Beroho, M., Ouallali, A., El Halimi, R., Spalevic, V., Essefiani, O., El Hamdouni, Y., & Aboumaria, K. (2024). Application of statistical functions for rainfall distribution modelling and SPI calculation in Mediterranean watershed. Agriculture and Forestry, 70(1), 303-323. https://doi.org/10.17707/AgricultForest.70.1.20

DOI: 10.17707/AgricultForest.70.1.20

Mohamed BEROHO, Abdessalam OUALLALI, Rachid EL HALIMI, Velibor SPALEVIC, Oumaima ESSEFIANI, Yassir EL HAMDOUNI, Khadija ABOUMARIA [1]

APPLICATION OF STATISTICAL FUNCTIONS FOR RAINFALL DISTRIBUTION MODELLING AND SPI CALCULATION IN MEDITERRANEAN WATERSHED

SUMMARY

Drought indices are one of the most widely used methods for drought monitoring because of their ease of application and interpretation. The most commonly used drought index method is the Standardized Precipitation Index (SPI) method, which uses precipitation as an input. The performance of the Standardized Precipitation Index (SPI) is affected by the choice of an incorrect probability distribution function, which can distort index values, exaggerating or minimizing drought severity. This study aims to test the suitability of the statistical distribution functions proposed by the Bootstrap model, which estimates the closest probability distribution for calculating the SPI at time scales (TS) of 3, 6, 9, and 12 months. Daily rainfall data collected at the Dar Chaoui meteorological station were used for the period 2000-2021. Distribution function parameters were estimated using the maximum likelihood (ML) method and the Kolmogorov-Smirnov method and then confirmed using the Bootstrap method. The results show several extremely dry picks, especially in the 9- and 12-month time scales.

Keywords: Rainfall distribution modelling, SPI calculation, Drought indices, Mediterranean, Bootstrap model, Kolmogorov-Smirnov method

INTRODUCTION

Like other natural phenomena closely linked to climate change, drought is increasingly affecting all throughout the world, more than other forms of disasters

[1] Mohamed Beroho (corresponding author: simo.beroho@gmail.com), Yassir El Hamdouni, Khadija Aboumaria, Department of Earth Sciences, Faculty of Sciences and Techniques, University of Abdelmalek Essaadi, PB. 416, 90000, Tangier, MOROCCO; Rachid El Halimi, Oumaima Essefiani, Department of Mathematics and Statistics, Faculty of Sciences and Techniques, University of Abdelmalek Essaadi, PB. 416, 90000, Tangier, MOROCCO; Abdessalam Ouallali, Department of Process Engineering and Environment, Faculty of Sciences and Techniques of Mohammedia, Hassan II University of Casablanca, B 146, Mohammedia 28806, MOROCCO; Velibor Spalevic, Biotechnical Faculty, University of Montenegro, 81000 Podgorica, MONTENEGRO.

Note: The authors declare that they have no conflicts of interest. Authorship Form signed online.
Received: 19/02/2024 Accepted: 22/03/2024

(Zarei et al. 2021) and is one of the most expensive natural disasters globally. It is often described as a slowly unfolding phenomenon (Sylla et al. 2016) due to its gradual onset as a natural hazard. Delayed identifications of natural hazards, which often take longer to manifest and have an impact, typically result in delayed or more expensive reactions compared to interventions made during the initial phase following timely identification (Sestras et al., 2023; Spalevic, 2011). Drought impacts vary across regions, depending on their unique climatic characteristics and socio-economic environments, as stated by Liu et al. (2012).

In relation to climatology and meteorology, drought is characterized by a significantly prolonged and severe lack of water, which falls below normal levels, resulting in adverse consequences for plants, animals, and society (Quenum et al., 2019). Higher temperatures, increased water evaporation, and decreased vegetation cover all contribute to exacerbating the phenomenon of drought, although occasional droughts have always been a part of Earth's natural phenomena (Ojha et al., 2021; Sabri et al., 2022).

There is no single globally accepted definition of drought (Wilhite and Glantz, 1985), as drought can be analyzed and interpreted from different angles and different perceptions (Liu et al. 2018). This is typically defined based on the circumstances in each specific area.

Drought monitoring requires a variety of approaches because of differences in local rainfall, seasonal cycles, and types of rainfall. This complexity in the accurate description of the phenomenon led researchers to define drought indexes, ranging from the simplest to the most complex. These indicators enable the characterization of droughts by their intensity, duration, spatial extent, probability of recurrence (Spinoni et al., 2014), and, as highlighted by Zhang and Li (2020), their detection at various stages of evolution, including location, time of occurrence, and termination.

A variety of drought indicators is in use, including the Palmer Drought Index (PDSI: Palmer, 1965), the Standardized Precipitation and Evapotranspiration Index (SPEI: Vicente-Serrano et al., 2010), and the Standardized Precipitation Index (SPI: McKee et al., 1993). The selection of these indicators depends on the specific impact to be evaluated within the framework of monitoring and comprehending changes in vulnerability to the phenomena. The Standardized Precipitation Index (SPI) is endorsed by the World Meteorological Organization as a standard for meteorological drought characterization (Hayes et al., 2011) due to its distinct advantages. It is flexible enough to be applied across various timescales (Fotse et al., 2024). It is applicable to all climate regimes and exhibits good spatial consistency, enabling comparison across different areas subject to varying climatic conditions (Pieper et al., 2020). Due to these exceptional advantages, the index has been demonstrated to be effective in detecting various historical drought events in numerous regions worldwide (Ndayiragije et al., 2022).

Promoters of the SPI have suggested using a gamma distribution to fit cumulative precipitation in the calculation of this index, but many studies have

shown the limitations of this distribution (Touma et al. 2015; Blain et al. 2018), and researchers have shown that the applicability of theoretical distributions to describe cumulative precipitation is inconsistent across different regions and climates (Raziei 2021). For this reason, in this study, we applied the Bootstrap model to find the closest distribution to the precipitation series, and then conducted tests to decide which the appropriate distribution is. We also applied extreme value theory to find the return period over the next 100 years.

MATERIAL AND METHODS

Study area. Morocco, officially known as the Kingdom of Morocco, is situated in the north-western part of Africa within the historically significant Maghreb region (Amraoui et al., 2023; Bouayad et al., 2023). The study area is the Tangier region (Figure 1), which is one of the twelve regions of Morocco and is located in its northernmost part (Ouallali et al., 2024; Badda et al., 2023). This part is known for its rich geological and environmental diversity.

The region exhibits a diverse geological landscape, with coastal areas featuring distinctive formations influenced by maritime processes. Tangier and Tetouan boast unique geological formations shaped by coastal erosion and sedimentation. In contrast, the mountain ranges, exemplified by Chefchaouen, exhibit distinct geological features formed through tectonic activity and erosion processes over millennia. The environmental diversity of the region is primarily attributed to its varied climate. Ranging from humid Mediterranean to sub-humid, the climate of the northern region of Morocco is influenced by its geographical location and proximity to both the Mediterranean Sea and the Atlantic Ocean.

Data used. Monthly precipitation data ranging from 2000 to 2021 were obtained from the database of the Loukkos Hydraulic Basin Agency (LHBA). They are from Dar Chaoui meteorological stations located in the Tangier region in northern Morocco. The geographical positions of this station and the topography of the domain are shown in Figure 1.

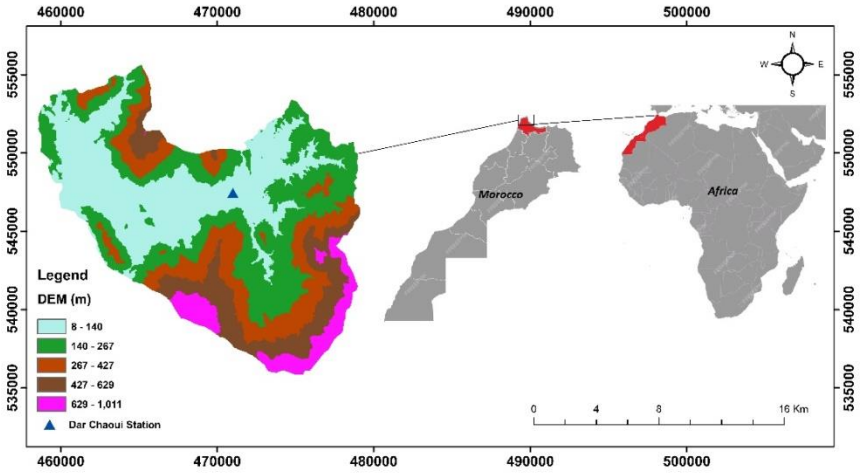

Figure 1: Study area with the geographical location of the station.

Computation of the Standardized Precipitation Index (SPI). To calculate SPI values, suitable probability density functions are fitted to frequency distributions of rainfall data. These distributions are aggregated for all selected timescales (3, 6, 9 &12 months) and subsequently transformed into standardized normal distributions (Raziei, 2021). The maximum likelihood (ML) estimation method was employed to determine the optimal parameters of distribution functions for testing purposes. Subsequently, the Kolmogorov-Smirnov (K-S) test was conducted to select the most suitable distribution from the bootstrapped functions (Raziei, 2021). The distribution with the lowest K-S statistic was identified as the best-fitting distribution. Subsequently, this distribution was utilized to construct the Cumulative Distribution Function (CDF). The CDF was transformed into normalized random variables and subsequently converted into Standardized Precipitation Index (SPI) values.

The duration of the Standardized Precipitation Index (SPI) varies depending on the specific type of drought under analysis and the intended applications (Gebremichael et al., 2022). Thus, the interpretation of the SPI indicates analyzing anomalies, which denote deviations from the average total rainfall observed within each specific period. High positive SPI values indicate excessively wet conditions, whereas high negative SPI values signify severe drought conditions. In the classification system proposed by McKee et al. (1993), various categories of drought are defined based on the SPI values, as illustrated in Table 1.

Table 1: Drought classification by SPI scores

SPI Value	Sequence of drought
SPI > 2	Extremely humid
1.5 < SPI < 1.99	Very humid
1 < SPI < 1.49	Moderately humid
-0.99 < SPI < 0.99	Near normal
-1 < SPI < -1.49	Moderately dry
-1.5 < SPI < -1.99	Very dry
SPI < -2	Extremely dry

Source: McKee et al, 1993; Cancelliere et al. 2007

The ML method, as introduced by Streit and Luginbuhl (1994), facilitates the estimation of parameters in a regression model, assuming knowledge of the true distribution law of these parameters. It involves maximizing the likelihood function, also known as the joint density function, with respect to the parameters for a given sample. The objective is to identify the parameter that has a high probability of reproducing the observed values of the sample, thereby closely matching the true values (Streit & Luginbuhl, 1994).

In simpler terms, the ML method seeks to determine the most likely value of a parameter for a population, based on a given sample (Horvath, 1993). When

applied to a dataset, it identifies the distribution parameter value that maximizes the likelihood function (Meng et al., 2014).

Random sample is $X_1,X_2,X_3,\ldots,X_n$ from a distribution $F(x;\theta_1,\theta_2,\ldots,\theta_p)$, where $\theta_1,\theta_2,\ldots,\theta_p$ are parameters of the distribution. The maximum likelihood (ML) estimators, denoted as $\hat{\theta}_1, \hat{\theta}_2, \ldots., \hat{\theta}_p$, are obtained as the solutions to the system of p equations:

$$\frac{\partial L(\theta_1,\theta_2,\ldots.\theta_p)}{\partial \theta_r} = 0 \quad (1)$$

For r = 1,2,…,p, where the likelihood function is defined as:

$$L(\theta_1, \theta_2, \ldots. \theta_p) = \prod_i^n f\left(X_i, \theta_1, \theta_2, \ldots. \theta_p\right) \quad (2)$$

Maximizing the log of the likelihood function is often preferred as it simplifies the calculations. Both methods, maximizing the probability function and maximizing its logarithm, lead to the same maximum value because the logarithm is a monotonically increasing function. Therefore, we maximize the log-likelihood function:

$$ln\ L(\theta_1, \theta_2, \ldots. \theta_p) = \prod_i^n ln\ f\left(X_i, \theta_1, \theta_2, \ldots. \theta_p\right) \quad (3)$$

This simplifies the maximization process and still yields the same maximum likelihood estimators.

the maximum likelihood (ML) method is widely regarded as an efficient estimator due to several favorable properties it exhibits.

Firstly, it typically yields estimators with lower variance compared to other methods, making it desirable for statistical inference. This property contributes to the precision and reliability of the estimates produced.

Moreover, the ML method tends to produce even more satisfactory results when applied to large datasets, particularly those with a sample size greater than 100 (n > 100). With larger sample sizes, the estimates tend to converge more closely to the true population parameters, enhancing the accuracy of the estimation process.

Furthermore, the ML estimator possesses several desirable properties of a good estimator. Firstly, it is consistent, meaning that it tends to converge to the true value of the parameter (θ) as the sample size increases. Additionally, the ML estimator is asymptotically unbiased, implying that the expected value of the estimator approaches the true parameter value as the sample size tends to infinity. Lastly, the ML estimator is asymptotically efficient, suggesting that it achieves the lowest possible variance among all consistent estimators, making it highly desirable for statistical inference (Horvath, 1993). These favorable properties contribute to the widespread use and popularity of the ML method in statistical analysis and modelling.

Bootstrap model application. The main advantage of the bootstrap resampling approach is that good estimates can be obtained regardless of the complexity of the data processing. In the context of density, the bootstrap method can be effectively used to estimate statistics such as skewness and kurtosis to explore density functions that closely represent the underlying reality of the data (Delignette-Muller and Dutang 2015). The following steps are typically taken to estimate skewness and kurtosis using Bootstrap:

Resample: Randomly sampling observations with replacement from the original data set to create multiple bootstrap samples.

Estimation: Compute skewness and kurtosis for each bootstrap sample.

Aggregation: Calculate the average skewness and kurtosis over all bootstrap samples.

By using bootstrapping in this way, obtained more reliable estimates of skewness and kurtosis help to better understand the shape and distribution of the data, leading to a more accurate density estimate that is closer to the real world scenario.

A general description of the basic principle of bootstrap methods are as follows: Suppose we are interested in estimating some parameter δ, and suppose we have observations Y_1...,Y_n from a distribution F that depends on δ.

Furthermore, we have a method for finding an estimate $\widehat{\delta}$ of δ, say $\hat{\delta}$ = T (Y_1,...,Y_n). The estimator T can be as simple as computing the skewness or kurtosis of the observations.

The main idea of bootstrapping is to replace the distribution F in the above study by the empirical distribution function $\hat{F}$.

We will show that sampling from a distribution means sampling by replacement from Y_1,..., Y_n. A bootstrapped sample has the same size as the original sampled data. It consists of the original observations, some of which may appear more than once, while others may not be included. We then apply the estimate T to each of them and obtain bootstrapping estimates $\widehat{\delta}_1$,..., $\widehat{\delta}_B$ of δ . To get an idea of the error and bias of T, or more generally of its sample distribution, we can then examine these bootstrap estimates.

Statistical distributions used to fit data

Gamma's law. Several studies have been carried out on the gamma law, and in particular (Choi and Wette, 1969) treat the gamma law in great detail. The X random variable follows a gamma distribution if its probit probability density functional (PDF) is:

$$f(x) = \frac{1}{\beta^{(\alpha)}\Gamma(\alpha)} x^{(\alpha-1)} \exp(-\frac{\alpha}{\beta}) \quad (4)$$

Proceed as follows to obtain the cumulative gamma function:

$$F(x) = \int_0^x f(x) \frac{1}{\beta^{(\alpha)}\Gamma(\alpha)} \int_0^x x^{(\alpha-1)} \exp(-\frac{\alpha}{\beta}) \quad (5)$$

With: $\alpha > 0$ is the parameter of shape

$\beta > 0$ is the parameter of scale

Γ is the mathematical gamma function

α and β are given by the ML method in the following way:

$$\begin{cases} \hat{\alpha} = \sqrt{1 + \frac{4A}{3}} \\ \hat{\beta} = \frac{-x}{\hat{\alpha}} \\ A = \ln(-x) - \frac{\Sigma \ln(x)}{n} \end{cases} \quad (6)$$

Where n is the number of years of observation. Note that this function is undefined for x=0, and its modified cumulative function has the form:

$$H(x) = q + (1 - q)F(x) \quad (7)$$

where q is the probability of zero precipitation at each station over the entire period under consideration.

Lognormal's law. If the logarithm of the random variable is normally distributed, then a positive random variable x follows a lognormal distribution. The PDF of a lognormal distribution is defined as (Mage and Ott 1984):

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} exp\left[\frac{-(\ln x - \mu)^2}{2\sigma^2}\right] \quad (8)$$

where x > 0, σ > 0 and -∞ < μ < +∞

μ is a scale parameter, stretching or shrinking a distribution, and σ2 is a shape parameter, affecting distribution shape. These can be estimated using the ML estimator method in the following way:

$$\begin{cases} \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} \ln x_i \\ \hat{\sigma}^2 = \frac{1}{n}(\sum_{i=1}^{n} \ln x_i - \hat{\mu})^2 \end{cases} \quad (9)$$

Weibull's law. According to Panahi and Asadi (2011), the PDF of a Weibull distribution for a random positive variable X is:

$$f(x, \alpha, \beta) = \alpha\beta x^{\alpha-1}\exp(-\beta x^{\alpha}) \quad (10)$$

Wu (2002) provides a detailed explanation of the shape and scale parameters derived by the ML approach mentioned above. Since there are no closed-form formulations for the parameters α and β, they are estimated by maximizing the equation's log-likelihood expression (Panahi and Asadi, 2011). Its complementary cumulative distribution function is a stretched exponential function, and its explicit form is provided by:

$$F(x) = 1 - \exp(-(\frac{x}{\alpha})^{\beta} \quad (11)$$

Gumbel's law. A random variable X is distributed according to a Gumbel law (Cooray 2010), also called a double exponential law or extreme value law, if its PDF is given by:

$$f(x) = \frac{1}{\beta} exp\left[-\exp(-\frac{x-\mu}{\beta})\right] \exp(-\frac{x-\mu}{\beta}) \quad (12)$$

With: μ > 0 is the position or mode parameter

$\beta > 0$ is the non-zero scale parameter, positive or negative

$-\infty < x < +\infty$

The ML method is used to estimate the terms μ and β. Their cumulative distribution function is given by:

$$F(x) = exp\left[-\exp(-\frac{x-\mu}{\beta})\right] \quad (13)$$

The maximum and minimum of a number of samples of normally distributed data is represented by Gumbel's law.

Exponential's law. The distribution of a random variable X is exponential if its PDF is defined as follows:

$$f(x) = \frac{1}{\beta}\frac{exp[-(x-\mu)]}{\beta} \quad (14)$$

Rahman and Pearson (2001) define $x \geq \mu$ and $\beta > 0$ as the location and scale parameters, respectively. Commonly referred to as the constant failure rate, the scaling parameter is $\lambda = \frac{1}{\beta}$. In this way, the PDF of the exponential rule can be represented as follows:

$$f(x) = \lambda\exp[-(x-\mu)]^{\lambda} \quad (15)$$

It is designed to distribute the following:

$$F(x) = 1 - exp(-(x-\mu))\,\lambda \quad (16)$$

A random and independent sample is used to estimate the parameters μ and λ. By taking the derivative of the logarithm of the likelihood function of the exponential law, the ML estimator is determined:

$$\hat{\lambda} = \frac{1}{\bar{x}}$$

Where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Logistic's law. If the PDF of a random variable X is given by (Pérez-Sánchez and Senent-Aparicio 2018), then the random variable X follows a logistic law:

$$f(x) = \frac{exp\frac{-(x-\alpha)}{\beta}}{(\alpha)(1+exp\frac{-(x-\alpha)}{\beta})^2} \quad (17)$$

$-\infty < x < +\infty$, where α is the shape parameter and β is the scale parameter that is nonzero and positive. Their cumulative distribution function is:

$$F(x) = \frac{1}{1+exp\frac{-(x-\alpha)}{\beta}} \quad (18)$$

The ML approach predicted the parameters α and β, which were used as the initial values of the program ($\alpha = 0$ and $\beta = 1$).

Burr's law. Burr's XII distribution is a continuous and widely known distribution, as it incorporates the characteristics of several well-known distributions, such as the Weibull and Gamma distributions (Pérez-Sánchez and

Senent-Aparicio 2018). A random variable X is said to follow a distribution of type XII of the Burr or Burr type if its PDF is:

$$f(x) = \frac{\alpha\gamma}{\lambda}(\frac{x}{\lambda})^{\alpha-1}(1+(\frac{x}{\lambda})^{\alpha})^{-\gamma-1} \quad (19)$$

Where:

$x > 0$

$\lambda > 0$ the scale parameter

$\alpha > 0$ the shape parameter

$\gamma > 0$ the shape parameter

ML is most often used to estimate these parameters (Ghitany and Al-Awadhi 2002). Their cumulative distribution function has the form:

$$F(x) = 1-(1+(\frac{x}{\lambda})^{\alpha})^{-\gamma} \quad (20)$$

The K-S fit test. This test is inspired by the (Kolmogorov 1933) distribution fitting statistic, as mentioned by (Stephens 1970). It is a measure of the extent that the data Xi (i=1,...n) follow a specific distributional rule. K-S-Test is a nonparametric test for comparing a sample to a reference probability distribution or for comparing two samples (Mitchell, 1971). It can be used for comparing a sample with a reference probability distribution or for comparing two samples (Mitchell, 1971). This difference is negligible, and the distribution of observations fits a pre-defined distribution according to the H_0 hypothesis. The better the law fits the data, the weaker the K-S test for a given data set and distribution. Thus, a law must have a significantly lower K-S test than the others for it to be the best. The K-S test is a measure of the difference between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution, or between the empirical distribution functions of two samples. The statistic (K-S) was defined by Stephens (1970) as follows:

$$D_n = max_x \|F_n(x) - F(x)\| \quad (21)$$

with $-\infty < x < +\infty$, and by means of the Glivenko-Cantelli theorem (Dehardt 1971):

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x)}(x_i) \quad (22)$$

With: n appears the observation parameter in population x.

$F_n(x)$ represents the empirical cumulative distribution function.

$I_{(-\infty,x)}$ is the indicator function for the event x.

$F(x)$ shows the theoretical cumulative distribution function.

Return Period Based on Extreme Value Theory. Extreme events play a crucial role in various natural processes. Knowledge of extreme events is required for the design and management of human activities in the environment. Therefore, to be able to make conclusions about extreme values of large magnitude associated with low probabilities of occurrence, the statistical modelling of extreme values is carried out. In order to refer to this type of value, the concept of a return period (T) has been introduced. The recurrence interval T

(Meylan et al. 2008) is the average length of time that, from a statistical point of view, an event of the same intensity occurs again. In hydrologic terms, the average time interval between two events of a certain intensity. There is a simple relationship between the probability of an event occurring and its return period.

Let X be the random variable associated with the precipitation series and "p" the probability of an extreme occurrence, p = P (X >=xT). The average time between two successive occurrences of the event (X =xT) is the return period T of the event. In the case of a yearly period, the return period T is related to this probability as p=1/T. Thus, the probability that an extreme event will not occur given a year is given by:

$$P\ (\mathrm{X} < x_T) = 1 - p = 1 - \frac{1}{\mathrm{T}} \quad (23)$$

The probability of the design rain not occurring for N years, the duration of our study, is:

$$P\ (\mathrm{X} < x_T) = \left(1 - \frac{1}{\mathrm{T}}\right)^{\mathrm{N}} \quad (24)$$

We define the return level zp as the distribution of this model given by the following equations:

$$GEV(z_p) = 1 - p \quad (25)$$

This will result in:

$$1 - p = \begin{cases} exp\left[-(1 + \varepsilon\frac{z_p - \mu}{\sigma})^{\frac{-1}{\varepsilon}}\right] & si\ \varepsilon \neq 0 \\ exp\left[-\exp(-\frac{z_p - \mu}{\sigma})\right] & si\ \varepsilon = 0 \end{cases} \quad (26)$$

We can then derive the expression for zp as follows:

$$z_p = \begin{cases} \mu - \frac{\sigma}{\varepsilon}[1 - \{-\ln(1-p)\}^{-\varepsilon}] & si\ \varepsilon \neq 0 \\ \mu - \sigma ln[-\ln(1-p)] & si\ \varepsilon = 0 \end{cases} \quad (27)$$

Our choice of p is small (unlikely value). By substituting the maximum likelihood estimators for the three model parameters in the formula, we obtain the maximum likelihood estimate of zp (likelihood invariance).

RESULTS

Determining Appropriate Distribution Functions

Bootstarp method. The unbiased estimation of skewness and kurtosis values is necessary for better decision making when a given observed sample is assumed to estimate the population distribution. In particular, these characteristics can be very useful to guide the choice of the most appropriate parametric distributions, since "a non-zero skewness reveals a lack of symmetry in the empirical distribution, while the kurtosis value quantifies the weight of the tails compared to the normal distribution, for which the kurtosis is equal to 3".

As a first step, we applied a bootstrap method to consider the uncertainty of the estimated values of kurtosis and skewness from the observed data. Bootstrapping is an efficient resampling technique used to estimate the variance of statistics, especially when the underlying data distribution is unknown or complex (for example, (DiCiccio and Efron 1996). The idea is to use the

observed sample to estimate the population distribution by computing the descriptive parameters of an empirical distribution, and to provide a skewness-kurtosis plot that has the square of the skewness on the x-axis and the kurtosis on the y-axis. The plot includes a point corresponding to the empirical distribution of the collected sample. It also includes bootstrapped values derived from random resampling.

To aid in the selection of distributions to fit the data, comparisons are made with values for various common distributions. For certain distributions such as normal, uniform, and logistic, where there is only one possible value for skewness (indicated by points with zero skewness) and kurtosis, these distributions are represented by distinct points on the graph. Other distributions exhibit ranges of possible values, depicted by lines (as seen with gamma and lognormal distributions) or larger areas (as observed with beta distributions). The Weibull distribution is often considered a close approximation of gamma and lognormal distributions.

In this paper, the "descdist" function of R, with boot = 1000, is employed to generate Cullen and Frey plots for the analyzed approaches, as depicted in Figure 2. Both observed and bootstrapped values exhibit notable deviations from the points representing symmetric distributions across all approaches. Consequently, to narrow our focus towards other potential distributions, we exclude symmetric distributions from consideration for fitting. Specifically, we focus on beta, log-normal, gamma, Weibull, and Burr distributions as candidate models.

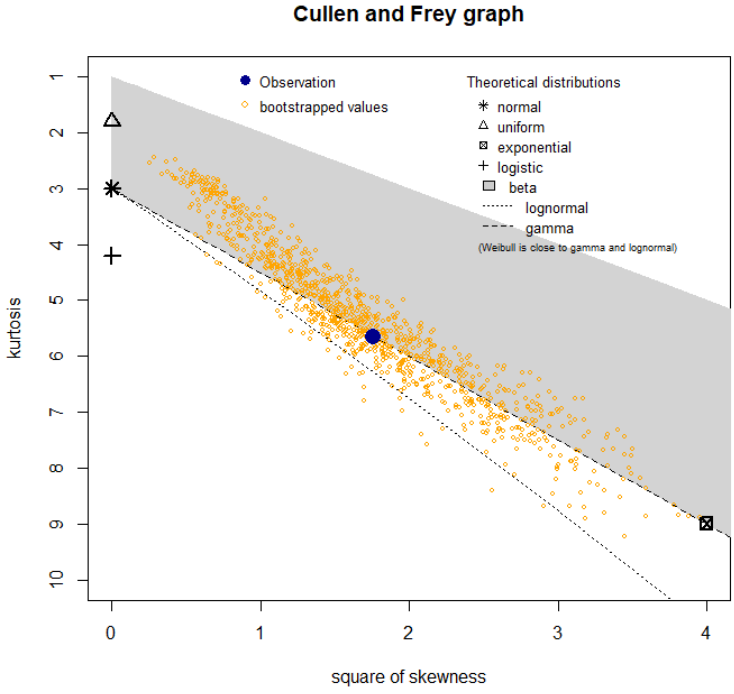

Figure 2: Determination of suitable distribution functions with Bootstrap model

Theoretical comparison. As we mentioned in the methodology section, the best fit law must have a low K-S test value, a low Akaike information criterion value, and a low Bayesian information criterion value. Figure 3 shows that the law with the lowest values of these goodness of fit tests is the Weibull law with K-S=0.03, AIC=1786.38 and BIC=1793.10; it is followed by Burr's law with K-S=0.04, AIC=1789.06 and BIC=1799.14; and Gumbel's law with K-S=0.06, AIC=1817.56 and BIC=1824.28.

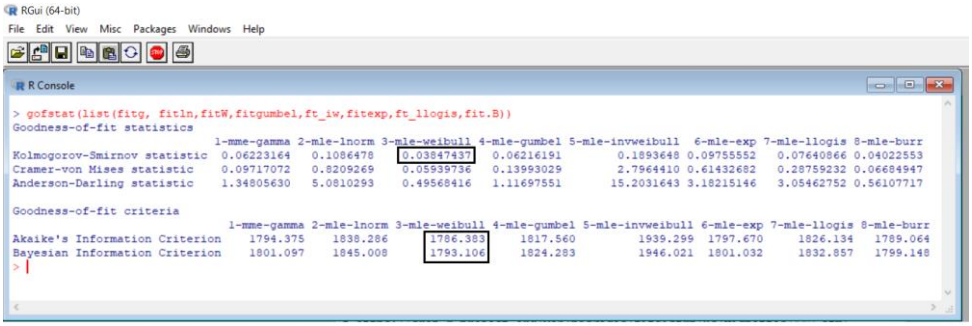

Figure 3: Theoretical determination of suitable distribution functions

Graphical comparison. Figure 4 displays a graphical comparison of all cumulative distribution functions for empirical rainfall alongside each of the test distribution functions. The results are presented for the study station and for a 12-month time series (TS). The findings corroborate those obtained from the theoretical comparison, indicating that the Weibull distribution provides the best fit, followed by the Burr and Gumbel distributions. The test comparing empirical and theoretical cumulative distribution functions (CDFs) reveals that the distributions best fitting the precipitation series and exhibiting nearly perfect alignment are Weibull's, Burr's, and Gumbel's distribution, respectively.

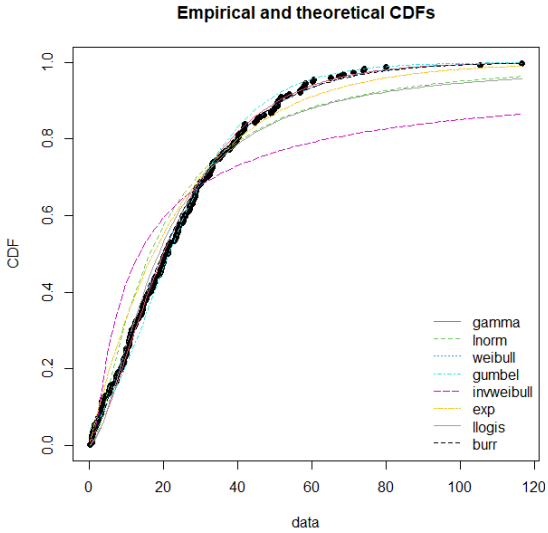

Figure 4: Graphical determination of suitable distribution functions

Weibull distribution fitting. To test hypotheses, we begin by stating a null hypothesis and an alternative hypothesis. For example, if we want to compare rates between two groups, we might say that the null hypothesizes that the rates will be the same, and the alternative hypothesizes that the rates will vary.

To test the truth of the null hypothesis, we then collect data. Specifically, the data allow us to calculate the p-value, which is defined as "the probability, under a particular statistical model, that a statistical summary of the data would be equal to or more extreme than its observed value" (Wasserstein and Lazar 2016), and is in effect a reflection of the degree of consistency of the data with the null hypothesis.

Usually, we reject the null hypothesis and accept the alternative hypothesis if the p-value is less than the 0.05 significance level. In our case, the P-value is equal to 0.89 (Figure 5), which is a very large value at 0.05. Therefore, the H0 hypothesis is accepted.

Also, the results obtained by fitting the Weibull distribution according to the equations cited in the Methodology section show that the value of shape = 1.232 and the value of scale = 26.586 (Figure 5).

From Figure 6, it is clear that the empirical and theoretical densities are very close to each other, and also that the empirical and theoretical probabilities, as well as the empirical and theoretical quantiles of the observations, are well aligned to the right. Thus, we can confirm that the Weibull distribution fits our data very well.

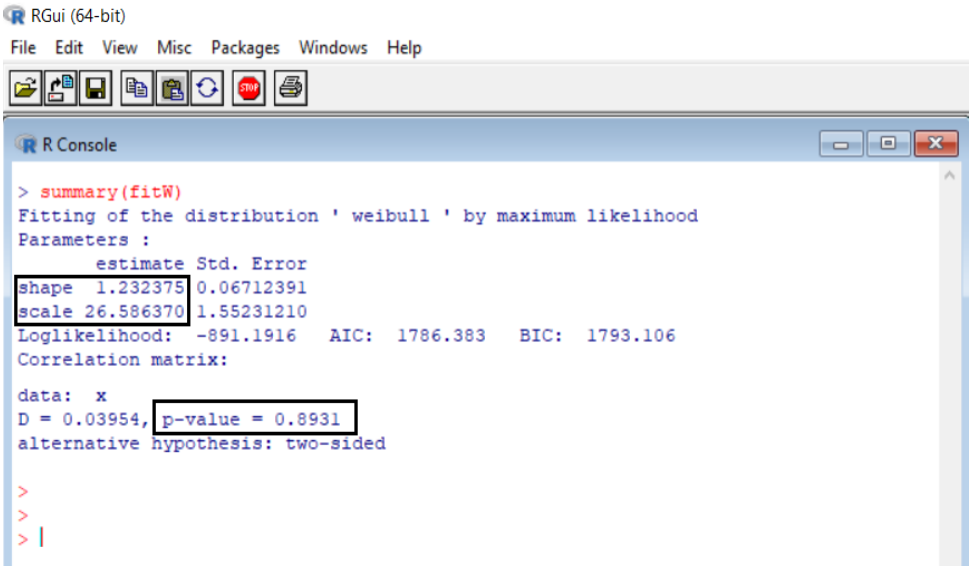

Figure 5: Summary results of Weibull's fit

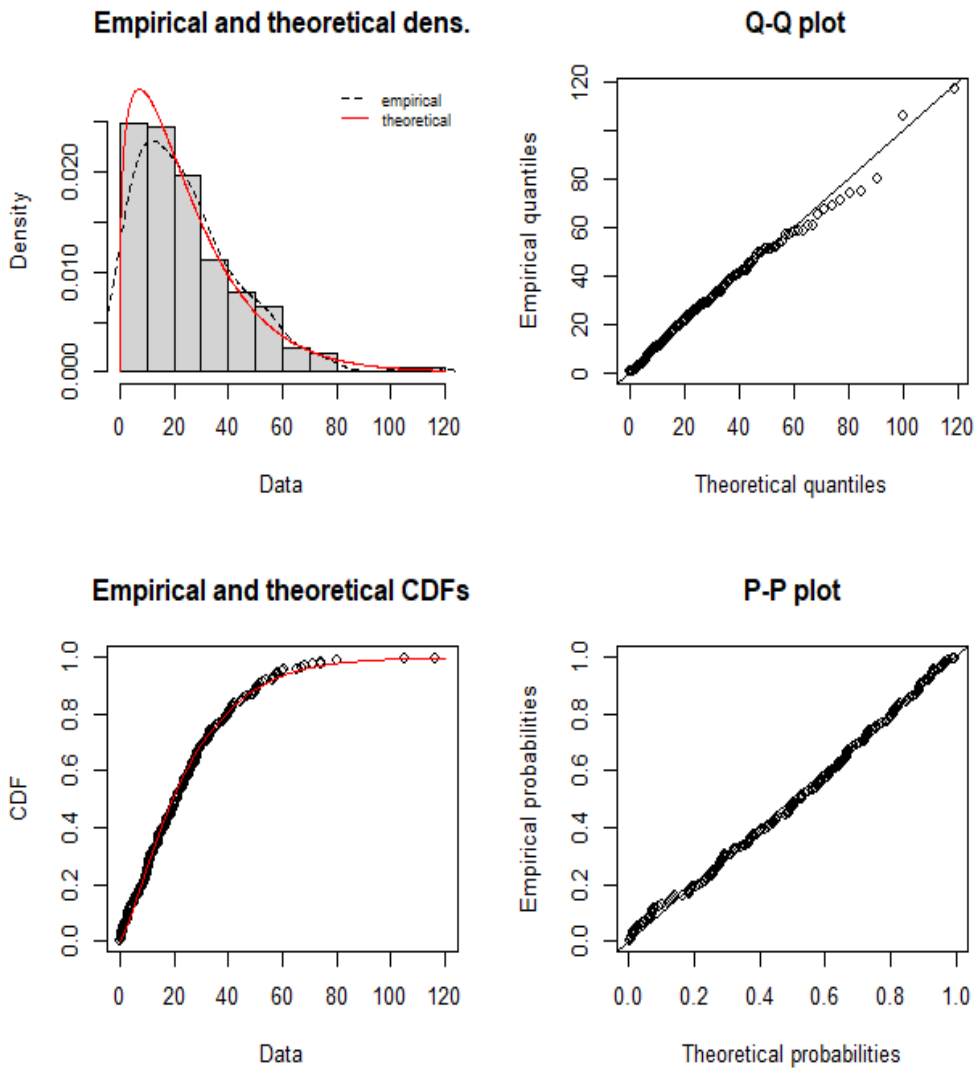

Figure 6: Empirical and theoretical CDFs of Weibull

Confirmation of Weibull Distribution Function with Bootstrap method

The results obtained by bootstrapping shape and scale values based on simulations of observed data without the intervention of distribution laws confirm the results obtained using Weibull distribution with shape value near 1.25 and scale value around 26.5 (Figure 7). Furthermore, all observations are within 95% confidence limits, which support our choice of distribution (Figure 8).

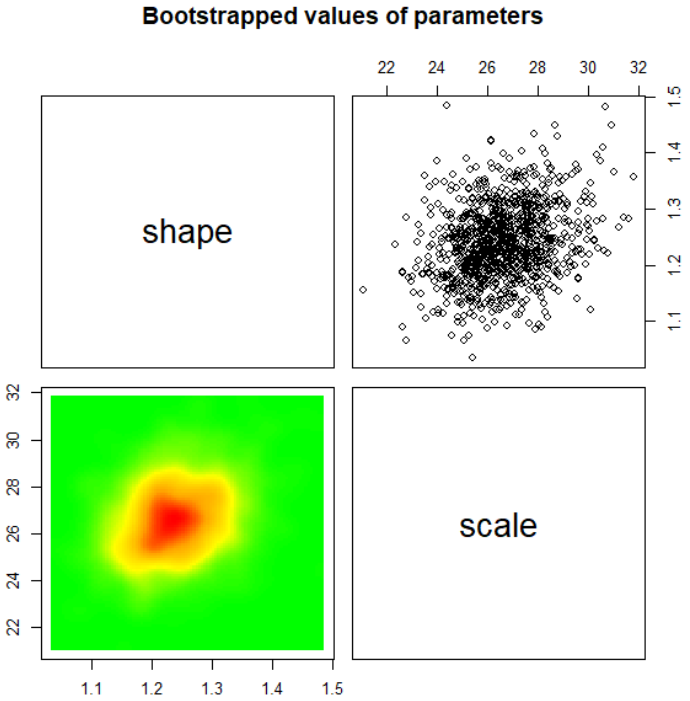

Figure 7: Shape and Scale of Weibull Distribution

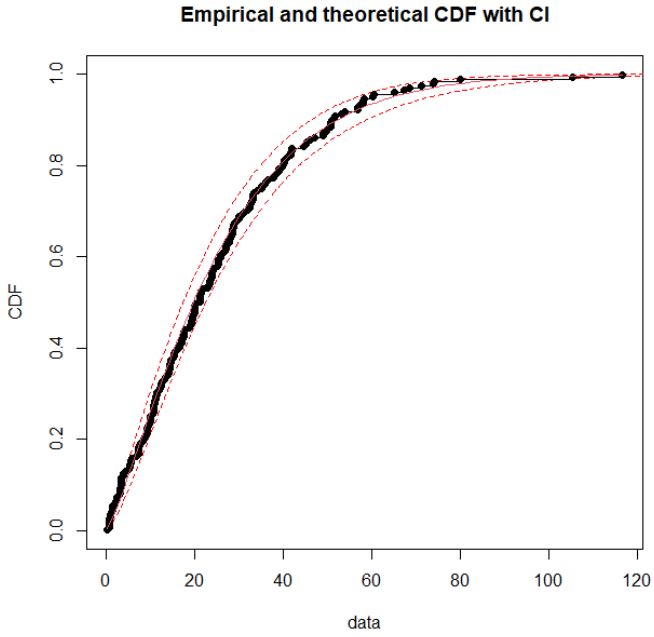

Figure 8: Weibull distribution with 95% confidence interval

Analysis of computed SPIs with adequate distributions. SPI time series were computed using the distribution most suitable for the station, and the results are illustrated in Figures 9 and 10. The 3-month SPI (Figure 9) exhibits a high frequency of drought episodes, ranging from mild to extreme. In the case of the 6-month SPI (Figure 9), there are 6 episodes of extreme drought observed. Analysis of the 9-month SPI (Figure 10) reveals 5 episodes of very severe drought and 1 episode of extreme drought recorded at the station. Furthermore, for the 12-month SPI index (Figure 10), the station experienced 3 episodes of very severe drought and 1 episode of extreme drought. Notably, the dramatic drought episodes in the years 2000 and 2021 are evident across all-time series, with an observable increase in drought duration, particularly from the 2010s onwards

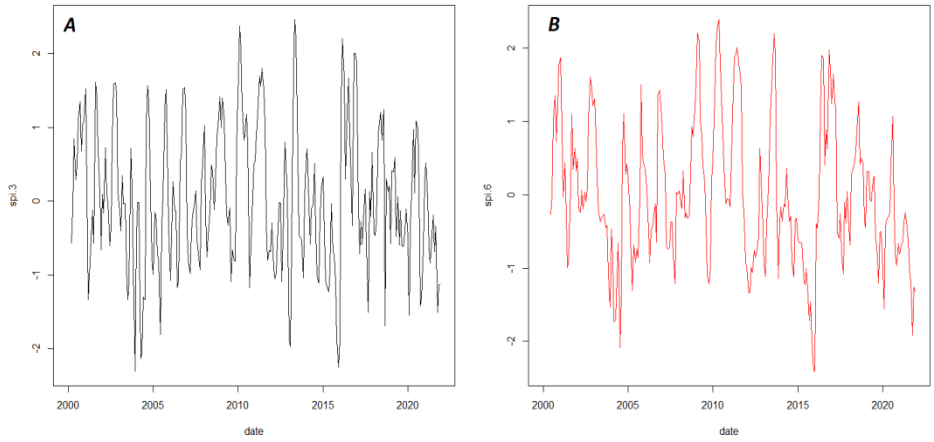

Figure 9: (A) SPIs at 3-month TS; (B) SPIs at 6-month TS

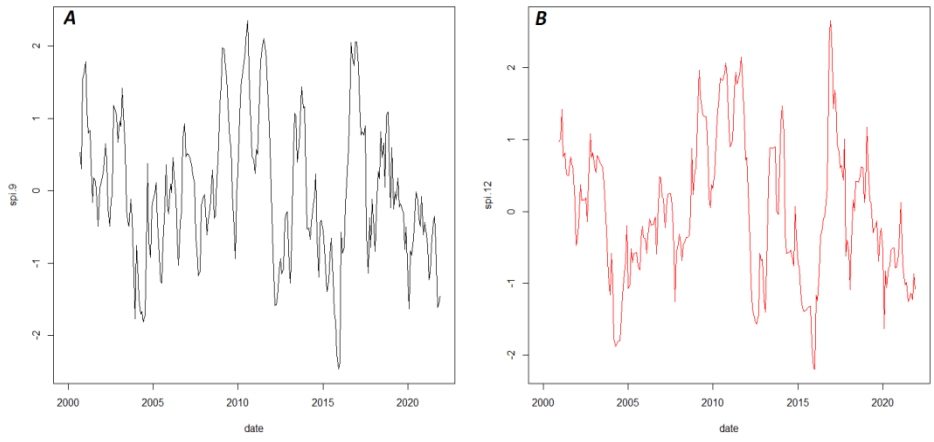

Figure 10: (A) SPIs at 9-month TS; (B) SPIs at 12-month TS

Return Period with Gumbell distribution. The station's Q-Q plot was created, and the fit of the distribution to the observed data was determined using the RMSE. The main aim of fitting the probability distribution here is to represent low-probability extreme events as accurately as possible. A Q-Q plot is used to study the level of fit of the extreme right tail (Alam et al. 2018). Any perfect match with the observed data points would fall on the [1:1] line. In Figure 11(A and B), the GEV distribution matches the data well, with the right tail close to the [1:1] line. The densities of the empirical and modelled data are very close to each other (figure 11(C)). Moreover, all observed values are within the 95% confidence interval (figure 11(B)).

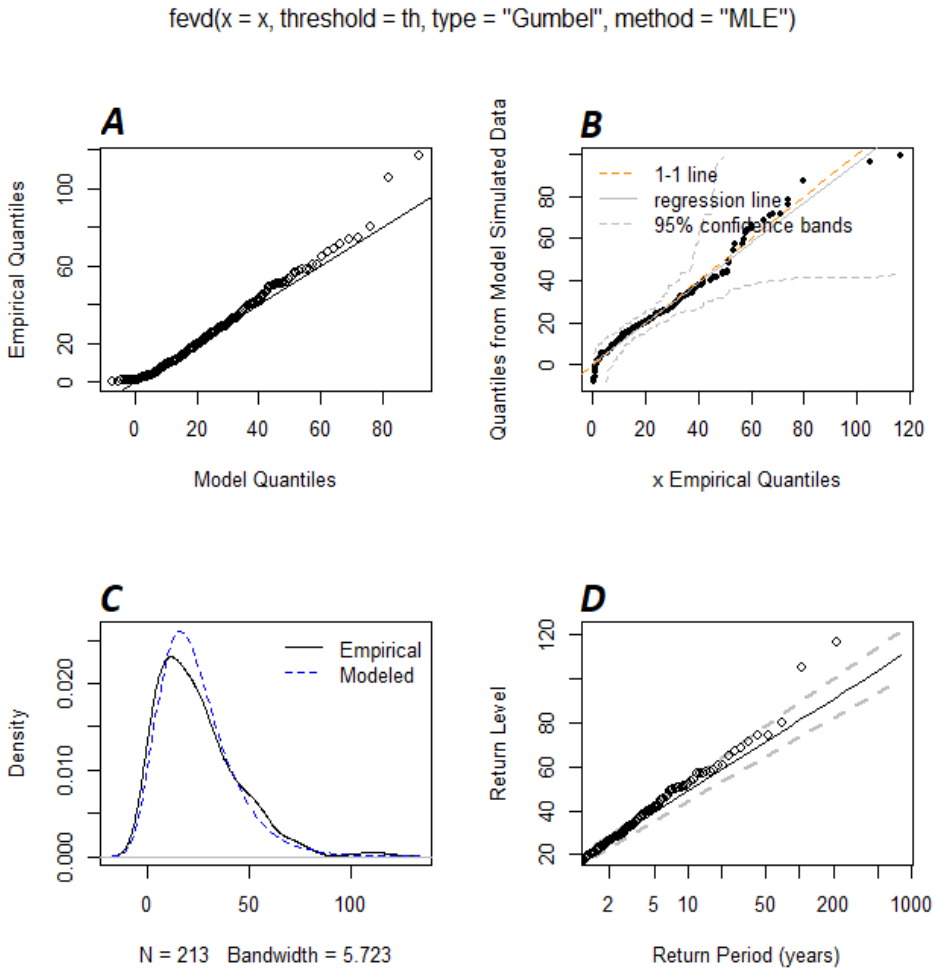

Figure 11: (A) Empirical and Model Quantiles; (B) Quantiles from model simulated data with 95% confidence interval; (C) Empirical and Modelled density; (D) Return Period.

DISCUSSION

Gamma is selected by default as the best fit without comparison with other distributions in most studies of SPI. In this study, we found that new functions (Weibull) are able to better fit the data in the station when a larger number of distribution functions are used. Therefore, depending on the geographical location of the station and the TS under consideration, the choice of the appropriate distribution function is important. These results are corroborated by our present study, which shows different distribution functions. The findings confirm those of (Fotse et al. 2024), who suggested that it is not possible to recommend a single, optimal distribution because the ratio of skewness and the coefficient of variation of the data rainfall could be the indicator for the choice of the most appropriate distribution for a particular region. Furthermore, (Angelidis et al. 2012) and (Stagge et al. 2015) considered that the appropriate probability distribution was associated with the TS of the rainfall data to be fit.

In a comparison of seven probability distributions, (Stagge et al. 2015) concluded that the gamma distribution produces the best fit for precipitation with long accumulations (> 6 months TS), while the Weibull distribution consistently performs the best for precipitation with short accumulations (1-3 months TS). In this study, weibull distribution gives best fit for rainfall with short time accumulation (3months TS), and longer time accumulation (>6months TS).

Significant variations in average precipitation patterns have a pronounced impact on the frequency and severity of droughts in the context of climate change particularly that attributed to global warming. This study shows that both intensity and duration of droughts show an increasing trend over different time scales. This observed phenomenon is likely due to reduced precipitation levels, a consequence attributed to climate change, as posited by scientific works such as those by (Beroho et al. 2020) and (Boulahfa et al. 2023). Specifically, temperature emerges as a key determinant of water availability dynamics, mainly by regulating evapotranspiration rates.

CONCLUSION

This research contributes to the improvement of mathematical methods for drought modelling, which is particularly relevant given its hazardous nature and the challenges associated with adaptation, especially in developing countries such as Morocco. In this context, the Standardized Precipitation Index (SPI) serves as a central drought indicator, prompting this study to investigate the efficacy of using alternative probability distribution functions to fit and characterize observed precipitation data - a crucial initial stage in the SPI calculation.

In this study, eight different statistical distribution functions were examined to determine the optimal fit for data from the Tangier region station, spanning the 2000-2021-time domain, at different time scales (TS) of 3, 6, 9, and 12 months. The Maximum Likelihood (ML) method was used to estimate the parameters of these distribution functions. The Kolmogorov-Smirnov (K-S) statistic served as a discriminating metric to identify the distribution functions

that best fit the observed station data, which were subsequently used in the SPI calculation. The results of this analysis were used to identify patterns of drought occurrence and to quantify discrepancies resulting from the use of mismatched distribution functions. The choice of an optimal distribution function for precipitation data depends on both the geographic location of the station and the temporal scope of the analysis, as defined by the number of months in the time series (TS). In particular, the Weibull probability distribution consistently demonstrated superior performance across all TS durations.

This investigation underscores the importance of conducting a careful preliminary assessment aimed at identifying the most appropriate distribution functions for data fitting, and then using them in the SPI calculation. Such an approach is critical to reducing error and improving the accuracy of results in drought modelling and assessment.

REFERENCES

Alam, M., Emura, K., Farnham, C., Yuan, J. (2018) Best-Fit Probability Distributions and Return Periods for Maximum Monthly Rainfall in Bangladesh. Climate 6:9. https://doi.org/10.3390/cli6010009

Amraoui, M., Mijanovic, D., El Amrani, M., Kader, S., & Ouakhir, H. (2023). Agriculture and economic development of the Ait Werra tribe during the French colonialism period and its local characteristics (1912-1956) within the Middle Atlas region of Morocco. Agriculture and Forestry, 69(4), 91-112. https://doi.org/10.17707/AgricultForest.69.4.07

Angelidis, P., Maris, F., Kotsovinos, N., & Hrissanthou, V. (2012). Computation of Drought Index SPI with Alternative Distribution Functions. Water Resources Management, 26, 2453–2473. https://doi.org/10.1007/s11269-012-0026-0

Badda, H., Cherif, E. K., Boulaassal, H., Wahbi, M., Yazidi Alaoui, O., Maatouk, M., Bernardino, A., Coren, F., & El Kharki, O. (2023). Improving the Accuracy of Random Forest Classifier for Identifying Burned Areas in the Tangier-Tetouan-Al Hoceima Region Using Google Earth Engine. Remote Sensing, 15, 4226. https://doi.org/10.3390/rs15174226

Beroho, M., Briak, H., El Halimi, R., Ouallali, A., Boulahfa, I., Mrabet, R., Kebede, F., & Aboumaria, K. (2020). Analysis and prediction of climate forecasts in Northern Morocco: application of multilevel linear mixed effects models using R software. Heliyon, 6(10), e05094. https://doi.org/10.1016/j.heliyon.2020.e05094

Blain, G. C., de Avila, A. M. H., & Pereira, V. R. (2018). Using the normality assumption to calculate probability-based standardized drought indices: selection criteria with emphases on typical events. International Journal of Climatology, 38, e418-e436. https://doi.org/10.1002/joc.5381

Bouayad, F. E., El Idrysy, M., Ouallali, A., El Amrani, M., Courba, S., Hahou, Y., Benhachmi, M. K., Spalevic, V., Kebede, F., & Briak, H. (2023). Assessing soil erosion dynamics in the Rmel watershed, northern Morocco by using the RUSLE model, GIS, and remote sensing integration. Agriculture and Forestry, 69(4), 173-194. doi:10.17707/AgricultForest.69.4.11

Boulahfa, I., ElKharrim, M., Naoum, M., et al. (2023). Assessment of performance of the regional climate model (RegCM4.6) to simulate winter rainfall in the north of Morocco: The case of Tangier-Tétouan-Al-Hociema Region. Heliyon, 9, e17473. https://doi.org/10.1016/j.heliyon.2023.e17473

Cancelliere, A., Mauro, D. G., Bonaccorso, B., & Rossi, G. (2007). Drought forecasting using the Standardized Precipitation Index. Water Resources Management, 21, 801–819. https://doi.org/10.1007/s11269-006-9062-y

Cooray, K. (2010). Generalized Gumbel distribution. Journal of Applied Statistics, 37, 171–179. https://doi.org/10.1080/02664760802698995

Cullen, A. C., & Frey, H. C. (1999). Probabilistic techniques in exposure assessment: A handbook for dealing with variability and uncertainty in models and inputs. Plenum Press, New York.

Dehardt, J. (1971). Generalizations of the Glivenko-Cantelli Theorem. Annals of Mathematical Statistics, 42, 2050–2055. https://doi.org/10.1214/aoms/1177693073

Delignette-Muller, M. L., & Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. Journal of Statistical Software, 64(4), 1–34. https://doi.org/10.18637/jss.v064.i04

DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. Statistical Science, 11(3), 189-228. https://doi.org/10.1214/ss/1032280214

Fotse, A. R. G., Guenang, G. M., Mbienda, A. J. K., & Vondou, D. A. (2024). Appropriate statistical rainfall distribution models for the computation of standardized precipitation index (SPI) in Cameroon. Earth Science Informatics, 17, 725–744. https://doi.org/10.1007/s12145-023-01188-0

Gebremichael, H. B., Raba, G. A., Beketie, K. T., & Feyisa, G. L. (2022). Temporal and spatial characteristics of drought, future changes and possible drivers over Upper Awash Basin, Ethiopia, using SPI and SPEI. Environment, Development and Sustainability, 26, 947–985. https://doi.org/10.1007/s10668-022-02743-3

Ghitany, M. E., & Al-Awadhi, S. (2002). Maximum likelihood estimation of Burr XII distribution parameters under random censoring. Journal of Applied Statistics, 29, 955–965. https://doi.org/10.1080/0266476022000006667

Hayes, M., Svoboda, M., Wall, N., & Widhalm, M. (2011). The Lincoln Declaration on Drought Indices: Universal Meteorological Drought Index Recommended. Bulletin of the American Meteorological Society, 92, 485–488. https://doi.org/10.1175/2010BAMS3103.1

Horvath, L. (1993). The Maximum Likelihood Method for Testing Changes in the Parameters of Normal Observations. Annals of Statistics, 21(2), 671-680. https://doi.org/10.1214/aos/1176349143

Liu, D., You, J., Xie, Q., et al. (2018). Spatial and Temporal Characteristics of Drought and Flood in Quanzhou Based on Standardized Precipitation Index (SPI) in Recent 55 Years. Global Environmental Protection, 6, 25–37. https://doi.org/10.4236/gep.2018.68003

Liu, L., Hong, Y., Bednarczyk, C. N., et al. (2012). Hydro-Climatological Drought Analyses and Projections Using Meteorological and Hydrological Drought Indices: A Case Study in Blue River Basin, Oklahoma. Water Resources Management, 26, 2761–2779. https://doi.org/10.1007/s11269-012-0044-y

Mage, D. T., & Ott, W. R. (1984). An evaluation of the methods of fractiles, moments and maximum likelihood for estimating parameters when sampling air quality data from a stationary lognormal distribution. Atmospheric Environment (1967), 18, 163–171. https://doi.org/10.1016/0004-6981(84)90239-7

Meng, X. P., Zhao, C. Q., & Huo, L. (2014). Maximum Likelihood Method for Parameter Estimation of Weibull Distribution Model Based on Fruit Fly Optimization Algorithm. Applied Mechanics and Materials, 602-605, 3508-3511. https://doi.org/10.4028/www.scientific.net/AMM.602-605.3508

Meylan, P., Favre, A.-C., & Musy, A. (2008). Hydrologie fréquentielle: Une science prédictive (pp. 1-184). Collection: Science et ingénierie de l'environnement. PPUR presses polytechniques.

Ndayiragije, J. M., Li, F., & Nkunzimana, A. (2022). Assessment of Two Drought Indices to Quantify and Characterize Drought Incidents: A Case Study of the Northern Part of Burundi. Atmosphere, 13, 1882. https://doi.org/10.3390/atmos13111882

Ojha, S. S., Singh, V., & Roshni, T. (2021). Comparison of Meteorological Drought using SPI and SPEI. Civil Engineering Journal, 7, 2130–2149.

Ouallali, A., Kader, S., Bammou, Y., Aqnouy, M., Courba, S., Beroho, M., Briak, H., Spalevic, V., Kuriqi, A., & Hysa, A. (2024). Assessment of the Erosion and

Outflow Intensity in the Rif Region under Different Land Use and Land Cover Scenarios. Land, 13, 141. https://doi.org/10.3390/land13020141

Pérez-Sánchez, J., & Senent-Aparicio, J. (2018). Analysis of meteorological droughts and dry spells in semiarid regions: A comparative analysis of probability distribution functions in the Segura Basin (SE Spain). Theoretical and Applied Climatology, 133, 1061–1074. https://doi.org/10.1007/s00704-017-2239-x

Pieper, P., Düsterhus, A., & Baehr, J. (2020). A universal Standardized Precipitation Index candidate distribution function for observations and simulations. Hydrology and Earth System Sciences, 24, 4541–4565.

Quenum, G. M. L. D., Klutse, N. A. B., Dieng, D., et al. (2019). Identification of Potential Drought Areas in West Africa Under Climate Change and Variability. Earth Systems and Environment, 3, 429–444.

Raziei, T. (2021). Performance evaluation of different probability distribution functions for computing Standardized Precipitation Index over diverse climates of Iran. Int. J. Climatol., 41, 3352–3373. https://doi.org/10.1002/joc.7023

Sabri, E., Spalevic, V., Boukdir, A., Karaoui, I., Ouallali, A., Mincato, R. L., & Sestras, P. (2022). Estimation of soil losses and reservoir sedimentation: a case study in Tillouguite sub-basin (High Atlas-Morocco). Agriculture and Forestry 68 (2), 207-220. doi:10.17707/AgricultForest.68.2.15

Sestras, P., Mircea, S., Cîmpeanu, S. M., Teodorescu, R., Roșca, S., Bilașco, Ș., Rusu, T., Salagean, T., Dragomir, L. O., Marković, R., et al. (2023). Soil Erosion Assessment Using the Intensity of Erosion and Outflow Model by Estimating Sediment Yield: Case Study in River Basins with Different Characteristics from Cluj County, Romania. Applied Sciences, 13(16), 9481.

Spinoni, J., Naumann, G., Carrao, H., et al. (2014). World drought frequency, duration, and severity for 1951-2010: World drought Climatology for 1951-2010. International J. Climatol., 34, 2792–2804. https://doi.org/10.1002/joc.3875

Stagge, J. H., Tallaksen, L. M., Gudmundsson, L., Van Loon, A. F., & Stahl, K. (2015). Candidate Distributions for Climatological Drought Indices (SPI and SPEI). Int. J. Climatol., 35, 4027–4040. https://doi.org/10.1002/joc.4267

Stephens, M. A. (1970). Use of the Kolmogorov–Smirnov, Cramér–Von Mises and Related Statistics Without Extensive Tables. Journal of the Royal Statistical Society: Series B (Methodological), 32, 115–122.

Streit, R. L., & Luginbuhl, T. E. (1994). Maximum likelihood method for probabilistic multihypothesis tracking. In O. E. Drummond (Ed.), Orlando, FL, pp. 394–405.

Sylla, M. B., Nikiema, P. M., Gibba, P., et al. (2016). Climate Change over West Africa: Recent Trends and Future Projections. In J. A. Yaro & J. Hesselberg (Eds.), Adaptation to Climate Change and Variability in Rural West Africa (pp. 25–40). Springer International Publishing.

Touma, D., Ashfaq, M., Nayak, M. A., et al. (2015). A multi-model and multi-index evaluation of drought characteristics in the 21st century. Journal of Hydrology, 526, 196–207. https://doi.org/10.1016/j.jhydrol.2014.12.011

Vicente-Serrano SM, Beguería S, López-Moreno JI (2010) A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. Journal of Climate 23:1696–1718. https://doi.org/10.1175/2009JCLI2909.1

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p -Values: Context, Process, and Purpose. The American Statistician, 70, 129–133. https://doi.org/10.1080/00031305.2016.1154108Wilhite DA, Glantz MH (1985) Understanding: the Drought Phenomenon: The Role of Definitions. Water International 10:111–120. https://doi.org/10.1080/02508068508686328

Zarei, A. R., Shabani, A., & Moghimi, M. M. (2021). Accuracy Assessment of the SPEI, RDI and SPI Drought Indices in Regions of Iran with Different Climate Conditions. Pure and Applied Geophysics, 178, 1387–1403.

Zhang, Y., & Li, Z. (2020). Uncertainty Analysis of Standardized Precipitation Index Due to the Effects of Probability Distributions and Parameter Errors. Frontiers in Earth Science, 8, 76. https://doi.org/10.3389/feart.2020.00076