Conditional Copula-Based Spatial–Temporal Drought Characteristics Analysis—A Case Study over Turkey

Mahdi Hesami Afshar [1,*], Ali Unal Sorman [2] and Mustafa Tugrul Yilmaz [1]

[1]   Department of Civil Engineering, Middle East Technical University, Ankara 06800, Turkey;
      tuyilmaz@metu.edu.tr
[2]   Department of Civil Engineering, Near East University, Mersin 99138, Turkish Republic of North Cyprus;
      ausorman@gmail.com
*   Correspondence: afshar.mahdi@metu.edu.tr; Tel.: +90-312-210-5413

Abstract: In this study, commonly used copula functions belonging to Archimedean and Elliptical families are fitted to the univariate cumulative distribution functions (CDF) of the drought characteristics duration (LD), average severity ($\bar{S}$), and average areal extent ($\bar{A}$) of droughts obtained using standardized precipitation index (SPI) between 1960 and 2013 over Ankara, Turkey. Probabilistic modeling of drought characteristics with seven different fitted copula functions and their comparisons with independently estimated empirical joint distributions show normal copula links drought characteristics better than other copula functions. On average, droughts occur with an average LD of 6.9 months, $\bar{S}$ of 0.94, and $\bar{A}$ of 73%, while such a drought event happens on average once in every 6.65 years. Results also show a very strong and statistically significant relation between $\bar{S}$ and $\bar{A}$, and drought return periods are more sensitive to the unconditioned drought characteristic, while return periods decrease by adding additional variables to the analysis (i.e., trivariate drought analysis compared to bivariate).

Keywords: copula; SPI; drought; multivariate; spatial; temporal

1. Introduction

Droughts are a climatic phenomenon that may impact large and small regions alike for long or short time periods and influence almost all aspects of society [1]. Historically, droughts have had negative effects on domestic and agricultural water supply needs [2] and consequently altered the behavior of ecosystems [3,4]. Plans, designs, and operational applications for water resources systems, therefore, require accurate estimation of different drought characteristics, like duration (LD), average severity ($\bar{S}$), average areal extent ($\bar{A}$), and re-occurrence rate under various scenarios.

Drought characteristics have been analyzed using various methods that have evolved over time. Earlier studies investigated the analytical or stochastic nature of past droughts using runs theory. Based on this theory, observed time series of drought-related variables are divided into wet and dry periods (positive and negative runs, respectively; see Figure 1) depending on a given threshold or critical level [5,6]. Later studies focused more on probabilistic drought estimation using ground station-based datasets [7–9]. The scarcity of ground stations and short recording lengths motivated the following studies to use synthetic datasets to analyze the frequencies of drought events [10,11]. Since the spatial–temporal relationships among drought characteristics are complex, recent studies have diversified their efforts by analyzing the return periods of droughts using various $\bar{S}$ or $\bar{A}$ scenarios [12–14].
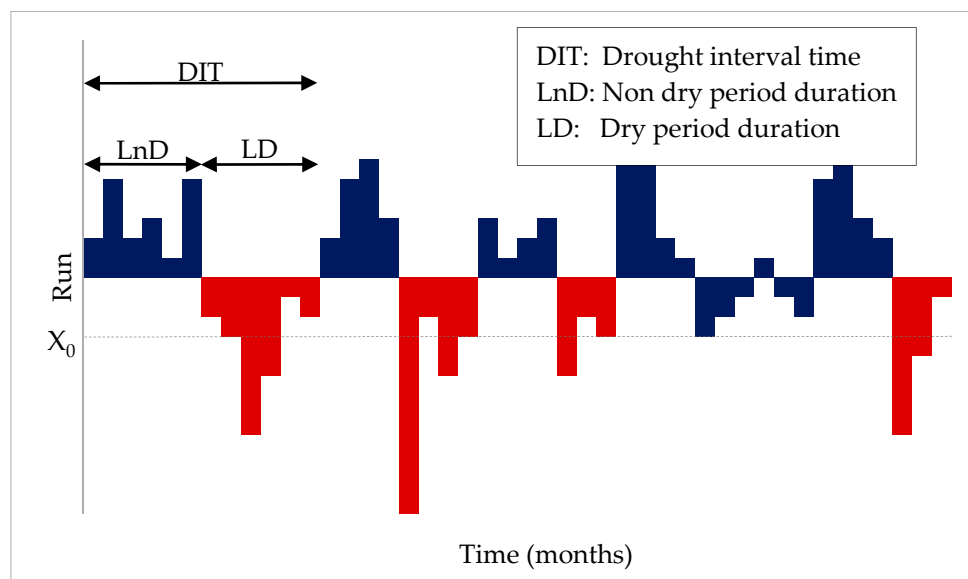
Figure 1. Illustration of drought characteristics using the run theory for a given threshold level of X_0 where the drought events are presented with the red bars over time.

The copula approach [15] provides an ideal test bed in which to analyze multivariate probabilistic problems without making explicit assumptions about the marginal or joint distributions of the variables involved in the estimation problem. Perhaps this ability has motivated studies to implement a multivariate probabilistic approach via copula functions to solve many different problems in hydrologic science and water resources-related applications [16–19].

Given that drought return periods can be immediately formulated in a probabilistic framework, copula-related methods have the necessary utility required to investigate different drought characteristics. Shiau [13] used bivariate copula functions to investigate bivariate drought analysis, including joint probabilities and return periods. Song and Singh [20] expanded this drought frequency analysis from bivariate to trivariate via a new method that constructs trivariate copulas to describe the joint distribution function of the temporal characteristics of meteorological drought events. Recently, Xu et al. [14] used a trivariate drought identification method within the copula approach to calculate the joint probability and return period of different drought spatial–temporal characteristic pairs. Overall, these studies demonstrated the advantages of the copula approach on bivariate and trivariate modeling of drought characteristics.

Certain characteristics of drought may precede others. For example, droughts with longer durations may tend to have higher maximum severity, or it may take longer to dissipate severe droughts than droughts with moderate severity. Such dependencies between drought characteristics can be immediately studied in a conditional copula framework, particularly when the occurrence of certain characteristics signals the occurrence of other characteristic functions in multivariate problems. Even though joint copula functions have been implemented in many studies, the implementation of conditional copulas remained limited in the literature [14,21]; hence there is still more room for more conditional copula-based applications to fully comprehend the utility of these functions.

Drought characteristics could be profoundly linked both in time and space. Accordingly, knowledge (i.e., observations) of some drought characteristics in time or space may help us predict others. This temporal and/or spatial dependency between different drought characteristics may immediately be exploited using copula functions, particularly using conditional probabilities. Even though spatial–temporal dependencies of drought characteristics have been investigated before using copula functions [22–24], such investigations have not been carried out in a conditional framework before.

The main objective of this study is to implement a copula-based methodology to analyze the joint and conditional dependency among various spatial–temporal characteristics of droughts, including LD, $\overline{S}$, and $\overline{A}$. The analysis utilizes univariate cumulative distribution functions (CDF) of drought characteristics obtained over Ankara, Turkey to establish joint and conditional relations between them, using bivariate and trivariate copula functions to calculate expected return periods of droughts with different characteristics scenarios.

2. Materials and Methods

In the following section, copula functions and their application to probabilistic drought characterizations are explained; later the use of bivariate and trivariate functions (i.e., the use of return periods) in drought analysis is elaborated.

2.1. Copula Functions

Copula functions link the univariate CDFs of random variables with their joint CDFs [13,15]. Estimations of these univariate CDFs are trivial and the sample datasets are often sufficient, while the analytical solution for the joint CDFs is not immediately clear. At this point, copula functions link univariate CDFs and create their multivariate (e.g., bivariate or trivariate) CDFs.

Based on Sklar's theorem, bivariate and trivariate joint CDFs denoted as $F(u, v)$ and $F(u, v, w)$, respectively, can be obtained in the form:

$$F(u, v) = C(F(u), F(v)), \tag{1}$$

$$F(u, v, w) = C(F(u), F(v), F(w)), \tag{2}$$

where u, v, and w are random variables (e.g., LD, $\overline{S}$, and $\overline{A}$); C( ) is the copula function; and univariate CDFs $F(u)$, $F(v)$, and $F(w)$ are inputs to these function. Here $F(u)$, $F(v)$, and $F(w)$ are the univariate CDFs of u, v, and w respectively, and they are defined as

$$F(u) = P(U \leq u), \tag{3}$$

$$F(v) = P(V \leq v), \tag{4}$$

$$F(w) = P(W \leq w), \tag{5}$$

where $P(U \leq u)$, $P(V \leq v)$, and $P(W \leq w)$ are the probabilities that random variables U, V, and W (i.e., different drought characteristics) takes values smaller than u, v, and w.

In Equations (1) and (2), the copula functions link the univariate CDF to bivariate and trivariate joint CDFs ($F(u, v)$ and $F(u, v, w)$). Using these joint CDFs, bivariate and trivariate conditional CDFs can be found by utilizing the multiplication rule:

$$F(u|v) = \frac{F(u, v)}{F(v)}, \tag{6}$$

$$F(u, v|w) = \frac{F(u, v, w)}{F(w)}. \tag{7}$$

In Equation (7), $F(u, v|w)$ is given as an example of the trivariate CDF, while many other trivariate CDF (e.g., $F(u|v, w)$, $F(v|u, w)$, etc.) can be found in a similar fashion.

In this study, the bivariate (Equation (1)) and trivariate (Equation (2)) joint CDF of LD, $\overline{S}$, and $\overline{A}$ are obtained using seven widely used copula functions belonging to Archimedean and Elliptical copula families. These copula functions and their parameter spaces are shown in Table 1. For further details about these copula functions, see the studies of Joe [25], Shiau [13], and Nelsen [26].

The required copula dependence parameters (θ and ρ) for the seven copula functions are separately estimated for each copula function using the inference function for margins (IFM) method [25]. This method involves estimating F(u) and F(v) for the variables and then estimating the θ or ρ parameter that maximizes the Kendall's tau correlation using these univariate CDFs. These obtained parameters are later used to make copula function-based joint CDF predictions, which are referred to as "theoretical joint CDF estimation" $(C(u,v)_t)$ in this study.

Table 1. Equations of Copula functions. Here u and v are two dependent univariate variables, df is the degree of freedom, θ and ρ are Copula dependence parameters, $\varnothing$ is the CDF of standard univariate Gaussian distribution, and t_{df} is the t–student distribution function. AMH and GH refer to the Ali–Mikhail–Haq and Gumbel–Hougaard distributions, respectively.

Function (Family)	Joint Cumulative Distribution Function, $C(u,v)$	Parameter Range
AMH (Archimedean)	$\frac{uv}{1-\theta(1-u)(1-v)}$	$-1 \le \theta \le 1$
Clayton (Archimedean)	$[\max(0, u^{-\theta} + v^{-\theta} - 1)]^{-1/\theta}$	$0 \le \theta$
Frank (Archimedean)	$\frac{-1}{\theta}\ln(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1})$	$\theta \neq 0$
GH (Archimedean)	$e^{[(-\ln u)^{\theta} + (-\ln v)^{\theta}]^{\frac{1}{\theta}}}$	$1 \le \theta$
Joe (Archimedean)	$1 - ((1-u)^{\theta} + (1-v)^{\theta} - (1-u)^{\theta}(1-v)^{\theta})^{\frac{1}{\theta}}$	$1 \le \theta$
Normal (Elliptical)	$\int_{-\infty}^{\varnothing^{-1}(u)} \int_{-\infty}^{\varnothing^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} e^{-\frac{u^2 - 2\rho uv + v^2}{2(1-\rho^2)}} \, du\,dv$	$-1 \le \rho \le 1$
t–student (Elliptical)	$\int_{-\infty}^{t_{df}^{-1}(u)} \int_{-\infty}^{t_{df}^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} (1 + \frac{u^2 - 2\rho uv + v^2}{df(1-\rho^2)})^{-\frac{df+2}{2}} \, du\,dv$	$-1 \le \rho \le 1$ $df \ge 1$

The above described copula function-based $C(u,v)_t$ values are later independently validated using "empirical joint CDF estimates" $(C(u,v)_e)$ obtained using rank-based joint distribution estimates obtained via the copula package [27] in the R programming language [28]. Datasets are split for parameter estimation and validation with the leave-one-out cross-validation method to avoid over-fitting. Validation efforts include calculation of root mean square error (RMSE) using $C(u,v)_t$ and $C(u,v)_e$:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (C(u,v)_t - C(u,v)_e)^2}{n}}, \tag{8}$$

where n is the length of the drought characteristic datasets. Among the seven copula functions, the one that yields the lowest RMSE value is later selected for further estimation of bivariate and trivariate return periods.

2.2. Return Periods

The expected value of the drought inter-arrival time (DIT; the time between two successive drought events) is one of the most commonly estimated drought characteristics in hydrological and water resources systems applications. T may be estimated for univariate (e.g., LD ≥ 10 months) as well as multivariate (e.g., LD ≥ 10 & $\bar{S}$ ≥ 1.0; or LD ≥ 10 | $\bar{S}$ ≥ 1.0) scenarios. Shiau and Shen [29] theoretically derived T for drought events with univariate drought characteristics (e.g., LD, $\bar{S}$, and $\bar{A}$) as a function of the expected DIT in the form

$$T_{U \geq u} = \frac{E(DIT)}{P(U \geq u)}, \tag{9}$$

where $E(DIT)$ is the expected value of DIT; and $T_{U \geq u}$ is the return period of a drought event, defined by the variable u. $E(DIT)$ is calculated as the ratio of the duration of entire study to the total number of droughts. Following Shiau [13], it is also possible to expand this univariate T estimation analysis to a bivariate case by characterizing events with bivariate joint or conditional probabilities in the forms:

$$T_{(U \geq u, V \geq v)} = \frac{E(DIT)}{P(U \geq u, V \geq v)} \tag{10}$$

$$T_{(U \geq u \text{ or } V \geq v)} = \frac{E(DIT)}{P(U \geq u \text{ or } V \geq v)} \tag{11}$$

$$T_{(U \geq u | V \geq v)} = \frac{E(DIT)}{P(U \geq u, | V \geq v)}. \tag{12}$$

Similarly, the trivariate case can be obtained in the forms:

$$T_{(U \geq u, V \geq v, W \geq w)} = \frac{E(DIT)}{P(U \geq u, V \geq v, W \geq w)} \tag{13}$$

$$T_{(U \geq u \text{ or } V \geq v \text{ or } W \geq w)} = \frac{E(DIT)}{P(U \geq u \text{ or } V \geq v \text{ or } W \geq w)} \tag{14}$$

$$T_{(U \geq u | (V \geq v, W \geq w))} = \frac{E(DIT)}{P(U \geq u | (V \geq v, W \geq w))}, \tag{15}$$

where the $P(\ldots)$ values in the denominators of Equations (10)–(12) are the bivariate, and of Equations (13)–(15) are the multivariate joint probabilities of drought characteristics u, v, and w; and $T_{(\ldots)}$ values on the left-hand side of Equations (10)–(15) are the corresponding return periods. By definition, the $P(\ldots)$ terms in Equations (9)–(15) can be found using the CDFs of the related drought characteristics. Below are some examples of univariate, bivariate, and trivariate cases:

$$P(U \geq u) = 1 - F_U(u) \tag{16}$$

$$P(U \geq u | V \geq v) = 1 - F_{U,V}(u|v) \tag{17}$$

$$P(U \geq u, V \geq v, W \geq w) = 1 - F_{U,V,W}(u, v, w). \tag{18}$$

Accordingly, the bivariate probability $P(\ldots)$ terms in Equations (10)–(12) can be found using the univariate CDF estimates and the bivariate copula functions in the forms:

$$P(U \geq u, V \geq v) = 1 - F_U(u) - F_V(v) + C(F_U(u), F_V(v)) \tag{19}$$

$$P(U \geq u \text{ or } V \geq v) = 1 - C(F_U(u), F_V(v)) \tag{20}$$

$$P(U \geq u | V \geq v) = \frac{P(U \geq u, V \geq v)}{P(V \geq v)} = \frac{1 - F_U(u) - F_V(v) + C(F_U(u), F_V(v))}{1 - F_V(v)}. \tag{21}$$

The trivariate probabilities in Equations (13)–(15) can be found using the univariate CDF estimates, and bivariate and trivariate copula functions. Below is an example of a trivariate conditional case, while many other forms could be obtained in a similar fashion:

$$P(U \geq u | (V \geq v, W \geq w)) = \frac{P(U \geq u, V \geq v, W \geq w)}{P(V \geq v, W \geq w)}, \tag{22}$$

where the bivariate and trivariate joint probability terms on the right-hand side can be expressed in terms of copula functions. Below are some examples of such expressions linking multivariate probabilistic terms with copula functions:

$$P(V \leq v, W \leq w) = C(F_V(v), F_W(w)) \tag{23}$$

$$P(V \geq v, W \geq w) = 1 - F_V(v) - F_W(w) + C(F_V(v), F_W(w)) \tag{24}$$

$$P(U \leq u, V \leq v, W \leq w) = C(F_U(u), F_V(v), F_W(w)) \tag{25}$$

$$\begin{aligned} P(U \geq u, V \geq v, \quad & W \geq w) \\ & = 1 - F_U(u) - F_V(v) - F_W(w) + C(F_U(u), F_V(v)) + C(F_U(u), F_W(w)) \\ & + C(F_V(v), F_W(w)) - C(F_U(u), F_V(v), F_W(w)) \end{aligned} \tag{26}$$

$$P(U \geq u \text{ or } V \geq v \text{ or } W \geq w) = 1 - C(F_U(u), F_V(v), F_W(w)). \tag{27}$$

Equations (9)–(15) describe how T values are estimated for univariate (9), bivariate (10)–(12), and trivariate (13)–(15) cases. These equations are flexible enough to accommodate investigations of various spatial–temporal drought characteristics like LD, $\overline{S}$, and $\overline{A}$ (e.g., T of a drought event with LD $\geq$ 10 | $\overline{S}$ >0.80 & $\overline{A}$ > 0.85 can be calculated). Once the multivariate joint probabilities of the drought characteristics are determined through these copula functions, T for these drought characteristics can be determined by plugging these probability estimates into Equations (10)–(15).

2.3. Standardized Precipitation Index (SPI)

In this study, SPI [30] is selected to analyze drought events. It is worth noting that this selection is arbitrary and inconsequential, while other drought indices (e.g., Percent of Normal, Palmer Drought Severity Index, Crop Moisture Index, Surface Water Supply Index, Reclamation Drought Index, etc.) could have been used as well.

SPI can be calculated for different time scales (i.e., 3, 6, 12, etc.), while the obtained time series (SPI-3, SPI-6, SPI-12, etc.) are used as drought indicators in applications with different goals. From a drought management point of view, managers may prefer SPI-3 as an indicator to monitor short-term anomalies, SPI-12 to monitor persistent events, or SPI-6 to have a flavor of both short-term anomalies and persistent events [31]. Accordingly, in this study SPI-6 is selected for the drought analysis to quantify the deficits of precipitation. For this purpose, all of the monthly precipitation time series are stored in a matrix $P_{s,m}$, where s refers to each station and m refers to each month. Later six-month moving-window precipitation sums ($PT_{s,m}$) are obtained in the form:

$$PT_{s,m} = \Sigma_{m-5}^{m} P_{s,m}. \tag{28}$$

These $PT_{s,m}$ values are spatially averaged to obtain representative values for the entire study area in the form:

$$PT_m = \sum_{s=1}^{D} P_{s,m}, \tag{29}$$

where D is the total number of stations. Following McKee et al. [30], these $PT_{s,m}$ and PT_m time series are later fitted to gamma distribution, where the probability of any given precipitation value is obtained using this fitted gamma distribution. These probability estimates obtained using $PT_{s,m}$ and PT_m are later converted into $SPI_{s,m}$ and SPI_m values, respectively, using the inverse of the standard normal distribution. Here, $SPI_{s,m}$ values are calculated for each station and month separately, while SPI_m is representative for the entire study area and calculated for each month.

2.4. Drought Characteristics

LD, $\overline{S}$, and $\overline{A}$ are the three drought characteristics considered in this study. The definition of the dry and the non-dry periods during the study period is made using SPI_m time series.

Following McKee et al. [30], the dry periods are defined to have continuously negative SPI values and have at least one SPI value below -1.0 (e.g., the time series $(-0.23, -0.76, -1.15, -0.52, -0.05)$ constitute a drought event, while the time series $(-0.23, -0.76, -0.99, -0.52, -0.05)$ does not). Accordingly, the periods that are not considered as dry are considered "non-dry" periods. LD and the length of non-dry periods (LnD) values are calculated as:

$$\text{LD} = (t_{de} - t_{ds}) + 1 \qquad (30)$$

$$\text{LnD} = (t_{nde} - t_{nds}) + 1, \qquad (31)$$

where t_{ds} and t_{de} are the months a particular drought period started and ended, respectively; and t_{nds} and t_{nde} are the months a particular non-dry period started and ended, respectively. DIT is obtained as

$$\text{DIT} = \text{LD} + \text{LnD}. \qquad (32)$$

Even though $\text{SPI}_{s,m}$ and SPI_m values are available for each month, drought characteristic values (LD, $\bar{S}$, and $\bar{A}$) are calculated for each drought period. Among these drought characteristics, LD values are calculated using SPI_m time series. For each drought period, $\bar{S}$ is calculated as:

$$\bar{S} = |\frac{1}{\text{LD}} \sum\nolimits_{m=t_{ds}}^{t_{de}} \text{SPI}_m|. \qquad (33)$$

Separate drought definitions are made for the region and the stations using SPI_m and $\text{SPI}_{s,m}$ values, respectively. If the given region is under drought conditions, then individual stations are assessed as to whether or not they individually satisfy the drought conditions using $\text{SPI}_{s,m}$ values. Then the number of months of each station that are under drought condition during the regional drought period is calculated. The fraction of station-based drought conditions compared to the duration of regional drought length is defined as $\bar{A}$. If the SPI values for a particular month are only available for four stations, then $\bar{A}$ is calculated as the fraction of these four stations that are under drought considering the length of the particular drought event. Once drought characteristics LD, $\bar{S}$, and $\bar{A}$ are obtained for each drought period separately, the univariate, bivariate, and trivariate T values of these characteristics are calculated using Equations (10)–(15).

2.5. Drought Frequency and Visual Analysis

The visual presentation of the multivariate CDFs of the drought characteristics is instructive for building knowledge about the drought events. However, such analysis requires long datasets for visually presentable plot generation. In the absence of a sufficiently long time series, it is trivial to synthetically generate sufficient datasets, while such simulations require the distribution of the particular drought characteristic (LD, $\bar{S}$, or $\bar{A}$) to be known in advance. In many cases, observed data do not follow a specific distribution, and their distributions likely vary depending on the variables and time series [21]. In this study, LD, $\bar{S}$, and $\bar{A}$ are respectively fitted to various distributions (the seven advanced with variable support, 23 non-negative, and eight bounded available distributions; Table 2) in the EasyFit program [32] to find the distribution that resulted in the best fit to the CDF of the observed drought characteristics data. The best CDF fit is separately found for each drought characteristic using the Maximum Likelihood method [33] by optimizing the Kolmogrov–Smirnov index [34] within the EasyFit program. The distributions that resulted in the best CDF fits are later used to synthetically generate sufficiently long datasets. Here it is stressed that these synthetically generated datasets are only used for visual representations but not in other analysis.

Table 2. Fitted distribution functions (total of 38 types).

Advanced with Variable Support	Non-Negative	Non-Negative	Non-Negative	Bounded
Gen. extr. val.	Burr	Gamma	Pareto	Beta
Gen. logistic	Chi-squared	Gen. gamma	Pareto 2	Johnson SB
Gen. Pareto	Dagum	Inv. Gaussian	Pearson 5	Kumaraswamy
Log–Pearson 3	Erlang	Levy	Pearson 6	Pert
Phased Bi–exp.	Exp.	Log-gamma	Rayleigh	Power func.
Phased bi–Weibull	F	Log-logistic	Rice	Reciprocal
Wakeby	Fatigue life	Lognormal	Weibull	Triangular
	Frechet	Nakagami		Uniform

2.6. Study Area and the Datasets

Analysis of drought events is very critical over arid and semi-arid regions, like Ankara, Turkey, where water management projects require knowledge of the water status during a particular drought period. Among the different types of droughts (meteorological, hydrological, agricultural, and socioeconomic), this study focuses on the analysis of the spatial–temporal characteristics of meteorological droughts using monthly precipitation records obtained over six meteorological stations (Beypazari, Esenboga, Kecioren, Kizilcahamam, Nallihan, and Polatli) distributed over Ankara, Turkey (Figure 2). These stations are maintained and the datasets are distributed by The General Directorate of the Turkish State Meteorological Service (TSMS). Precipitation datasets over these six stations between January 1960 and February 2014 are obtained from TSMS. Precipitation time series for the first 60 months and the last two months for Nallihan and Esenboga stations, respectively (out of a total of 650 months). These missing data are ignored for these missing periods, while they constitute only <2% of the total datasets. More information about these stations can be found in the study of Sonmez [35].
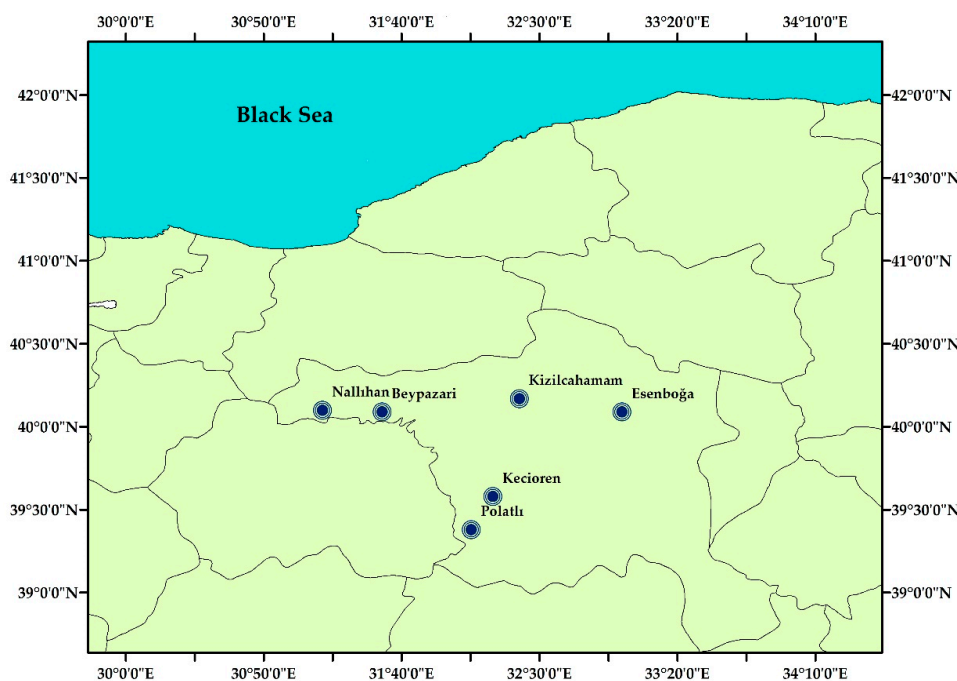

Figure 2. Locations of the six meteorological stations where the data used in this study were obtained.

3. Results and Discussion

The drought characteristics LD, $\overline{S}$, and $\overline{A}$ were calculated for each drought period separately (Table 3). Accordingly, higher $\overline{S}$ values indicate more severe drought conditions and values closer to 0 indicate near to normal conditions. For example, the drought event starting in

July 1985 and ending in January 1986 (Table 3) has an LD of seven months, $\bar{S}$ of 1.05, and $\bar{A}$ of 0.74. Here $\bar{S}$ is calculated as the absolute value of the average of seven SPI values: $|-(0.28 + 1.04 + 1.50 + 1.42 + 1.60 + 1.33 + 0.17)/7| = 1.05$; and $\bar{A}$ is calculated as the fraction of total months under drought condition (i.e., the regional average SPI time series show the drought duration is seven months, while the six separate SPI time series over each station show 31 months are under drought conditions out of a total of six stations × seven months = 42 months, implying $\bar{A} = 31/42 = 0.74$). Overall results show that drought conditions influenced the study area during 39% of the total study period (summation of all LD/650 months), while these drought events, on average, have $\bar{S}$ of 0.94, last for 6.92 months, and impact 73% of the area covered by the stations (Table 3).

Table 3. Drought characteristics defied by using SPI series values.

Drought	Start of Drought		End of Drought		LD	$\bar{S}$	$\bar{A}$
	Year	Month	Year	Month			
1	1960	9	1961	1	5	0.9	0.8
2	1961	8	1962	1	6	0.68	0.67
3	1962	7	1962	11	5	0.94	0.76
4	1965	8	1965	12	5	1.16	0.92
5	1966	9	1967	2	6	0.82	0.78
6	1967	9	1968	1	5	1.05	0.87
7	1969	9	1969	12	4	1.28	0.96
8	1970	8	1971	2	7	0.75	0.64
9	1973	3	1974	4	14	0.84	0.82
10	1974	11	1975	1	3	0.84	0.44
11	1976	7	1976	12	6	0.39	0.17
12	1977	6	1978	1	8	1.03	0.81
13	1978	8	1978	12	5	0.79	0.67
14	1979	7	1979	12	6	1.07	0.5
15	1980	8	1980	12	5	0.95	0.9
16	1981	8	1981	11	4	0.79	0.5
17	1982	10	1983	5	8	0.93	0.94
18	1984	9	1985	3	7	1.28	1
19	1985	7	1986	1	7	1.05	0.74
20	1986	7	1986	12	6	0.86	0.81
21	1987	7	1988	2	8	0.91	0.88
22	1988	9	1989	10	14	0.82	0.92
23	1991	8	1992	2	7	0.53	0.48
24	1992	6	1993	2	9	0.52	0.48
25	1993	7	1995	2	20	1.05	0.86
26	1996	8	1996	12	5	0.48	0.33
27	1998	10	1999	1	4	0.68	0.33
28	2000	9	2001	4	8	1.28	0.94
29	2001	8	2001	11	4	0.76	0.67
30	2003	7	2003	12	6	1.43	0.97
31	2004	7	2005	2	8	1.34	0.98
32	2005	10	2006	1	4	0.78	0.46
33	2007	3	2008	2	12	1.08	0.86
34	2008	6	2008	12	7	0.86	0.48
35	2009	9	2009	12	4	0.93	0.63
36	2012	7	2012	12	6	1.33	1
37	2013	7	2014	2	8	1.46	0.91
Statistics of the above given 37 drought periods				Min	3	0.39	0.17
				Mean	6.92	0.94	0.73
				Max	20	1.46	1

Droughts tend to start during drier months (July–September) and end during wetter winter months (November–February). This result is expected as SPI-6 considers six-monthly accumulated

precipitation values (Equation (28)): SPI-6 values for July–September are calculated using relatively drier months and values for November–February are calculated using relatively wetter months, where the summer and the winter months are climatologically dry and wet, respectively. This systematic pattern is consistent with the drought definition of many water resources managers who are interested in the lack of water. Removal of this strong SPI-6 seasonality may result in reduced SPI-6 sensitivity to actual water deficit (i.e., anomalies obtained during drier months may not necessarily mean the same "water deficit" during wetter months), hence SPI-6 seasonality is retained in this study.

Strong linear dependence is found between $\overline{S}$ and $\overline{A}$ (correlations are statistically significant at 95% confidence level), while the relationships between other drought characteristics are much weaker (Table 4). This implies that predictability studies involving $\overline{S}$ and $\overline{A}$ may contain high predictive skill, while particularly LD prediction using LD–$\overline{S}$ or LD–$\overline{A}$ relations may not yield accurate results. Here, drought characteristics LD and $\overline{A}$ have boundaries (0 and 1.0, respectively); however, there are not many values close to these boundary values, hence the use of correlation statistics may not be problematic assuming these drought characteristics are approximately normal.

Table 4. Kendall's tau and Pearson's correlations between drought characteristics LD, $\overline{S}$, and $\overline{A}$.

	LD$\overline{S}$	LD$\overline{A}$	$\overline{S}\overline{A}$
Kendall's tau	0.15	0.27	0.63
Pearson's	0.10	0.32	0.79

The univariate CDFs of drought characteristics are necessary to fit the copula functions $[C(u,v)_t]$, so the best fitting copula function can be used for further analysis. Here, the best-fitting copula is obtained after performing validation efforts utilizing $C(u,v)_e$ as the truth. For the purpose of fitting copulas, univariate empirical CDF values are used while the theoretical univariate CDF values are used in the synthetic simulations described below (Figure 3). Among copula functions, by definition, the nature of the AMH requires Kendall's tau between variables to be bounded between -0.18 and 0.33 [27]. Since the $\overline{S}$–$\overline{A}$ pair does not satisfy this requirement (Table 4), an AMH function could not be fitted to this pair. Leave-one-out type validation of copula parameters show that the elliptical copula family results in better performance in bivariate and trivariate joint CDF estimation analysis than other copula families, while normal copula seems to have the best overall performance (Table 5).

Table 5. Parameters (θ or ρ) and RMSE values of bivariate and trivariate copula fitting simulations, where the three smallest RMSE values are written in bold for each analysis.

Copula Function	Variables of Bivariate Copulas						Variables of Trivariate Copula	
	LD & $\overline{S}$		LD & $\overline{A}$		$\overline{S}$ & $\overline{A}$		LD & $\overline{S}$ & $\overline{A}$	
	(θ/ρ)	RMSE	(θ/ρ)	RMSE	(θ/ρ)	RMSE	(θ/ρ)	RMSE
AMH	0.514	0.062	0.895	0.056	-	-	-	-
Clayton	0.280	0.064	0.728	0.058	2.395	0.043	0.930	0.067
Frank	1.272	0.061	2.425	0.058	8.170	0.030	3.026	0.065
GH	1.115	0.065	1.271	0.064	2.514	0.029	1.425	0.070
Joe	1.120	0.070	1.269	0.075	2.966	0.042	1.538	0.084
Normal	0.242	**0.059**	0.426	**0.055**	0.839	**0.026**	0.528	**0.054**
t-Student	0.244	0.061	0.426	0.056	0.838	**0.026**	0.528	0.056

Analysis of drought period counts show a drought event is expected on average every 18 months (i.e., $E(DIT) = 650$ months/37 droughts). Univariate drought T calculations using the univariate CDF of variables (Figure 3) show on average drought events with LD $\geq$ 6.92 are expected every 4.15 years, $\overline{S} \geq 0.94$ every 3.06 years, and $\overline{A} \geq 0.73$ every 2.53 years (Table 6). However, in many cases these univariate T estimates may not be sufficient for various drought investigations requiring bivariate

or trivariate T information. These multivariate drought T values are calculated using the best fitting copula functions for the bivariate and the trivariate cases. As an example, T is calculated for the bivariate case as 7.13 years for the drought event with $\overline{S} \geq 0.94$ and LD ≥ 6.92; as 3.41 years for a drought event with $\overline{A} \geq 0.73$ and $\overline{S} \geq 0.94$; and for the trivariate case T is calculated as 6.65 years for a drought event with $\overline{S} \geq 0.94$, $\overline{A} \geq 0.73$, and LD ≥ 6.92.
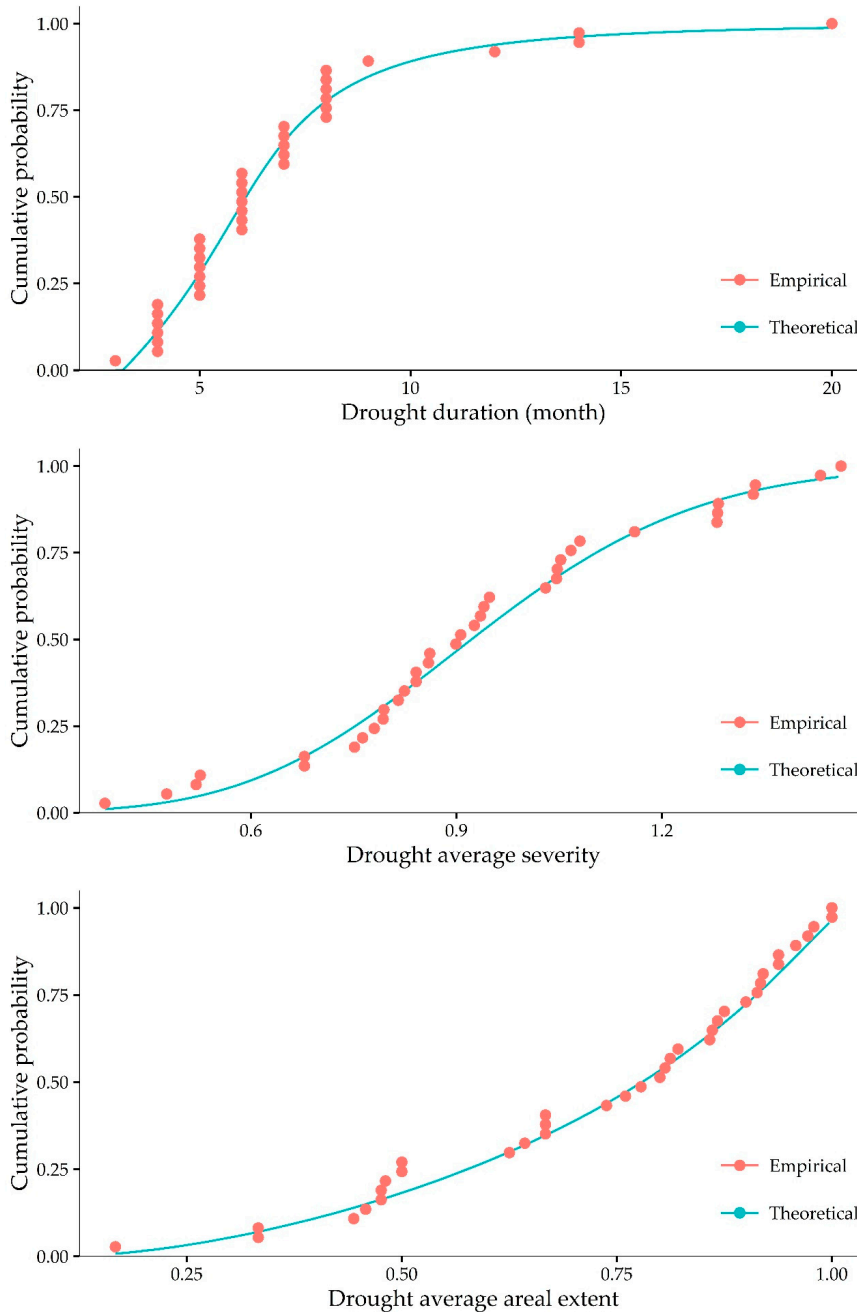
Figure 3. Empirical and theoretical univariate CDF of drought characteristics LD, $\overline{S}$, and $\overline{A}$. The theoretical CDF values in the top (LD), middle ($\overline{S}$), and bottom ($\overline{A}$) panels are calculated using Wakeby, Pearson 5, and Johnson SB distributions, respectively.

The examples given above could be reproduced for infinitely many drought characteristic pairs or triplets. However, plots showing their multivariate variations often are necessary to understand the exact nature of their multivariate relations. However, to generate such plots the lengths of the

available 37 drought events are not sufficient. Instead, the drought characteristics are fitted to different distributions and the best fitting distributions are later used to generate synthetic drought characteristic datasets that are used to generate multivariate T plots (Figures 4 and 5). Parameters result in the smallest Kolmogorov–Smirnov statistics (Table 7) are used to generate the relevant synthetic datasets. Results show log-logistic, normal, and logistic distributions result in the best fit (i.e., lowest errors) for the drought characteristics of LD, $\bar{S}$, and $\overline{A}$, respectively (Table 7). Hence, these distributions are used to create synthetic LD, $\bar{S}$, and $\overline{A}$ data.

Table 6. Univariate drought return periods (T).

T (Years)	Drought Characteristics		
	LD (Month)	$\bar{S}$ (-)	$\overline{A}$ (-)
1.46	3.00	-	-
1.47	3.24	-	0.17
1.48	3.26	0.39	0.18
2.53	5.71	0.87	0.73
3.06	6.21	0.94	0.79
4.15	6.92	1.02	0.86
40.57	14.20	1.44	1.00
48.19	15.01	1.46	-
116.83	20.00	-	-

Table 7. Kolmogorov–Smirnov statistics and parameters of best theoretical CDF distributions (Figure 3) for the drought characteristics LD, $\bar{S}$, and $\overline{A}$.

Drought Characteristic	Distribution	Parameters	Test Statistic
LD	Wakeby	$\alpha = 6.575$; $\beta = 3.344$; $\gamma = 1.470$; $\delta = 0.34$; $\xi = 3.176$	0.10
$\bar{S}$	Pearson 5	$\alpha = 167.61$; $\beta = 563.17$; $\gamma = -2.44$	0.08
$\overline{A}$	Johnson SB	$\gamma = -0.733$; $\delta = 0.711$; $\lambda = 0.937$; $\xi = 0.088$	0.08

Three drought characteristics (LD, $\bar{S}$, and $\overline{A}$) are considered in this study, hence three bivariate drought characteristic pairs ($\bar{S}$–LD, $\overline{A}$–LD, and $\overline{A}$–$\bar{S}$) are analyzed. Additionally, three bivariate drought operators ("and", "or", "conditional") that are suited for various drought applications are also considered. Accordingly, a total of nine bivariate T maps were obtained for visual presentation efforts (Figure 4). In Figure 4, the left, middle, and right columns show "and", "or", and "conditional" scenarios, respectively; the top, middle, and bottom rows show the $\bar{S}$–LD, $\overline{A}$–LD, and $\overline{A}$–S pairs, respectively.

The "and" scenario shows that the bivariate T has a higher sensitivity to LD than $\bar{S}$ or $\overline{A}$: When LD is included in the bivariate T maps (left middle and top panels), T values are more impacted by LD than $\overline{A}$ or $\bar{S}$, particularly for higher values of LD, while the sensitivity difference between $\overline{A}$–$\bar{S}$ pairs is marginal. If a line is drawn from the lower left corner to the upper right corner of the $\overline{A}$–$\bar{S}$ bivariate T figure (lower-left panel of Figure 4), then the area to the left of this line shows the region where T is more sensitive to changes in $\overline{A}$ than $\bar{S}$ while the area on the right side of the line shows the region where T is more sensitive to changes in $\bar{S}$ than $\overline{A}$. These results suggest that the overall sensitivity of T to these drought characteristics can be compared via LD > $\overline{A}$–$\bar{S}$. On average, bivariate T values are much higher, particularly for higher values of LD (LD > 10). Similar to the "and" case, the "or" case also show pareto front type plots but the pareto front direction is reversed (i.e., the middle column): on average almost one in 2.33 years a mild drought event with "$\overline{A} \geq 0.73$ or $\bar{S} \geq 0.94$" or "LD ≥ 6.91 or $\bar{S} \geq 0.94$" is observed. As expected, the average T value for the "or" case is much lower than for the "and" case as the probability of an "or" case is higher than an "and" case. In the case of "conditional" drought T plots, the T values are most sensitive to the unconditioned value rather than the given observation

(i.e., the lines are vertical), while this sensitivity decreases as the drought impact increases (higher LD, $\overline{S}$, or $\overline{A}$). On average, the dynamic range of T values for conditional case are more similar to the "or" case, particularly for less severe drought events (e.g., for the bivariate case they range between 1.5 and 42 years). These conditional T values could be particularly useful if certain observations could be obtained. For example, if $\overline{S} \geq 0.94$, then T values range between 1.86 and 9.95 years for $5 \leq LD \leq 10$ (Figure 4, top right panel).
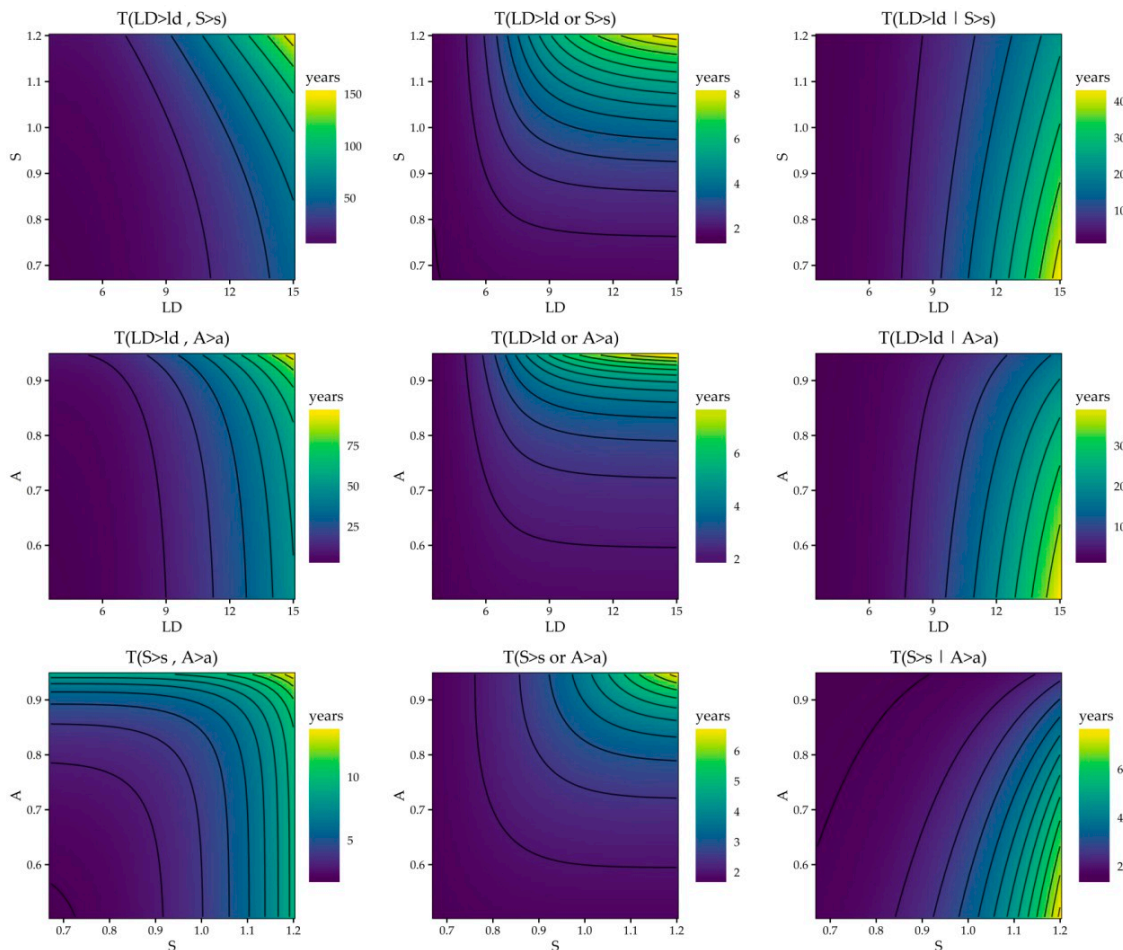
Figure 4. Bivariate T of drought characteristics obtained using synthetic datasets, where LD is given in months and $\overline{S}$ and $\overline{A}$ are shown as S and *A*, respectively.

Figure 5 shows that the trivariate T values are conditioned on both $\overline{S}$ and $\overline{A}$. Given that it would be harder to digest 3D plots, trivariate cases are plotted for four separate $\overline{A}$ values (0.65, 0.75, 0.85, and 0.95). Similar to conditional bivariate plots, conditional trivariate plots also show that T values are sensitive to the unconditioned variables (i.e., the plots are perpendicular to the unconditioned variable). Additionally, the conditional trivariate contour plots show that when additional observations are available compared to the bivariate case, then T values decrease (i.e., if $\overline{A} \geq 0.65$, it is more likely that drought events with longer duration will happen). For example, the bivariate conditional plot for LD$|\overline{S} \geq 0.7$ shows T values ranging between 1.95 and 42.04, while the trivariate conditional plot for LD$|\overline{S} \geq 0.7$, $\overline{A} \geq 0.95$ shows T values ranging between 2.05 and 13.69. The conditional T values have much higher sensitivity to LD when $\overline{A} \geq 0.75$ than when $\overline{A} \geq 0.95$ (Figure 5); for greater $\overline{A}$ values, the return periods are much smaller for conditional LD compared to smaller $\overline{A}$ values.
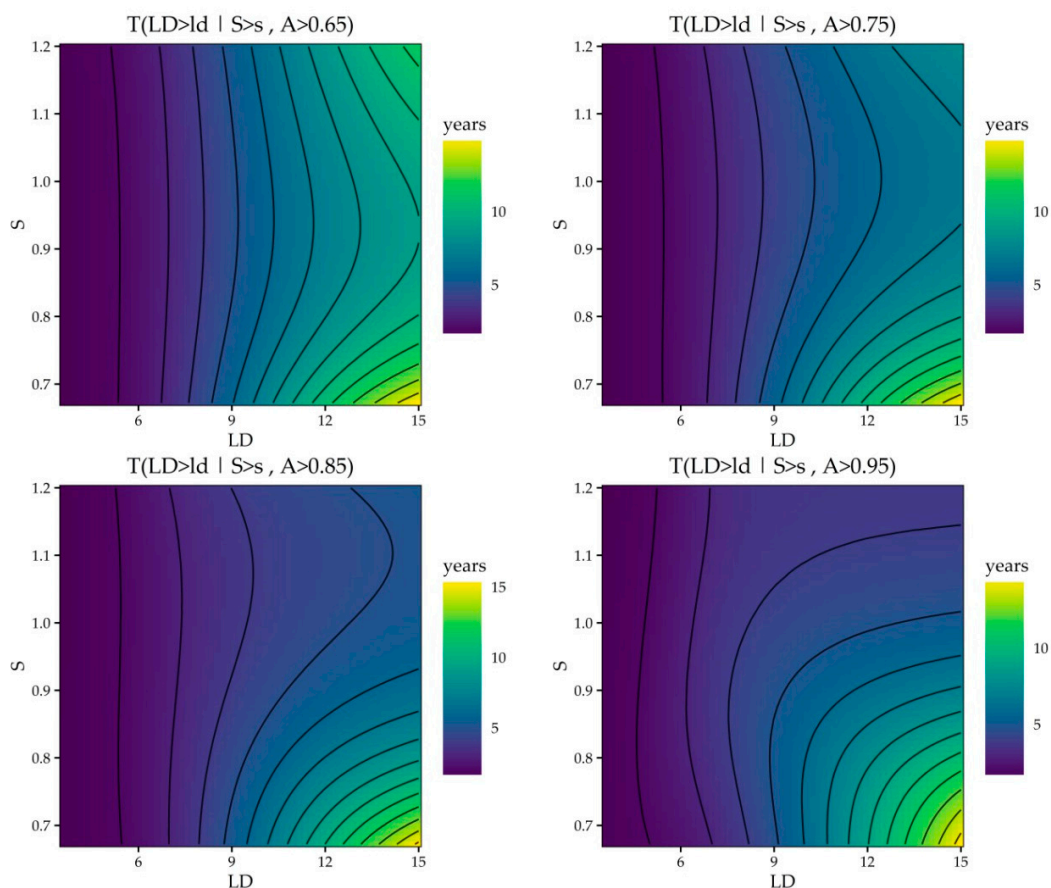
Figure 5. T of trivariate drought characteristics (for the cases of $\overline{A}$ greater than 75, 85, and 95 percent), where LD is given in months and for brevity $\overline{S}$ and $\overline{A}$ are shown as S and A, respectively.

4. Conclusions

Droughts have many different spatial–temporal characteristics; hence, a better way to describe them is to use joint probability-based functions, rather than building drought-related analyses on univariate datasets as joint distribution of these drought characteristics often do not follow a particular known distribution. The current study offers a methodology to analyze the spatial–temporal multivariate aspect of droughts via copula functions in a probabilistic approach.

The results show that, among the drought characteristics, $\overline{S}$ and $\overline{A}$ are found to have significant linear dependence, implying their predictions using each other may yield reasonable forecasts, while the same may not be true for other relationships including LD. On the other hand, bivariate T is more sensitive to variability in LD than $\overline{S}$ or $\overline{A}$, implying that predictions of bivariate T could be more accurate with the presence of LD than $\overline{S}$ or $\overline{A}$. Nevertheless, observations of $\overline{S}$ or $\overline{A}$ narrow the variability window of trivariate T predictions, particularly for higher values of $\overline{S}$ or $\overline{A}$; implying that $\overline{S}$ or $\overline{A}$ observations have utility in predictions of T even though such predictions are more sensitive to variability in LD.

The flexibility of T calculation for different multivariate drought scenarios clearly illustrates the power of the introduced methodology. Additionally, certain drought characteristics may precede others, hence such observed drought characteristics may prove very valuable when making inferences about non-observed drought characteristics, while such relations can be easily studied using the multivariate conditional copula methodology presented in this study. In particular, trivariate drought analysis scenarios could prove very helpful in estimating a particular drought characteristic given availability of other drought characteristic info.

Drought characteristics often have dataset-specific distributions. Even though elliptical family-based copula functions yielded better performance in this study, other studies found that different copula families perform better [13,20]. This implies that the optimality of the copula functions could be dataset- and case-specific; hence inter-comparisons of copula functions could be necessary before estimation of return periods in other studies. However, this result should be confirmed via independent studies before it can be generalized for copula function-related drought analyses.

In this study, ground station-based datasets are utilized in all of the analyses. While they provide the longest datasets over the most locations, these station-based datasets are spatially limited, often contain gaps in their time series, and may have user-dependent data quality. On the other hand, remote sensing-based datasets prove very powerful alternatives to station-based datasets as these spatially and temporally consistent datasets are available everywhere, including remote locations where installation of stations is not economically feasible. Therefore, remotely sensed datasets could be recommended as alternative reanalysis products to be used as such model datasets may provide consistent and sufficiently long time series for drought analysis.

Author Contributions: Ali Unal Sorman designed the experiments, provided the data, and contributed to the analysis tools; Mahdi Hesami Afshar and Mustafa Tugrul Yilmaz designed the experiments, analyzed the data, contributed to the analysis tools, and wrote the paper.

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

References

1. Tadesse, T.; Wilhite, D.A.; Harms, S.K.; Hayes, M.J.; Goddard, S. Drought monitoring using data mining techniques: A case study for nebraska, USA. *Nat. Hazards* **2004**, *33*, 137–159. [CrossRef]
2. Zhong, S.; Shen, L.; Sha, J.; Okiyama, M.; Tokunaga, S.; Liu, L.; Yan, J. Assessing the water parallel pricing system against drought in China: A study based on a CGE model with multi-provincial irrigation water. *Water* **2015**, *7*, 3431–3465. [CrossRef]
3. Avilés, A.; Célleri, R.; Solera, A.; Paredes, J. Probabilistic forecasting of drought events using markov chain- and bayesian network-based models: A case study of an andean regulated river basin. *Water* **2016**, *8*, 37. [CrossRef]
4. Karavitis, C.A.; Alexandris, S.; Tsesmelis, D.E.; Athanasopoulos, G. Application of the standardized precipitation index (SPI) in Greece. *Water* **2011**, *3*, 787–805. [CrossRef]
5. Guerrero-Salazar, P.; Yevjevich, V. *Analysis of Drought Characteristics by the Theory of Runs*; Colorado State University: Fort Collins, CO, USA, 1975.
6. Yevjevich, V. *An Objective Approach to Definitions and Investigations of Continental Hydrologic Droughts*; Colorado State University: Fort Collins, CO, USA, 1967.
7. Clausen, B.; Pearson, C. Regional frequency analysis of annual maximum streamflow drought. *J. Hydrol.* **1995**, *173*, 111–130. [CrossRef]
8. Rossi, G.; Benedini, M.; Tsakiris, G.; Giakoumakis, S. On regional drought estimation and analysis. *Water Resour. Manag.* **1992**, *6*, 249–277. [CrossRef]
9. Şen, Z. Regional drought and flood frequency analysis: Theoretical consideration. *J. Hydrol.* **1980**, *46*, 265–279. [CrossRef]
10. Henriques, A.; Santos, M. Regional drought distribution model. *Phys. Chem. Earth B Hydrol. Oceans Atmos.* **1999**, *24*, 19–22. [CrossRef]
11. Pelletier, J.D.; Turcotte, D.L. Long-range persistence in climatological and hydrological time series: Analysis, modeling and application to drought hazard assessment. *J. Hydrol.* **1997**, *203*, 198–208. [CrossRef]
12. Serinaldi, F.; Bonaccorso, B.; Cancelliere, A.; Grimaldi, S. Probabilistic characterization of drought properties through copulas. *Phys. Chem. Earth A/B/C* **2009**, *34*, 596–605. [CrossRef]

13. Shiau, J. Fitting drought duration and severity with two-dimensional copulas. *Water Resour. Manag.* **2006**, *20*, 795–815. [CrossRef]

14. Xu, K.; Yang, D.; Xu, X.; Lei, H. Copula based drought frequency analysis considering the spatio–temporal variability in Southwest China. *J. Hydrol.* **2015**, *527*, 630–640. [CrossRef]

15. Sklar, M. *Fonctions de Répartition à n Dimensions et Leurs Marges*; Université Paris: Paris, France, 1959; Volume 8.

16. De Michele, C.; Salvadori, G. A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. *J. Geophys. Res. Atmos.* **2003**, *108*. [CrossRef]

17. Ellouze-Gargouri, E.; Bargaoui, Z. Runoff estimation for an ungauged catchment using geomorphological instantaneous unit hydrograph (GIUH) and copulas. *Water Resour. Manag.* **2012**, *26*, 1615–1638. [CrossRef]

18. Favre, A.C.; El Adlouni, S.; Perreault, L.; Thiémonge, N.; Bobée, B. Multivariate hydrological frequency analysis using copulas. *Water Resour. Res.* **2004**, *40*. [CrossRef]

19. Grimaldi, S.; Serinaldi, F. Design hyetograph analysis with 3-copula function. *Hydrol. Sci. J.* **2006**, *51*, 223–238. [CrossRef]

20. Song, S.; Singh, V.P. Frequency analysis of droughts using the Plackett copula and parameter estimation by genetic algorithm. *Stoch. Environ. Res. Risk Assess.* **2010**, *24*, 783–805. [CrossRef]

21. Saghafian, B.; Mehdikhani, H. Drought characterization using a new copula-based trivariate approach. *Nat. Hazards* **2014**, *72*, 1391–1407. [CrossRef]

22. Klein, B.; Meissner, D.; Kobialka, H.-U.; Reggiani, P. Predictive uncertainty estimation of hydrological multi-model ensembles using pair-copula construction. *Water* **2016**, *8*, 125. [CrossRef]

23. Li, H.; Shao, D.; Xu, B.; Chen, S.; Gu, W.; Tan, X. Failure analysis of a new irrigation water allocation mode based on copula approaches in the Zhanghe Irrigation District, China. *Water* **2016**, *8*, 251. [CrossRef]

24. Zhang, L.; Singh, V.P. Bivariate rainfall and runoff analysis using entropy and copula theories. *Entropy* **2012**, *14*, 1784–1812. [CrossRef]

25. Joe, H. *Multivariate Models and Multivariate Dependence Concepts*; CRC Press: Boca Raton, FL, USA, 1997.

26. Nelsen, R.B. *An Introduction to Copulas*; Springer: New York, NY, USA, 2013; Volume 139.

27. Hofert, M.; Kojadinovic, I.; Maechler, M.; Yan, J. Copula: Multivariate dependence with copulas. R package version 0.999-5, 2012. Available online: http://CRAN.R-project.org/package=copula (accessed on 24 September 2016).

28. The R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2015.

29. Shiau, J.-T.; Shen, H.W. Recurrence analysis of hydrologic droughts of differing severity. *J. Water Resour. Plan. Manag.* **2001**, *1*, 30–40. [CrossRef]

30. McKee, T.B.; Doesken, N.J.; Kleist, J. The Relationship of Drought Frequency and Duration to Time Scales. In Proceedings of the 8th Conference on Applied Climatology, Anaheim, CA, USA, 17–22 January 1993; American Meteorological Society: Boston, MA, USA, 1993; pp. 179–183.

31. Wilhite, D.A. *Drought and Water Crises: Science, Technology, and Management Issues*; CRC Press: Boca Raton, FL, USA, 2005.

32. Schittkowski, K. *Data Fitting and Experimental Design in Dynamical Systems with EASY-FIT^{ModelDesign}—User's Guide*; University of Bayreuth: Bayreuth, Germany, 2009.

33. Richards, F.S.G. A method of maximum-likelihood estimation. *J. R. Stat. Soc. Ser. B Methodol.* **1961**, *23*, 469–475.

34. D'Agostino, R.B. *Goodness-of-Fit-Techniques*; CRC Press: Boca Raton, FL, USA, 1986; Volume 68.

35. Sönmez, İ. Quality control tests for western turkey Mesonet. *Meteorol. Appl.* **2013**, *20*, 330–337. [CrossRef]