Application of probability decision system and particle swarm optimization for improving soil moisture content

Panpan Zhu, Hossein Saadati and Majid Khayatnezhad

a School of Economics and Management, Henan Agricultural University, Zhengzhou 450046, Henan, China
b Department of Environmental Sciences, Ardabil Branch, Islamic Azad University, Ardabil, Iran

*Corresponding author. E-mail: zpp@henau.edu.cn

ABSTRACT

Drought is one of the natural disasters having the highest degrees in comparison to the other natural disasters in terms of rate, intensity, incident duration, region expansion, life losses, economic damages, and long-term effects. Hence, the prediction of drought as a meteorological phenomenon should be evaluated to determine the groundwater exploitation strategies in agriculture. The present study aims at investigating the impact of the drought duration and severity on soil moisture supplement for agricultural activities in Baghmalek plain, Khuzestan province, Iran. For this objective, a non-dimensional index of precipitation depth was defined for quantifying the drought characteristics. Furthermore, marginal distribution functions, correlation coefficients and joint functions were incorporated to a probabilistic decision-making framework to predict the variables in different return periods from 2-year to 100-year periods. Results showed that t copula was the best function for constructing the multivariate distribution in the study area based on the goodness-of-fit tests. Moreover, soil moisture content in the root zone achieved by the predetermined amounts of precipitation could be increased in the seasonal average.

Key words: Archimedean, copula, drought, frequency analysis, precipitation

HIGHLIGHTS

- A new probabilistic framework was developed for sustainable water allocation.
- Fuzzy particle swarm optimization was employed for generating the optimal solution.
- A daily time-step model was simulated using mathematical formulation of crop growth process.
- Developed framework increased the water use efficiency in the cropping pattern of Baghmalek as the study area.

INTRODUCTION

Rainfall status and water resources constraints and climate conditions in Iran indicate the fact that there must be a plan for drought phenomena and its effects and consequences must be seriously addressed at the time of occurrence (Amirataee et al. 2018; Salas & Obeysekera 2019; Ebadi et al. 2020; Lalehzari & Kerachian 2020b). Drought and increasing water demands decrease the groundwater resources for food security in arid and semi-arid areas (Huang et al. 2021a; Li et al. 2021; Ren & Khayatnezhad 2021). The effects of this phenomenon are gradually observed in water resources, agriculture, environment and society (Chao et al. 2018; Vicente-Serrano et al. 2018; Zhang et al. 2019).

Several researches have been conducted for evaluating the drought indexes in different fields of water resource planning and management (Amirataee et al. 2018; Vicente-Serrano et al. 2018). Percent of normal index was applied for considering the drought severity in Urmia Lake in Iran. Results showed that the drought severity could be occurring in duration of increasing agricultural activities (Nikbakht et al. 2013). The characteristics of drought are significantly correlated with each other. This means that changes in one characteristic affect other characteristics of drought. In these circumstances, univariate analyses do not provide an accurate interpretation of the drought characteristics (Kwon & Lall 2016). The main problem of the initial bivariate frequency functions was the application of the same family for marginal frequency distributions (Banibayat et al. 2021).

Considering the previous studies focus on investigation of the effect of drought phenomena on precipitation and temperature variables, less attention has been paid to the soil water demands in agriculture (Yang et al. 2015, 2020a, 2020b; Lalehzari & Kerachian 2020a; Xu et al. 2021). Conditional value at risk, fuzzy set theory,
and Monte Carlo simulation were the techniques of uncertainty analysis in the previous studies (Chen & Guo 2019; Alam et al. 2020; Lalehzari & Kerachian 2020b; Simos and Tsitouras 2020; Chen et al. 2021). Application of these techniques is unable to present a predicted pattern of hydrological phenomena for generating a multivariate estimation to define the future events. Drought severity and duration analysis, as the main components of water management in food security policies, is necessary, using probabilistic functions to achieve the predetermined pattern to decrease the water deficit in the root zone. Bivariate distribution functions such as copula have been widely implemented in drought analysis (Kao & Govindaraju 2010; Kwon & Lall 2016; Kong et al. 2018; Dehghani et al. 2019). The main idea of copula functions was first proposed by Sklar (1959) and used to derive the distribution of random variables with abnormal marginal distributions. Salvadori & De Michele (2004) first used these functions in hydrological studies to create a bivariate model describing the intensity and duration of storms.

Therefore, developing a probabilistic decision system could be used to incorporate the drought characteristics in the determination of soil moisture content. Consideration of soil moisture balance, farmer’s priorities and agricultural constraint are the main factors in decision-making. Drought severity and duration were analyzed using bivariate joint functions to estimate the return periods of water deficit. Finally, the impact of drought characteristics on improving the soil moisture content in agriculture was evaluated.

MATERIAL AND METHODS

Daily precipitation depths during the statistical period of 1978–2020 were analyzed and examined for accuracy in Baghmalek plain, located in Khuzestan province, Iran. This region has an area of approximately 62.4 km² between 49° 39′ to 50° 11′ north longitudes and 31° 22′ to 31° 42′ east latitudes. The mean elevation of the plain is 743.5 m above sea level. Mean precipitation and evaporation are 596 mm/year and 1,445 mm/year, respectively. Furthermore, the average maximum and minimum monthly temperature are 41.8 °C in July and 5.3 °C in January in the farming year, respectively (Lalehzari et al. 2020). Drought characteristics including severity (Ds) and duration (Dd) were extracted using the analyzed precipitation depth (Pr) based on the non-dimensional form as follows.

\[ \text{NPD} = \frac{(Pr - M)}{st} \]  

where, NPD is the non-dimensional precipitation depth, M and st are the mean and standard deviation of long-term data. With the defined concept, a drought period begins when the NRI is less than zero and the severity of drought is equal to the absolute value of the sum of the index values in a drought period (Ayantobo et al. 2019). Fitting of the marginal distribution functions for each drought characteristic, determination of the correlation between drought characteristics, creating a two-dimensional joint function and calculating return periods were the main steps to achieve the probabilistic values of drought. Frequency analysis functions applied to determine the appropriate marginal distribution for each drought characteristic were formulated, as shown in Table 1.

Kolmogorov-Smirnov and Anderson-Darling tests were used to select the superior distribution (Abdi et al. 2017; Ayantobo et al. 2018). \( r \) (Equation (11)) and \( \rho \) (Equation (12)) coefficients were incorporated to evaluate the correlation between drought characteristics. These coefficients can be written as (Chen & Guo 2019):

\[ \tau = P[\{x_i < x_j, y_i < y_j\} \text{or} (x_i > x_j, y_i > y_j)] - P[\{x_i < x_j, y_i < y_j\} \text{or} (x_i < x_j, y_i > y_j)] \]  

\[ \rho = 1 - \frac{6(x_i - y_i)^2}{n(n^2 - 1)} \]  

\( \tau \) = Kendall correlation coefficient and \( \rho \) = Spearman correlation coefficient. \( x \) and \( y \) are the drought variables.

Copula function, as a bivariate technique, is expected to lead to more accurate estimation of return periods in the drought frequency analysis. A copula function links marginal distribution functions to a flexible multivariate distribution. If \( X_1, X_2, \ldots, X_d \) are random variables, and \( F \) is a \( d \)-dimensional distribution, copula \( C \) can be obtained as follows (Kong et al. 2018):

\[ F(x_1, \ldots, x_d) = C(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)) \]
Numerous studies have used optimization methods to estimate and calibrate the coefficients of the marginal distribution functions based on an objective function that minimizes the difference between observed and predicted drought variables. Furthermore, return periods of the drought severity and duration are estimated based on the joint distribution function. Therefore, particle swarm optimization (Shi & Eberhart 1999) was incorporated to estimate the coefficients of the marginal distribution functions (De Michele et al. 2005):  

\[ T_{(ar)} = e / 1 - F(s, d) \]  

where \( F(s, d) \) is the copula function; \( e \) is equal to 1 for the annual data set. A suitable joint function to create a bivariate distribution function can be determined based on five steps, including: 1. fitting the appropriate marginal distribution on each of the studied variables; 2. estimating the coefficients of the marginal distribution functions using the maximum likelihood criteria; 3. evaluating the correlation coefficient between two marginal distribution functions; and 4. selecting the best joint function of different joint functions to connect the univariate functions. For this subject, Akaike information criterion (AIC) and Bayesian information criteria (BIC) can be implemented to evaluate fitting biases of various copulas (Zhang & Singh 2007). The AIC and BIC can be expressed as following equations:  

\[ AIC = 2m - 2 \ln(L) \]  

\[ BIC = N \ln \left( \frac{1}{N} \sum_{i=1}^{N} (P_{ei} - P_{i})^2 \right) + m \ln(N) \]  

where, \( P_{ei} \) and \( P_{i} \) are the empirical probability and theoretical probability, respectively; \( m \) is the number of parameters; and \( L \) is the maximized value of the likelihood function for the estimated model.

An intelligent search engine is needed to estimate the coefficients of the marginal distribution functions based on an objective function that minimizes the difference between observed and predicted drought variables. Numerous studies have used optimization methods to estimate and calibrate the coefficients (Huang & Wang 2021; Huang et al. 2021b). Therefore, particle swarm optimization (Shi & Eberhart 1999) was incorporated to the model structure to find the coefficients of the marginal functions (\( \beta, k, \lambda \) and \( \gamma \)). Developing the calibration

### Table 1 | Marginal distribution functions for univariate frequency analysis

<table>
<thead>
<tr>
<th>Name</th>
<th>Cumulative distribution function (CDF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>( F(x) = 1 - \exp \left( -\frac{(x - \gamma)x}{\alpha} \right) )</td>
</tr>
<tr>
<td>Gamma</td>
<td>( F(x) = \frac{\Gamma(x-\gamma/a)}{\Gamma(a)} )</td>
</tr>
<tr>
<td>Generalized Extreme Values</td>
<td>( F(x) = \exp \left( -\exp \left( k^{-1} \ln \left( 1 - \frac{k(x - \gamma)}{a} \right) \right) \right) )</td>
</tr>
<tr>
<td>Log-Logistic</td>
<td>( F(x) = \left( 1 + \left( \frac{\beta}{x - \gamma} \right) ^{-1} \right) ^{-l} )</td>
</tr>
<tr>
<td>Normal</td>
<td>( F(x) = \Phi \left( -k^{-1} \ln \left( 1 - \frac{k(x - \gamma)}{a} \right) \right) )</td>
</tr>
<tr>
<td>Inv-Gaussian</td>
<td>( F(x) = \Phi \left( \frac{\lambda}{x - \gamma} \right) + \Phi \left( -\sqrt{\frac{\lambda}{x - \gamma}} \right) )</td>
</tr>
<tr>
<td>Log-Normal</td>
<td>( F(x) = \Phi \left( \frac{\ln (x - \gamma) - \mu}{\sigma} \right) )</td>
</tr>
<tr>
<td>Log-Pearson</td>
<td>( F(x) = \frac{\Gamma_{\ln(x-\gamma)/\alpha}}{\Gamma(\alpha)} )</td>
</tr>
<tr>
<td>Weibull</td>
<td>( F(x) = 1 - \exp \left( -\left( \frac{x - \gamma}{\beta} \right)^{\alpha} \right) )</td>
</tr>
</tbody>
</table>

where \( a, \beta, \gamma, k \) are the calibrated parameters; \( \mu \) = mean; \( \sigma \) = standard deviation.
frameworks for linear and non-linear problems using metaheuristic algorithms has been successfully carried out in recent years (Lalehzari et al. 2016; Moradzadeh et al. 2020).

RESULTS AND DISCUSSION

The marginal distribution functions used to the frequency analysis of the drought severity and duration are summarized in Table 2. The results obtained by Kolmogorov-Smirnov and Anderson-Darling tests showed that the gamma and generalized extreme values (GEV) functions were the best options to predict the severity and duration of drought, respectively.

For evaluation of the bivariate model performance, correlation coefficient (CC) were computed for the calibration period (from 1978 to 2020) between the drought characteristics, for which \( \gamma_k \) and \( \rho_1 \) were obtained as 0.54 and 0.62, respectively (Equations (11) and (12)). In the next step, to determine the superior joint function, the results of AIC, BIC and maximum likelihood estimator (MLE) were presented in Table 3.

The copula function with the smaller AIC and BIC values is the better one. The results showed that the t copula fits better than others according to the table. The t copula function was defined as follows:

\[
C(u_1, u_2, \theta_1, \theta_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi(1 - \theta^2)^{1/2}} \times \left( 1 + \frac{(s^2 - 2\theta st + t^2)}{v(1 - \theta^2)} \right)^{-\theta_1^2/2} dsdt
\]

where \( \theta_1 \) and \( \theta_2 \) are the correlation parameters and \( v \) = degree of freedom.

Graphical comparison presented in Figure 1 showed a good agreement between the cumulative probability of drought characteristics and random pattern generated by t copula function.

<table>
<thead>
<tr>
<th>Marginal functions</th>
<th>Drought severity</th>
<th>Drought duration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kolmogorov-Smirnov</td>
<td>Anderson-Darling</td>
</tr>
<tr>
<td>Gamma</td>
<td>0.08 1</td>
<td>0.19 2</td>
</tr>
<tr>
<td>GEV</td>
<td>0.11 2</td>
<td>0.22 3</td>
</tr>
<tr>
<td>Log-logistic</td>
<td>0.19 4</td>
<td>0.23 4</td>
</tr>
<tr>
<td>Normal</td>
<td>0.31 7</td>
<td>1.13 7</td>
</tr>
<tr>
<td>Inv-Gaussian</td>
<td>0.27 6</td>
<td>1.65 8</td>
</tr>
<tr>
<td>Log-normal</td>
<td>0.22 5</td>
<td>0.26 5</td>
</tr>
<tr>
<td>Log-Pearson</td>
<td>0.16 5</td>
<td>0.15 1</td>
</tr>
<tr>
<td>Weibull</td>
<td>0.46 9</td>
<td>0.37 6</td>
</tr>
<tr>
<td>Exponential</td>
<td>0.33 8</td>
<td>1.85 9</td>
</tr>
</tbody>
</table>

Table 3 | Goodness-of-fit criteria to evaluate the bivariate functions

<table>
<thead>
<tr>
<th>Bivariate functions</th>
<th>AIC</th>
<th>BIC</th>
<th>MLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clayton</td>
<td>-52.1</td>
<td>-53.1</td>
<td>24.1</td>
</tr>
<tr>
<td>Frank</td>
<td>-62.2</td>
<td>-63.1</td>
<td>21.5</td>
</tr>
<tr>
<td>Gamble</td>
<td>-59.4</td>
<td>-58.2</td>
<td>24.1</td>
</tr>
<tr>
<td>Gamble-Hoggard</td>
<td>-68.4</td>
<td>-59.0</td>
<td>19.1</td>
</tr>
<tr>
<td>Joe</td>
<td>-65.2</td>
<td>-61.9</td>
<td>17.6</td>
</tr>
<tr>
<td>t</td>
<td>-70.3</td>
<td>-68.1</td>
<td>16.7</td>
</tr>
</tbody>
</table>
The drought severity and duration are calculated using the frequency analysis and in different return periods based on the (or) operator (Figure 2). Furthermore, the fitted equations calibrated to estimate the non-dimensional precipitation depth based on the time variable are summarized in Table 4. The presented relationships could be incorporated as the predetermined information for irrigation planning, soil water simulation and groundwater resources management.

Decreasing rainfall and increasing the severity and duration of drought in agriculture increase the water needs for irrigation. Increasing the water productivity in higher return periods has decreased the total agricultural water demand (Figure 3). Because, optimized water demands have been able to increase water efficiency by changing the time and amount of groundwater allocation strategies.
Soil moisture index (SMI) was defined to indicate the effect of drought parameters on providing the crop water requirement in the root zone area. SMI is defined as follows:

\[
SMI_i = \frac{SMC_i - SMC_{pwp}}{SMC_s - SMC_{pwp}}
\]  

(18)

**Table 4** | Determination of the non-dimensional precipitation depth based on time

<table>
<thead>
<tr>
<th>Return period</th>
<th>Non-dimensional precipitation depth</th>
<th>Min</th>
<th>Max</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-year</td>
<td>( NPD = 2 \times 10^{-3}I^4 - 8 \times 10^{-6}I^2 + 0.0008I^2 - 0.0274I - 0.1 )</td>
<td>–1</td>
<td>1</td>
<td>0.83</td>
</tr>
<tr>
<td>5-year</td>
<td>( NPD = 2 \times 10^{-3}I^4 - 8 \times 10^{-6}I^2 + 0.0009I^2 - 0.0318I - 0.14 )</td>
<td>–1</td>
<td>0.68</td>
<td>0.87</td>
</tr>
<tr>
<td>10-year</td>
<td>( NPD = 2 \times 10^{-3}I^4 - 8 \times 10^{-6}I^2 + 0.001I^2 - 0.0331I - 0.18 )</td>
<td>–1</td>
<td>0.42</td>
<td>0.90</td>
</tr>
<tr>
<td>25-year</td>
<td>( NPD = 2 \times 10^{-3}I^4 - 8 \times 10^{-6}I^2 + 0.0013I^2 - 0.0338I - 0.14 )</td>
<td>–1</td>
<td>0.19</td>
<td>0.91</td>
</tr>
<tr>
<td>50-year</td>
<td>( NPD = 2 \times 10^{-3}I^4 - 8 \times 10^{-6}I^2 + 0.0014I^2 - 0.0346I - 0.23 )</td>
<td>–1</td>
<td>0.05</td>
<td>0.92</td>
</tr>
<tr>
<td>100-year</td>
<td>( NPD = 2 \times 10^{-3}I^4 - 8 \times 10^{-6}I^2 + 0.0016I^2 - 0.0359I - 0.24 )</td>
<td>–1</td>
<td>–0.1</td>
<td>0.93</td>
</tr>
</tbody>
</table>

\( I \) — days from the beginning of study year.

**Figure 3** | Supplied water by irrigation and rainfall in the return periods.

Soil moisture index (SMI) was defined to indicate the effect of drought parameters on providing the crop water requirement in the root zone area. SMI is defined as follows:

\[
SMI_i = \frac{SMC_i - SMC_{pwp}}{SMC_s - SMC_{pwp}}
\]  

(18)

**Figure 4** | Ranges of soil moisture content with/without irrigation in the wheat root zone for a growing season.
where \( \text{SMC}_i \) = soil moisture content in day \( i \) (mm), \( \text{SMC}_s \) = soil moisture content in field capacity point (mm), and \( \text{SMC}_{pwp} \) = soil moisture content in permanent wilting point (mm). As shown in Figure 4, probabilistic values of SMI in the root zone illustrated for wheat are improved by irrigation to increase the biomass production. The ranges of variation for each scenario show the uncertainty amounts in the water availability.

**CONCLUSION**

Considering the rainfall characteristics is one of the key components in developing a decision-making system to assess the effects of drought on agriculture. Variables including duration and severity of drought are determined from long-term recorded precipitation depth information. In this study, the effect of duration and severity of drought characteristics on the estimation of root zone moisture was evaluated to determine probabilistic estimates for irrigation planning. The results showed that in an interconnected programming system, metaheuristic algorithms can be used to fit the coefficients of probabilistic equations and marginal functions can be incorporated directly to determine the bivariate return period. The \( t \) Copula joint function with Pearson correlation coefficient of 0.62 was the first choice for predicting drought characteristics. As obtained results, the application of the probabilistic model in irrigation planning was able to introduce a component into the decision-making system that provides predictive capabilities for calculating water use efficiency. Future studies could focus on other characteristics of drought in agricultural water supply and how to use them in water management.

**DATA AVAILABILITY STATEMENT**

All relevant data are included in the paper or its Supplementary Information.

**REFERENCES**


Received 22 January 2021; accepted in revised form 21 May 2021. Available online 4 June 2021