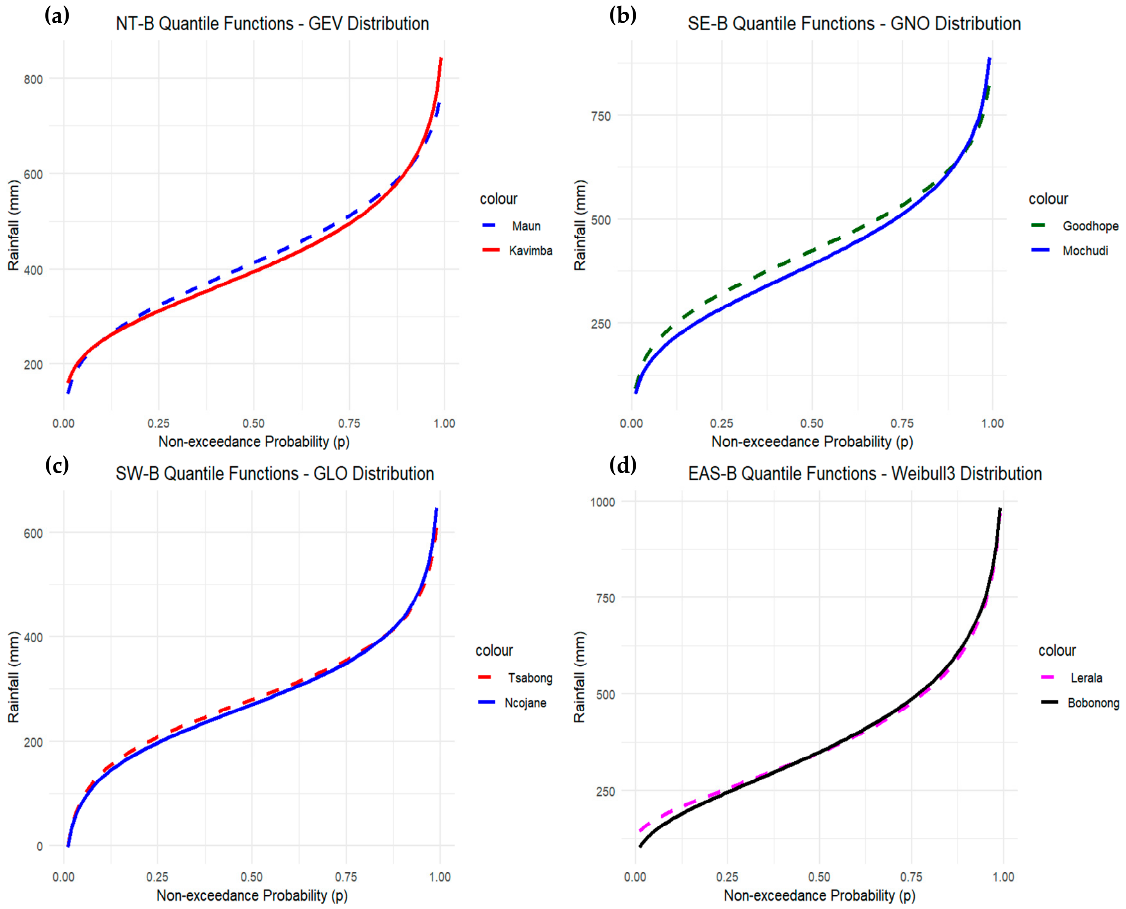


Figure 19. Quantile function plots for two representative gauge stations at the four sub-regions: (**a**) NT-B; (**b**) SE-B; (**c**) SW-B; and (**d**) EAS-B.

As p approaches one, both curves exhibit a distinct upward trajectory, which suggests a Fréchet-type generalized extreme value distribution and a positive shape parameter ($\xi > 0$). This suggests that Kavimba's tail is heavier than Maun's. The red curve's robust terminal ascent suggests that Kavimba's ξ is more positive.

Figure 19b depicts the quantile functions for Goodhope (green, dashed) and Mochudi (blue, solid) in relation to the non-exceedance probability p. The discrepancies in the graphs indicate station-specific scaling through L-moments, despite both stations being fitted to the same regional GNO distribution. The figure shows that the two curves markedly converge between 125 and 150 mm in the lower tail ($p < 0.10$) of the SE-B sub-region plots. Regional homogeneity at low return periods is enhanced by the almost identical low-probability quantiles generated by each station.

Divergence occurs near the mid-range ($0.10 < p < 0.75$), as seen from the graph (see Figure 19b). In the mid-range, the Goodhope plot consistently surpasses Mochudi's plot by 100–150 mm at $p = 0.50$. The median precipitation for Goodhope is from 450 to 480 mm, but for Mochudi it ranges from 330 to 350 mm. A probable reason for the discrepancy is that

Goodhope has a higher average yearly precipitation. Mochudi exceeds Goodhope, reaching around 875 mm when p approaches 1.00 in the upper tail ($p > 0.90$), when the plots converge and cross at $p \approx 0.95$–0.97. As p approaches 1.00, both curves have significant upward curvature, indicating, within the GNO framework, a negatively skewed shape parameter that results in a more noticeable upper tail than the standard normal distribution.

Figure 19c depicts the quantile functions of the generalized logistic (GLO) distribution for two representative stations in the SW-B region: Tsabong (red dashed) and Ncojane (blue solid). The figure illustrates that the stations converge closely, yielding virtually identical quantile values throughout the range of around 0–200 mm in the lower tail. This signifies robust regional uniformity at low-to-moderate probability. The calibrated GLO parameters at both sites exhibit similar lower-bound rainfall characteristics. The divergence is minimal, aligning with a well-fitted regional frequency model inside the distribution's body. In the intermediate range ($0.25 < p < 0.75$), a distinct and enduring divergence emerges. Tsabong (red) is positioned above Bray (blue) within this area by roughly 30–50 mm at any specified probability. This indicates that Tsabong's fitted GLO distribution has a greater center tendency possibly because of a bigger location parameter (ξ) or scale parameter (α). This implies that median-range rainfall events are consistently more intense in Tsabong compared to Bray.

Figure 19c shows that both curves have pronounced, increasing upward curvature as p approaches 1.0, indicative of GLO's substantial upper tail. In contrast to the GEV or Gumbel distributions, the GLO distribution has a shape parameter κ that, when negative, results in an unbounded upper tail, indicating that quantile values increase indefinitely as p approaches one.

Figure 19d shows the three-parameter Weibull annual rainfall quantile functions for two representative gauge stations (Lerala and Bobonong) in the EASB sub-region. The curves monotonically increase and are concave, rising gradually through the central probabilities and steepening moderately in the upper tail. For the EASB sub-region specifically, the eastern rainfall regime, which is influenced by the Indian Ocean moisture flux and the Inter-tropical Convergence Zone, tends to produce a moderately right-skewed annual rainfall distribution, which the three-parameter Weibull distribution captures well through a shape parameter κ typically in the range of 1.5 to 3.0 for stations in this sub-region.

In the lower tail region ($p < 0.25$), the lower bound $\xi > 0$ is the most distinctive feature of the three-parameter Weibull fit. Unlike unbounded distributions, three-parameter Weibull distribution asserts that there is a physically meaningful minimum annual rainfall below which the station will not fall. For EASB stations, this lower bound typically sits in the range of 80–180 mm, consistent with the semi-arid to sub-humid transition zone in eastern Botswana (Figure 19d).

Figure 19d shows that at the central range region ($0.25 \leq p \leq 0.75$), the three-parameter Weibull fit is the most reliable, and this is where the two stations should behave most similarly if EASB has been correctly delineated as a homogeneous region. The quantile curve is approximately linear on a probability scale, indicating that median rainfall years are well-described and predictable. The median rainfall Q (0.5) for EASB stations typically falls between 350 and 550 mm, consistent with eastern Botswana's long-term records.

In the upper tail region ($p > 0.75$), the curve continues rising but does not diverge wildly, a characteristic that makes the three-parameter Weibull distribution conservative relative to GLO or GEV for the same rainfall data. For EASB, occasional anomalously wet years do occur due to tropical cyclone remnants.

4.12. Performance Indices for the Best-Fitting Distributions

Figure 20 shows that GEV displays strong performance in the NTB sub-region regarding correlation (mean r = 0.987) and satisfactory performance concerning error size (mean CV-RMSE ≈ 5.08%); however, it reveals a consistent and significant systematic underestimate (mean bias ≈ −0.49). The zonal-level fit is unequivocally superior to the station-level fit across all indices, as anticipated due to spatial aggregation; nonetheless, it should not obscure the constraints inherent at the station scale. Any return period estimations or design values obtained from this GEV fitting must be bias corrected prior to practical use.

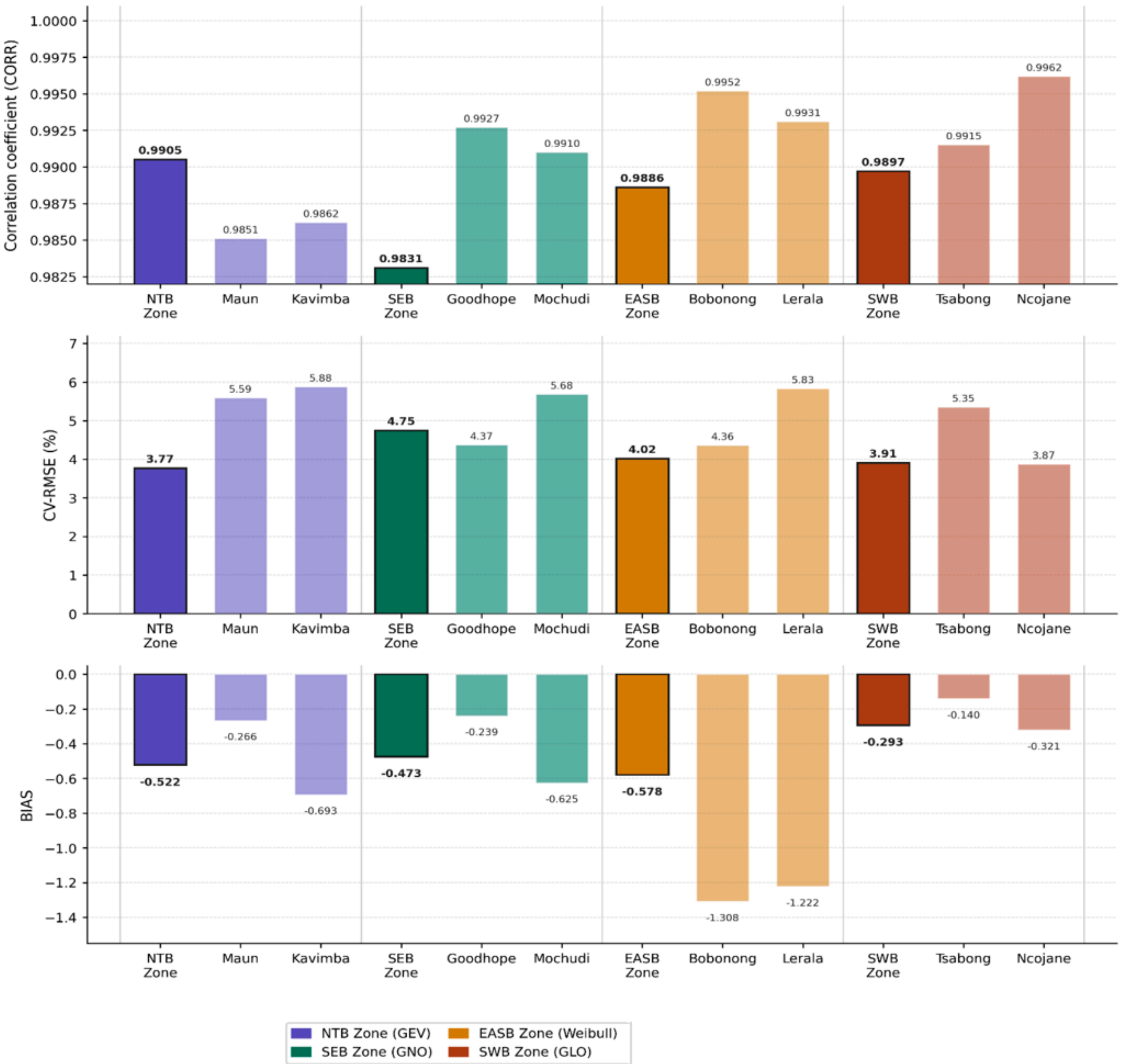


Figure 20. Probability distribution function performance by zone and representative gauge stations.

The average performance indices demonstrate a good match of the Weibull distribution with the mean correlation coefficient of 0.9923, mean CV-RMSE of 4.74%, which is well below the 10% criterion, and the mean bias of −1.036 mm. The underestimation constitutes a valid deficiency that requires consideration.

The GNO distribution demonstrates a superior match across the SEB sub-region according to all three average performance metrics (refer to Figure 20). The correlation is robust, the spread error is well managed, and the bias is very little. This distribution performs well in this zone. The table indicates that the GNO distribution effectively reflects the observed annual rainfall data patterns across the SEB sub-region and its representative stations, with a mean correlation value of 0.9889. The correlation between fitted and observed values is consistently high throughout all three sites.

A mean CV-RMSE of 4.93% the SEB sub-region is far below the 10% excellence level, indicating that the GNO distribution accurately replicates observed values with little relative error on average. Like the correlation coefficient, this statistic remains consistently reliable throughout the zone and both stations. A mean bias of −0.446 is quite minor, suggesting that systematic underestimating is minimal (Figure 20).

Figure 20 demonstrates that the GLO distribution provides an exceptional match for the SWB sub-region across all three average performance measures. The GLO distribution is notable for its bias control, with a mean CORR of 0.9925, a mean CV-RMSE of 4.38%, and a mean bias of just −0.252 mm, the lowest seen among all zones and distributions examined to date. The spread error is among the most competitive, confirming that the GLO distribution faithfully follows known wind speed patterns with minimum systematic distortion.

According to all three criteria, the three-parameter Weibull distribution exhibits outstanding performance in the EASB sub-region, as seen in Figure 20. The negative bias is the only substantial limitation. The mean correlation value of 0.9923 is very high, indicating that the three-parameter Weibull distribution well represents the overall shape and pattern of the rainfall data across the region. This is hardly a cause for celebration. The mean CV-RMSE of 4.74% is below the widely recognized 10% criterion for good model fit, although it is alarmingly near that limit. A mean bias of −1.036 mm indicates a consistent underestimate by the Weibull distribution over the whole area.

4.13. Bias Correction of CHIRP (Ver. 2.0) Data

PT Bias Correction of Monthly CHIRPS (Ver. 2.0) Data

Section 4.12 indicates that CHIRPS (ver. 2.0) consistently underestimate precipitation data from gauge stations across all sub-regions. PT bias correction has been employed to adjust the variance in satellite data for improved alignment with ground-measured monthly data. This sub-section demonstrates the bias correction of CHIRPS data using the SEB sub-region as a case study, with the potential for broader application to other sub-regions. The outcomes of the PT bias correction are depicted in Figure 21. The figure reveals that both the raw and bias-adjusted CHIRPS data for SEB exhibit a pronounced seasonal pattern, marked by a peak rainfall period from January to March, a cessation from April to June, and a dry season from July to August.

Figure 21 shows that raw CHIRPS data (red) consistently shows lower magnitudes compared to the gauge measurements This verifies a persistent PBIAS underestimate of −9.56% as illustrated in Table 7. Performance improvements are seen after the power transformation bias adjustment, particularly during the peak precipitation months (December to March) and during the cessation period. The corrected data closely correspond to the gauge data, signifying enhanced magnitude representation (Figure 21). This indicates that the bias-corrected CHIRPS (version 2.0) is appropriate for hydrological applications in the SEB sub-region.

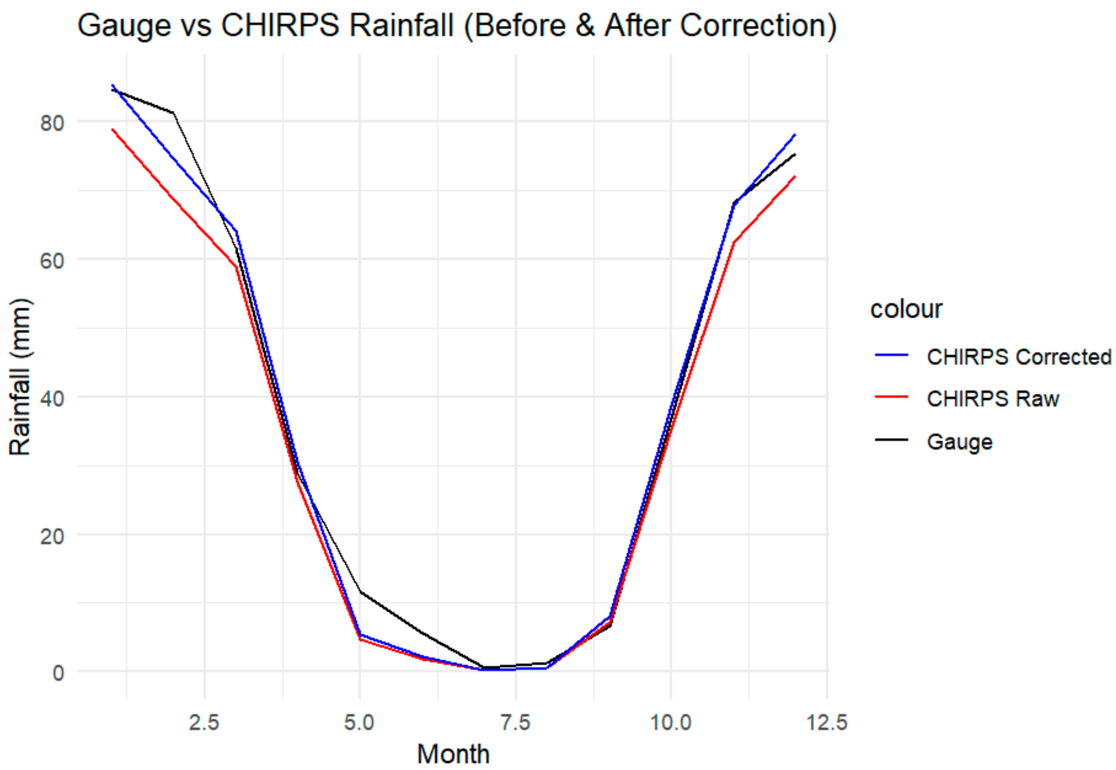


Figure 21. Power transformation bias correction of mean monthly CHIRPS data.

Table 7. Bias correction of SE-B sub-region average monthly CHIRPS precipitation data.

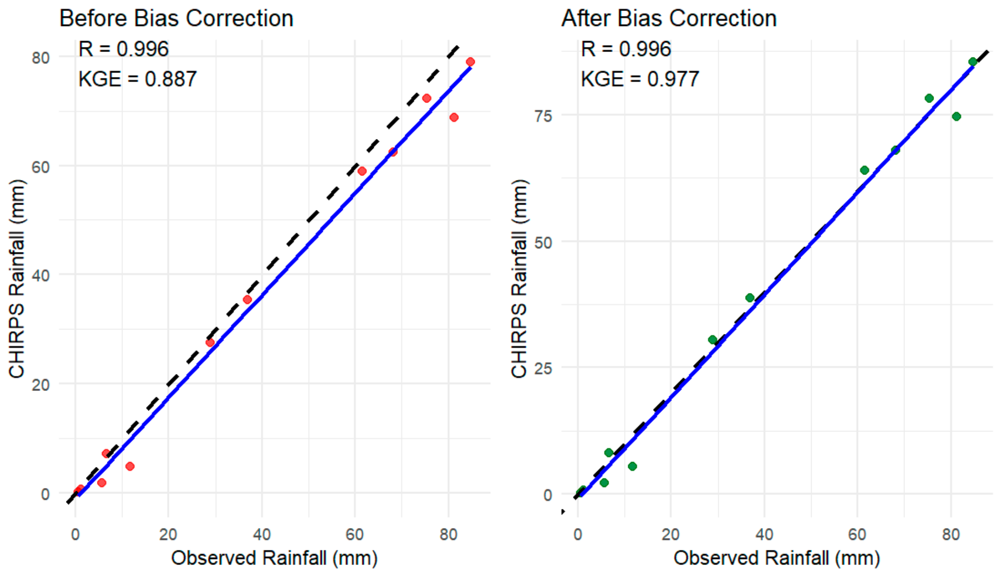
Index:	Before Correction	After Correction	Change
RMSE (mm)	5.065	3.160	A significant decrease
CV-RMSE (%)	13.130	8.180	A significant decrease
PBIAS (%)	−9.560	−1.550	A significant reduction of underestimation
Correlation (R)	0.996	0.996	Unchanged
KGE	0.887	0.977	Strong improvement

The PT bias correction made a considerable contribution to increasing the accuracy of rainfall estimations in the sub-region. The analysis of indices shows that an improvement can be observed for all important characteristics. As can be seen from Table 7, there is a considerable decrease in errors since the RMSE has become smaller, meaning that the total error has been lowered. Moreover, the CV-RMSE is reduced, which means improved stability of measurements.

In Table 7, PBIAS improved from −9.56% to −1.55%, indicating an initial tendency for the model to predict low rainfall using CHIRPS data. The correction has significantly reduced the bias (a bias of less than ±5% is considered excellent). The above conclusion reveals that the bias correction process was relatively successful. However, what needs to be pointed out is that even after bias correction, the correlation coefficient is extremely high (approaching 0.996).

Table 7 shows that the Kling–Gupta efficiency (KGE) enhanced markedly from 0.887 to 0.977. The KGE is significant as it assesses the correlation (R), bias ratio (β), and variability ratio (γ). Prior to correction, CHIRPS (ver. 2.0)exhibited a favorable albeit minor imbalance in bias and variability; after correction, it demonstrated nearly flawless reproduction of rainfall features. KGE affirms that enhancement is evident not only in magnitude but also in variability, distribution, and hydro-meteorological consistency. Nevertheless, there are potential issues associated with the implemented bias correction method, as the RMSE remains significant post-correction.

Figure 22 further illustrates how the CHIRPS data consistently underestimates observed precipitation. The regression line (blue) in the figure, which is below the 1:1 dashed line, clearly confirms the underestimation. After bias correction, the blue regression line exactly lines up with the 1:1 dashed line throughout the whole range. Furthermore, by showing that the discrepancy is more apparent at high rainfall levels, the figure depicts conditional bias, in which errors increase with magnitude. The dispersion at low and high precipitation levels is depicted in both panels (Figure 22). The bias adjustment has not corrected the dispersions. The potential challenges associated with extreme rainfall events are depicted in the illustration.



(**a**) (**b**)

Figure 22. Scatter plots showing the performance of monthly CHIRPS (ver. 2.0) data (**a**) before bias correction and (**b**) after bias correction using PT.

Figure 23a,b show the comparison of observed SEB sub-region gauge rainfall, the raw CHIRPS estimates, and bias-corrected CHIRPS data using power transformation and quantile mapping methods. The gauge rainfall dataset was used as the benchmark for assessing the efficiency of the proposed corrections.

As observed from Figure 23, the raw CHIRPS dataset shows significant systematic biases in estimating annual rainfall when compared to observed gauge rainfall. This means that the satellite-derived dataset consistently underestimates rainfall values in relation to gauge observations. Moreover, this underestimation effect is even higher during wet periods with gauge peaks around 650–700 mm being significantly lower at about 500–600 mm on the CHIRPS record (see Figure 23).

Using PT correction for rainfall data resulted in improving the dataset in question (Figure 24a). Specifically, PT produced an upward shift in rainfall estimates in relation to the raw CHIRPS dataset. As a result, the PT-corrected CHIRPS data become more similar to gauge measurements, showing an improved representation of mean values and increasing interannual variability.

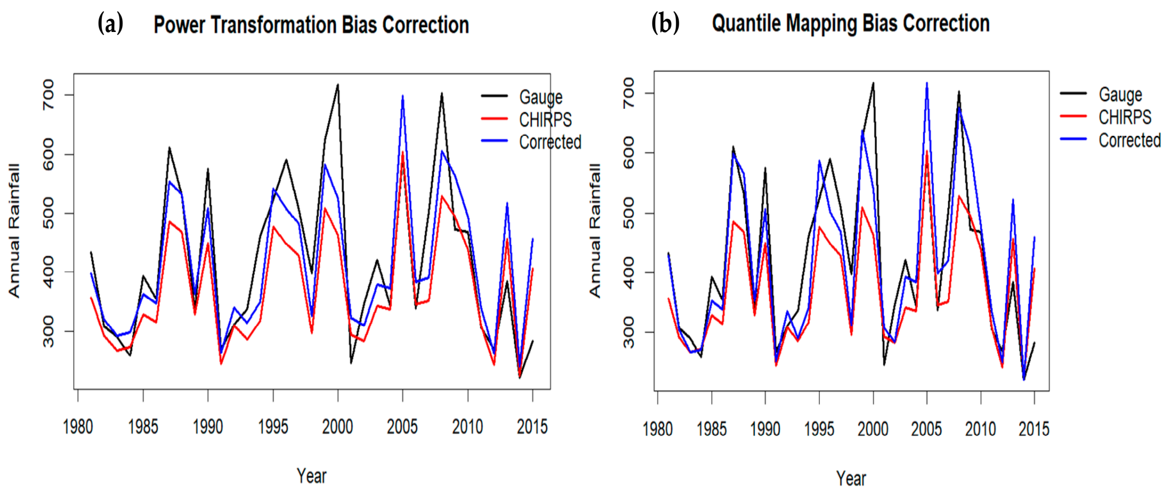


Figure 23. Time series plots for the comparison of spatial averaged gauge station data with average raw and corrected CHIRPS (ver. 2.0) data using (**a**) P T and (**b**) Q M methods.

Quantile mapping showed the best results compared to other approaches (Figure 23b). Thus, QM was able to reproduce low and high rainfall values much closer to the corresponding observed data points (Figure 23b). Specifically, QM managed to improve the estimates of peaks in particular years, especially 1999–2000 and 2005.

Moreover, QM helped retain the temporal variation in rainfall recorded in gauge data, matching peaks and valleys to those in observed datasets. As a result, QM demonstrated the ability to adjust not only the mean rainfall value but also its variability, including extremes. Comparing the two methods described above, it can be observed that quantile mapping has superior characteristics to power transformation. While PT was able to address issues related to systematic biases and average rainfall values, QM provided a better representation of rainfall peaks and variability.

Figure 24 shows the scatter plots of CHIRPS rainfall estimates versus gauge observations before and after using PT and QM bias-correction approaches. A solid black line is the 1:1 line, while the dashed lines represent regression fits.

As seen from Figure 24a, there is significant scatter between satellite estimates and observations. There is systematic bias because all points are situated to the right of the 1:1 reference line. This is especially noticeable when dealing with higher rainfall values; hence, there is an underestimation bias. The significant scatter also means that there are random errors associated with high inconsistency. The deviation of the regression line from the 1:1 reference further proves low correlation between satellite estimates and observations (Figure 24a).

After using the PT bias correction technique, there was a marked increase in consistency between CHIRPS estimates and gauge observations (Figure 24b). Data points cluster around the 1:1 line much closer compared to the previous case, which means that both systematic and random errors are decreased. However, deviations still occur, especially when dealing with larger rainfalls. Hence, PT can decrease bias but cannot eliminate differences (Figure 24b).

Using QM as a method of bias correction achieves the best results because the regression line almost coincides with the 1:1 line (Figure 24c). Thus, there is high consistency between estimates and observations. In addition, the scatter of points is noticeably smaller than in other cases. This confirms the efficiency of the QM technique for reducing errors.

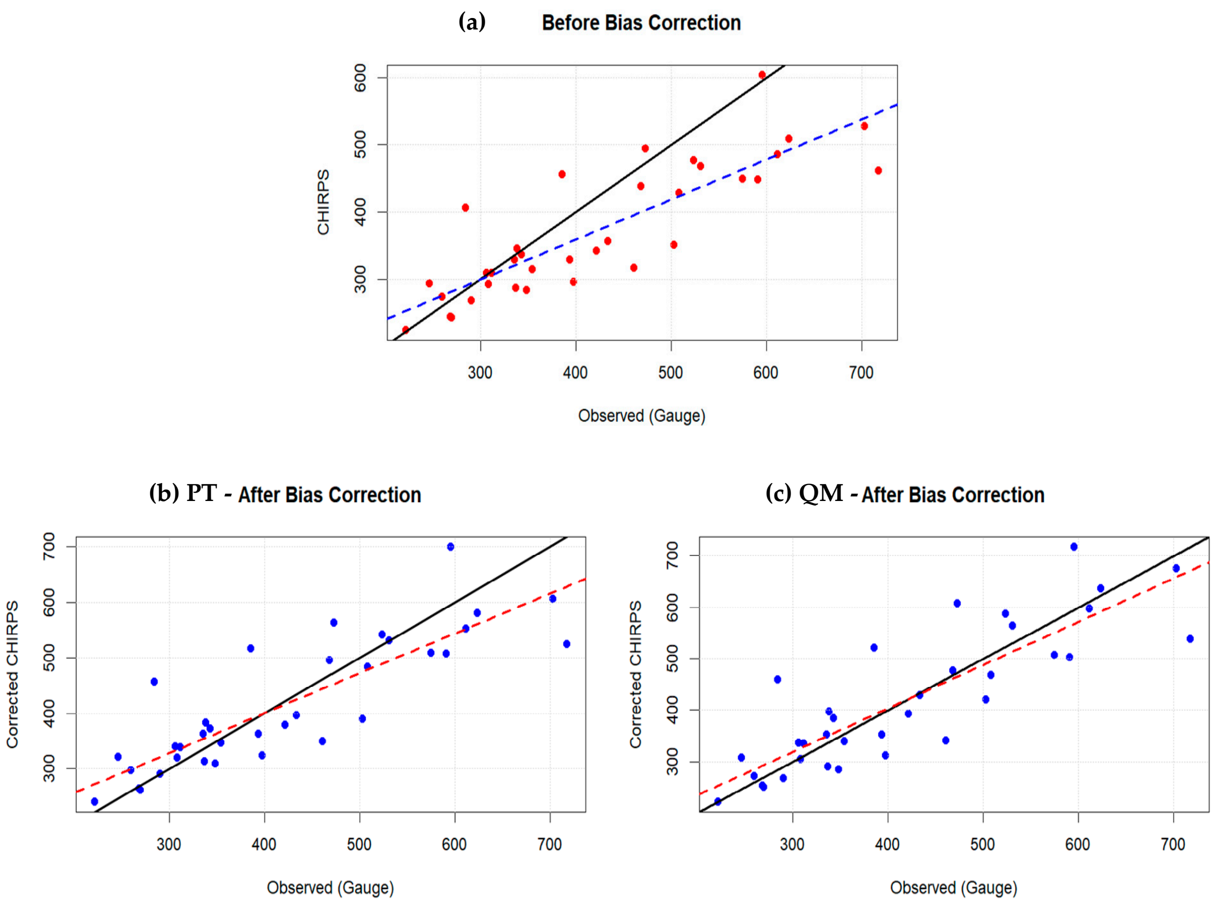


Figure 24. Scatter plots showing the performance of CHIRPS (ver. 2.0) (**a**) before bias correction, (**b**) after bias correction using PT, and (**c**) after bias correction using QM methods.

Therefore, the conclusion can be drawn that both techniques used to correct bias are effective because they result in increasing accuracy of estimates. The original dataset demonstrates an underestimation bias and high stochastic variability. Bias correction techniques decrease these but to different degrees. The most successful is the QM approach.

Bias correction results demonstrate that the two techniques differ in efficiency by strength. Figure 25 shows that PT significantly mitigates the negative bias for the SEB sub-region down to only 0.08% from −11.71%, eliminating systematic bias present in CHIRPS data. The figures shows that QM decreases the bias from −17.35% to −9.55%; still, underestimation is present, which may be due to inaccurate quantiles or a limited sample size.

The PT method transforms and corrects the mean bias and does not influence the correlation or the rank order (KGE components); QM keeps the correlation (R = 0.993) constant and even brings R up to 0.996, benefiting the correlation. KGE benefits more from the QM technique.

Figure 25 also shows that PT increases from 0.53 to 0.75 (+41.5%), while QM increases from 0.77 to 0.89 (+15%). Therefore, QM proves itself to be more efficient in KGE, R, and other aspects. PT notably succeeds only in eliminating biases. Bias correction significantly improves CHIRPS data. PT proves more efficient in bias elimination; QM demonstrates high performance in all metrics including KGE, bias, and RMSE.

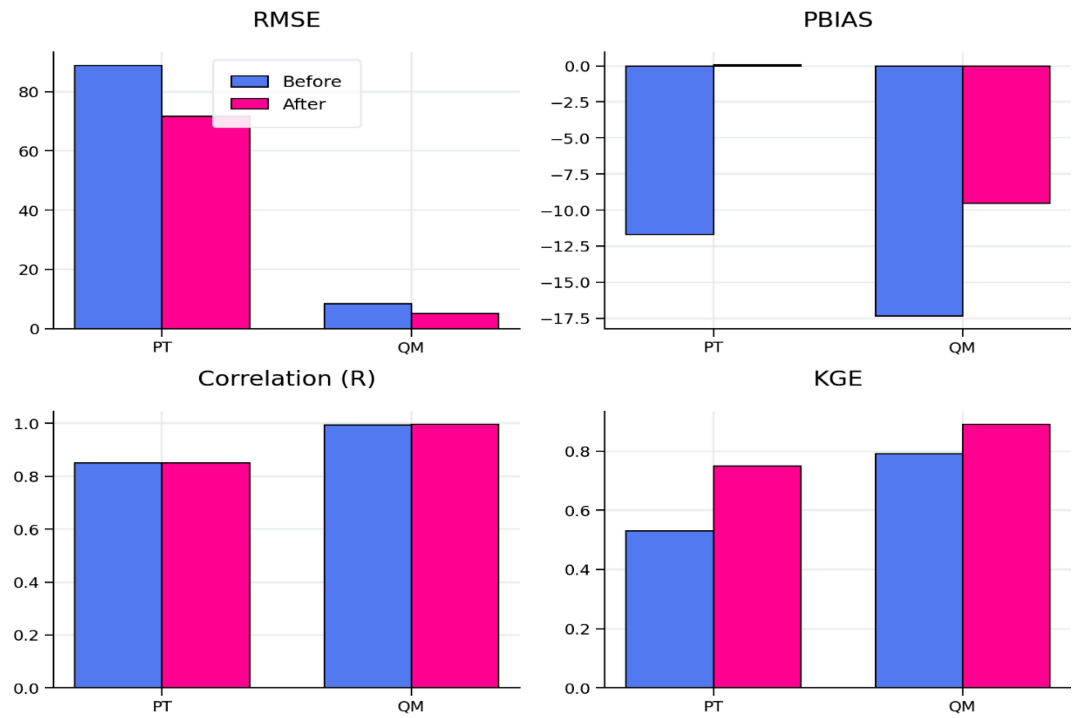


Figure 25. Performance metrics for the CHIRPS (ver. 2.0) dataset before and after bias correction.

5. Conclusions

This research examined significant patterns of regional and temporal variation in monthly rainfall in semi-arid Botswana by eigenvector analysis. The initial section of the study focused on evaluating the effectiveness of CHIRPS (ver. 2) as a substitute for gauge station data in Botswana at a monthly time scale. The satellite precipitation estimation product demonstrated reliability in estimating in situ precipitation data. The product exhibits a robust correlation ($r > 0.97$) with the observed monthly data, alongside minimal MAE values of 0.64 mm/month. However, the correlation coefficient is diminished (<0.95) in dry and drought-prone areas, which are marked by vegetation degradation.

This is distinctly demonstrated by the KGE index. Relatively low values of KGE (<0.83) were observed in the vegetation-eroded areas of the Kalahari, Makgadikgadi pans, and Bobirwa region, as well as in the urbanized southeastern portion of Botswana. On average, high values in the KGE index (0.94), low negative bias (−1.57 mm/month), and a robust spatial correlation (0.97) exist between monthly CHIRPS (ver. 2) and gauge station data over the whole country. Consequently, because of its high efficacy, CHIRPS (ver. 2) was used as a substitute for gauge station data in the country's rainfall variability analysis.

Using CHIRPS (ver. 2) gridded data for eigenvector analysis, three principal components with a cumulative percentage variation of 88% were retained for use as inputs for the non-hierarchical K-means clustering procedure. The three main components reflected intra-annual fluctuations in rainfall across the country. Pc1 had the highest percentage variance of 51.3% and depicted the highest association with months from April to September. The months represented the cessation of the rainy season. Pc2, with a percentage variance of 28.6%, had the highest association with the months from December to March. This aligns with precipitation that occurs during the late summer. The final component with a percentage variance of 7.9%, Pc3, corresponds to the start of summer precipitation.

As per the requirement of K-means clustering, the optimal number of homogeneous precipitation clusters was determined using the average silhouette width index. The indices calculated that the optimum number of homogeneous zones is four (4). This was consistent with a previous investigation conducted by Nkoni et al. [6]. The clusters were referred to as

NT-B, SW-B, SE-B, and EAS-B. The distinct rainfall patterns of the sub-regions are a result of the varying large-scale oceanic and atmospheric moisture sources. The incursion extent of the ITCZ influences precipitation in the NT-B and SW-B sub-regions, while easterly waves affect rainfall in both the EAS-B and SW-B zones. The westerly troughs have an impact on SE-B.

Author Contributions: G.A.N. developed the analytical methodology and conducted the data analysis. K.M.M. and N.C.M. provided supervision to the main author, and together with S.H.S., offered insightful feedback on the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: The research was funded by the Department of Staff Welfare and Training at the University of Botswana, and the APC was funded by Office of Research and Development (ORD), University of Botswana.

Data Availability Statement: The data used in this study was obtained from the Department of Meteorological Services in Botswana (gauged station rainfall) and can be made available upon request. The CHIRPS (ver.2.0) and CHIRTS daily data (Tmin and Tmax) were obtained online via their respective web sites: https://data.chc.ucsb.edu/products/CHIRPS-2.0/ (accessed on 14 April 2025). CHIRTS data is obtainable from the website: https://data.chc.ucsb.edu/products/CHIRTSdaily/ (accessed on 20 May 2026).

Acknowledgments: The authors express gratitude to the Department of Staff Welfare and Training at the University of Botswana for providing financial support to the main author. They are also thankful to the Botswana Department of Meteorological Services (BDMS) and the University of California for providing gauge station data and CHIRPS ver. 2 gridded dataset.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Kotir, J.H. Climate change and variability in Sub-Saharan Africa: A review of current and future trends and impacts on agriculture and food security. *Environ. Dev. Sustain.* **2011**, *13*, 587–605. [CrossRef]
2. Kim, S.; Shin, J.-Y.; Heo, J.-H. Assessment of Future Rainfall Quantile Changes in South Korea Based on a CMIP6 Multi-Model Ensemble. *Water* **2025**, *17*, 894. [CrossRef]
3. Satyanarayana, P.; Srinivas, V.V. Regionalization of precipitation in data sparse areas using large scale atmospheric variables—A fuzzy clustering approach. *J. Hydrol.* **2011**, *405*, 462–473. [CrossRef]
4. Baeriswyl, P.A.; Rebete, M. Regionalization of precipitation in Switzerland by means of principal component analysis. *Theor. Appl. Climatol.* **1997**, *58*, 31–41. [CrossRef]
5. Domroes, M.; Kaviani, M.; Schaefer, D. An Analysis of Regional and Intra-annual Precipitation Variability over Iran using Multivariate Statistical Methods. *Theor. Appl. Climatol.* **1998**, *61*, 151–159. [CrossRef]
6. Nkoni, G.; Mphale, K.; Mbangiwa, N.; Samuel, S.; Molosiwa, R. Use of multivariate techniques to regionalize rainfall patterns in semiarid Botswana. *Discov. Environ.* **2024**, *2*, 77. [CrossRef]
7. Hosking, J.R.M.; Wallis, J.R. Regional frequency analysis. In *An Approach Based on L-Moments*; Cambridge University Press: Cambridge, UK; New York, NY, USA; Melbourne, Australia, 1997; p. 224.
8. Goyal, M.K.; Sharma, A. A fuzzy c-means approach regionalization for analysis of meteorological drought homogeneous regions in western India. *Nat. Hazards* **2016**, *84*, 1831–1847. [CrossRef]
9. Roushangar, K.; Alizadeh, F.; Adamowski, J. Exploring the effects of climatic variables on monthly precipitation variation using a continuous wavelet-based multiscale entropy approach. *Environ. Res.* **2018**, *165*, 176–192. [CrossRef] [PubMed]
10. MacQueen, J. Some methods for classification and analysis of multivariate observations. In Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 21 June–18 July 1965; pp. 281–296.
11. Eslamian, S.S.; Feizi, H. Maximum monthly rainfall analysis using L-moments for an arid region in Isfahan province, Iran. *J. Appl. Meteorol. Climatol.* **2007**, *46*, 494–503. [CrossRef]
12. Parida, B.P.; Moalafhi, D.B. Regional rainfall frequency analysis for Botswana using L-moments and radial basis function network. *Phys. Chem. Earth Parts A/B/C* **2008**, *33*, 614–620. [CrossRef]
13. Satyanarayana, P.; Srinivas, V.V. Regional frequency analysis of precipitation using large-scale atmospheric variables. *J. Geophys. Res. Atmos.* **2008**, *113*, D24110. [CrossRef]

14. Nam, W.; Shin, H.; Jung, Y.; Joo, K.; Heo, J.-H. Delineation of the climatic rainfall regions of South Korea based on a multivariate analysis and regional rainfall frequency analyses. *Int. J. Climatol.* **2015**, *35*, 777–793.
15. Hazarika, J.; Sarma, A.K. Importance of regional rainfall data in homogeneous clustering of data-sparse areas: A study in the upper Brahmaputra valley region. *Theor. Appl. Climatol.* **2021**, *145*, 1161–1175. [CrossRef]
16. Dikbas, F.; Firat, M.; Koc, A.C.; Gungor, M. Classification of precipitation series using fuzzy cluster method. *Int. J. Climatol.* **2012**, *32*, 1596–1603. [CrossRef]
17. Gocic, M.; Velimirovic, L.; Stankovic, M.; Trajkovic, S. Regional precipitation-frequency analysis in Serbia based on methods of L-moment. *Pure Appl. Geophys.* **2021**, *178*, 1499–1511. [CrossRef]
18. Guttman, N.B. The use of L-moments in the determination of regional precipitation climates. *J. Clim.* **1993**, *6*, 2309–2325. [CrossRef]
19. Regenmortel, G.V. Regionalization of Botsowana Rainfall during the 1980s Using Principal Component Analysis. *Int. J. Clim.* **1995**, *5*, 313–323. [CrossRef]
20. Main, J.P.L.; Hewitson, B.C. Regionalization of daily precipitation in Botswana, 1972–1989. *S. Afr. Geogr. J.* **1995**, *77*, 51–55. [CrossRef]
21. Chipanshi, A.C.; Maphanyane, J.G. Nature of rainfall variability in Botswana over the period 1961–1990. *J. East. Afr. Res. Dev.* **1997**, *27/28*, 299–317.
22. Jamshidi, Z.; Samani, N. Mapping the spatiotemporal diversity of precipitation in Iran using multiple statistical methods. *Theor. Appl. Climatol.* **2022**, *150*, 893–907. [CrossRef]
23. Sahu, R.; Verma, M.; Ahmad, I. *Regional Frequency Analysis Using L-Moment Methodology—A Review*; Springer: Singapore, 2021; pp. 811–832.
24. Sahu, R.T.; Verma, M.K.; Ahmad, I. Some non-uniformity patterns spread over the lower Mahanadi River Basin, India. *Geocarto Int.* **2021**, *37*, 8792–8814. [CrossRef]
25. Darand, M.; Mansouri Daneshvar, M.R. Regionalization of Precipitation Regimes in Iran Using Principal Component Analysis and Hierarchical Clustering Analysis. *Environ. Process.* **2014**, *1*, 517–532. [CrossRef]
26. Boluwade, A. Spatial-Temporal Assessment of Satellite-Based Rainfall Estimates in Different Precipitation Regimes in Water-Scarce and Data-Sparse Regions. *Atmosphere* **2020**, *11*, 901. [CrossRef]
27. Kessabi, R.; Hanchane, M.; Caloiero, T.; Pellicone, G.; Addou, R.; Krakauer, N.Y. Analyzing Spatial Trends of Precipitation Using Gridded Data in the Fez-Meknes Region, Morocco. *Hydrology* **2023**, *10*, 37. [CrossRef]
28. Funk, C.; Peterson, P.; Landsfeld, M.; Pedreros, D.; Verdin, J.; Shukla, S.; Husak, G.; Rowland, J.; Harrison, L.; Hoell, A.; et al. The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Sci. Data* **2015**, *2*, 150066. [CrossRef] [PubMed]
29. Kummerow, C. The Status of the Tropical Rainfall Measuring Mission (TRMM) after Two Years in Orbit. *J. Appl. Meteorol. Climatol.* **2000**, *39*, 1965–1982. [CrossRef] [PubMed]
30. Paredes-Trejo, F.J.; Barbosa, H.A.; Kumar, T.L. Validating CHIRPS-based satellite precipitation estimates in Northeast Brazil. *J. Arid Environ.* **2017**, *139*, 26–40. [CrossRef]
31. Berihu, T.; Chen, W.; Wang, L. Rainfall variability and its teleconnection with atmospheric circulation anomalies over southern and southeastern region, Ethiopia. *Theor. Appl. Clim.* **2024**, *155*, 5819–5834. [CrossRef]
32. du Plessis, J.A.; Kibii, J.K. Applicability of CHIRPS-based satellite rainfall estimates for South Africa. *J. S. Afr. Inst. Civ. Eng.* **2021**, *63*, 43–54. [CrossRef]
33. Dlamini, W.M.D.; Tfwala, S.S. The performance of a high-resolution satellite-derived precipitation product over the topographically complex landscape of Eswatini. *Geosci. Data J.* **2025**, *12*, e278. [CrossRef]
34. Almazroui, M.; Saeed, S. Contribution of extreme daily precipitation to total rainfall over the Arabian Peninsula. *Atmos. Res.* **2020**, *231*, 104672. [CrossRef]
35. Howard, E.; Washington, R. Drylines in southern Africa: Rediscovering the Congo air boundary. *J. Clim.* **2019**, *32*, 8223–8242. [CrossRef]
36. Nashwan, M.S.; Shahid, S.; Wang, X. Assessment of satellite-based precipitation measurement products over the hot desert climate of Egypt. *Remote Sens.* **2019**, *11*, 555. [CrossRef]
37. Funk, C.C.; Peterson, P.J.; Landsfeld, M.F.; Pedreros, D.H.; Verdin, J.P.; Rowland, J.D.; Romero, B.E.; Husak, G.J.; Michaelsen, J.C.; Verdin, A.P. A quasi-global time series for drought monitoring. *U.S. Geol. Surv. Data Ser.* **2014**, *834*, 4. [CrossRef]
38. Keoagile, K.H.; Wiston, M.; Mbangiwa, N.C. Assessing Crop Yield Variability Using Meteorological Drought Indices for Agricultural Drought Monitoring in Botswana. *Climate* **2026**, *14*, 77. [CrossRef]
39. Bai, P.; Liu, X. Evaluation of five satellite-based precipitation products in two gauge-scarce basins on the Tibetan Plateau. *Remote Sens.* **2018**, *10*, 1316. [CrossRef]
40. Horn, J.L. A rationale and test for the number of factors in factor analysis. *Psychometrika* **1965**, *30*, 179–185. [CrossRef] [PubMed]
41. Cattell, R.B. The Scree Plot Test for the Number of Factors. *Multivar. Behav. Res.* **1966**, *1*, 140–161. [CrossRef] [PubMed]

42. Roushangar, K.; Alizadeh, F. A multiscale spatio-temporal framework to regionalize annual precipitation using k-means and self-organizing map technique. *J. Mt. Sci.* **2018**, *15*, 1481–1497. [CrossRef]
43. Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [CrossRef]
44. Dudek, A. Silhouette Index as Clustering Evaluation Tool. In *Classification and Data Analysis. SKAD 2019. Studies in Classification, Data Analysis, and Knowledge Organization*; Jajuga, K., Batóg, J., Walesiak, M., Eds.; Springer: Cham, Switzerland, 2020. [CrossRef]
45. Modarres, R.; Sarhadi, A. Statistically-based regionalization of rainfall climates of Iran. *Glob. Planet Change* **2011**, *75*, 67–75. [CrossRef]
46. Hosking, J.R.M.; Wallis, J.R. Some statistics useful in regional frequency analysis. *Water Resour. Res.* **1993**, *29*, 271–281. [CrossRef]
47. Noto, L.V.; La Loggia, G. Use of L-moments approach for regional flood frequency analysis in Sicily, Italy. *Water Resour. Manag.* **2009**, *23*, 2207–2229. [CrossRef]
48. Almuhayfith, F.E.; Kachour, M.; Daghestani, A.F.; Rehman, Z.U.; Hussain, T.; Bakouch, H.S. A Lower-Bounded Extreme Value Distribution for Flood Frequency Analysis with Applications. *Mathematics* **2025**, *13*, 3378. [CrossRef]
49. Liu, S.; Han, X.; Zhang, X.; Zhao, B.; Xie, L. Parameter Estimation of the Three-Parameter Weibull Distribution Based on an Iterative CDF Method. *Mathematics* **2026**, *14*, 649. [CrossRef]
50. Shepard, D. A Two-Dimensional Interpolation Function for Irregularly-Spaced Data. In Proceedings of the 1968 ACM National Conference, New York, NY, USA, 27–29 August 1968; pp. 517–524. [CrossRef]
51. Bhattachan, A.; D'Odorico, P.; Okin, G.S.; Dintwe, K. Potential dust emissions from the southern Kalahari's dunelands. *J. Geophys. Res. Earth Surf.* **2013**, *118*, 307–314. [CrossRef]
52. Sebego, R.J.; Gwebu, T.D. Patterns, determinants, impacts and policy implications of the spatial expansion of an African capital city: The greater gaborone example. *Int. J. Sustain. Built Environ.* **2013**, *2*, 193–208. [CrossRef]
53. Alsilibe, F.; Bene, K.; Bilal, G.; Alghafli, K.; Shi, X. Accuracy Assessment and Validation of Multi-Source CHIRPS Precipitation Estimates for Water Resource Management in the Barada Basin, Syria. *Remote Sens.* **2023**, *15*, 1778. [CrossRef]
54. Knoben, W.J.M.; Freer, J.E.; Woods, R.A. Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrol. Earth Syst. Sci.* **2019**, *23*, 4323–4331. [CrossRef]
55. Hart, N.C.G.; Reason, C.J.C.; Fauchereau, N. Cloud bands over southern Africa: Seasonality, contribution to rainfall variability and modulation by the MJO. *Clim. Dyn.* **2013**, *41*, 1199–1212.
56. Dezfuli, A.K. Spatiotemporal variability of seasonal rainfall in western equatorial Africa. *Theor. Appl. Climatol.* **2011**, *104*, 57–69.
57. Das, S. Assessing the Regional Concept with Sub-Sampling Approach to Identify Probability Distribution for at-Site Hydrological Frequency Analysis. *Water Resour. Manag.* **2020**, *34*, 803–817. [CrossRef]
58. Wang, Z.; Yan, J.; Zhang, X. Incorporating spatial dependence in regional frequency analysis. *Water Resour. Res.* **2014**, *50*, 9570–9585. [CrossRef] [PubMed]
59. Politis, D.N.; Romano, J.P.; Wolf, M. *Subsampling*; Springer: New York, NY, USA, 1999.
60. Hu, Y.-M.; Liang, Z.-M.; Li, B.-Q.; Yu, Z.-B. Uncertainty Assessment of Hydrological Frequency Analysis Using Bootstrap Method. *Math. Probl. Eng.* **2013**, *724632*. [CrossRef]
61. Haberle, J.; Chuchma, F.; Raimanova, I.; Wollnerova, J. Agroclimatic Zoning of Temperature Limitations for Growth of Stubble Cover Crops. *Climate* **2025**, *13*, 15. [CrossRef]