Copula-based joint probability distribution of water supply and demand in Luhun irrigation district

Jinping Zhang, Xixi Shi, Jian Li and Fawen Li

ABSTRACT

The potential of the copula method to construct the joint probability distribution of three hydrological variables characterizing water supply and demand (WSD) is explored for the Luhun irrigation district of China. The marginal distributions of rainfall, reference crop evapotranspiration (ET_0) and irrigation water are simulated by the corresponding best-fitting cumulative distribution functions. Furthermore, the correlations between every pair of variables are quantified. On this basis, the two-dimensional joint distributions of rainfall and (ET_0) (representing natural WSD), and irrigation water and (ET_0) (representing man-made WSD), and the three-dimensional joint distribution of rainfall, irrigation water, and (ET_0) (representing natural–man-made WSD) are established. The results reveal that the best-fitting marginal distributions for rainfall and (ET_0) and irrigation water are the normal distribution and the Weibull distribution. Moreover, for rainfall and (ET_0), the Student's *t* copula is applied to obtain the joint distribution, while the corresponding copula for (ET_0) and irrigation water is the Clayton copula. Finally, the three-dimensional Student's *t* copula is selected to explore the dependence structure among rainfall, irrigation water, and (ET_0). Therefore, these joint distributions provide an efficient approach to assess water shortage risks in the irrigation district.

Key words | copula function, irrigation water, rainfall, reference crop evapotranspiration

Jinping Zhang
Xixi Shi
Jian Li (corresponding author)
College of Water Conservancy and Environment, Zhengzhou University, Zhengzhou 450001, China
E-mail: *lijian_6plus@163.com*

Fawen Li
State Key Laboratory of Hydraulic Engineering Simulation and Safety, Tianjin University, Tianjin 300072, China

INTRODUCTION

The water supply to an irrigation district comprises rainfall and irrigation water. Rainfall can be called natural water, while irrigation water may be denoted as man-made water. Rainfall and irrigation water are consumed by crops to meet their water requirement. Typically, for crops in an irrigation district, reference crop evapotranspiration (ET_0) indicates the crop water demand. Affected, Rainfall and ET_0 are random variables as they are affected by the meteorological system and underlying surface conditions. From the perspective of hydrostatistics, irrigation water that represents an artificial water resource in an irrigation district can also be considered as a random variable. In order to describe the associated dependence structure among water supply and demand (WSD) in an irrigation district, Ding *et al.* (2011) and Zhang *et al.* (2013a, 2013b) used the Frank copula to construct the joint probability of rainfall and ET_0 in an irrigation district. Zhang *et al.* (2016) indicated that it is feasible to apply the Student's *t* copula for the statistical analyses of natural–man-made water supply and water demand in an irrigation district, using the data series of rainfall, ET_0, and irrigation water.

Currently, extensively applied copula functions involve Symmetry Archimedean copula (Song *et al.* 2012) and Elliptical copula (Genest *et al.* 2007), which can reveal the dependence structure by connecting the univariate marginal distributions with the multivariate joint distribution. Notably, copula functions can be used flexibly because all variables need not obey the same marginal probability

doi: 10.2166/ws.2018.144

distribution or need not be transformed to follow the normal distribution (Zhang & Singh 2007).

In the context of joint probability distribution (JPD), the simplicity and capacity of the One-parameter Symmetry Archimedean copula (Song et al. 2012) and Elliptical copula (Genest et al. 2007) are evident from studies on the probabilistic characterization of drought (Guttman 1998; Dalezios et al. 2000; Shiau 2006; Kao & Govindaraju 2010; Xu et al. 2015), sea storm analysis (Michele et al. 2007; Corbella & Stretch 2013; Montes-Iturrizaga & Heredia-Zavoni 2015), flood risk analysis (Fu & Butler 2014; Masina et al. 2015; Yang et al. 2017), runoff and sediment (Xiong et al. 2014; Zhou et al. 2014; Huang et al. 2017; Wang et al. 2017) and streamflow simulation (Chen et al. 2015; Jeong & Lee 2015; Liu et al. 2017). Zhang et al. (2013a, 2013b) and Fan et al. (2016) provided a detailed theoretical background and methodological descriptions for applying copula to hydrology and water resources. Similarly, in this paper, the JPDs of the natural, man-made and natural–man-made water supply and demand for an irrigation district are explored using Symmetry Archimedean copula and Elliptical copula.

The main objective of this study is to explore the most appropriate copula function to establish the multivariate JPDs of WSD under different water supply conditions using the data series of rainfall, ET_0, and irrigation water from 1970 to 2013 in the Luhun irrigation district of Henan Province, China. The paper is structured as follows: the section titled 'Methods' introduces eight cumulative probability distribution (CPD) functions as well as the copula methods. The data series of rainfall, ET_0, and irrigation water from 1970 to 2013 in the Luhun irrigation district are presented in 'Data series'. The section titled 'Results and analysis' estimates the univariate marginal distributions, and the JPDs of WSD in the irrigation district under different water supply conditions. The final section presents the conclusions.

METHODS

Cumulative distribution function

The most popular eight CPD functions that have been widely applied for hydrological variables include the normal, exponential, two-parameter gamma, Pearson III Type (P-III), generalized extreme value (GEV), Generalized Pareto (GP), lognormal and Weibull distributions (Song et al. 2012).

Correlation of variables

Prior to introducing the copula method, the correlation of variables needs to be estimated. The three most widely used coefficients are Pearson's correlation coefficient γ, Kendall's correlation coefficient τ, and Spearman's correlation coefficient ρ.

Pearson's correlation coefficient γ denotes the linear relationship between random variables. Usually, $\gamma \in [-1, 1]$. If the absolute value of γ is higher, the linear relationship between variables is more significant. The Pearson's correlation coefficient γ can be calculated by (Song et al. 2012):

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{1}$$

where (x_i, y_i) is the sample, $i = 1,2, \ldots, n$; $\bar{x}$, $\bar{y}$ are the mean values of x and y, respectively; and n is the total number of observations.

Kendall's rank correlation coefficient τ is used to measure both the linear and nonlinear correlation of variables, and it can be calculated from the following expression (Song et al. 2012):

$$\tau = (C_n^k)^{-1} \sum_{i<j} sign[(x_i - x_j)(y_i - y_j)] \tag{2}$$

where C_n^k denotes 'n choose k', n is the total number of observations (x_i, y_i); and i, $j = 1,2,3, \ldots, n$. The $sign(\bullet)$ denotes the symbol function, which is defined as

$$sign[(x_i - x_j)(y_i - y_j)] = \begin{cases} 1 \cdots\cdots (x_i - x_j)(y_i - y_j) > 0 \\ 0 \cdots\cdots (x_i - x_j)(y_i - y_j) = 0 \\ -1 \cdots\cdots (x_i - x_j)(y_i - y_j) < 0 \end{cases} \tag{3}$$

Spearman's correlation coefficient ρ can be computed by (Song et al. 2012):

$$\rho = \frac{\sum_{i=1}^{n}(R_i - \overline{R})(S_i - \overline{S})}{\sqrt{\sum_{i=1}^{n}(R_i - \overline{R})^2}\sqrt{\sum_{i=1}^{n}(S_i - \overline{S})^2}} \quad (4)$$

where n is the sample size; R_i, S_i are the ranks of x and y, respectively; $\overline{R}$, $\overline{S}$ are the means of R_i and S_i, respectively; and $i = 1, 2, \ldots, n$.

Copula method

According to the Sklar theorem proposed in 1959 (Nelsen 2006), copulas are functions that characterize the dependence structure among variables by linking univariate marginal distributions to form multivariate joint distribution functions. If H is an n-dimensional distribution function of n-dimensional random variables $X_1, X_2, \cdots, X_n$ defined in R, and $F_1(x_1), F_2(x_2), \cdots, F_n(x_n)$ are their marginal distributions respectively, for any $x_i \in R_n$ (x_i is a specific value of X_i, $i = 1, 2, \ldots$, n) an n-dimensional copula function $C()$ exists and can be expressed as follows:

$$\begin{aligned} H(x_1, x_2, \cdots, x_n) &= C_\theta(F_1(x_1), F_2(x_2), \cdots, F_n(x_n)) \\ &= C(u_1, u_2, \cdots, u_n) \end{aligned} \quad (5)$$

where n is the number of variables, $u_i = F_i(x_i)$ ($i = 1, 2, \ldots$, n), $C()$ is the copula function, and θ is the parameter of interest. If $F_i(x_i)$ is continuous and U_i ($i = 1, 2, \ldots$, n) subjects to the uniform distribution on [0, 1], C is uniquely determined.

Currently, three types of Symmetry Archimedean copulas are often applied in the hydrology and water resources fields (Song *et al.* 2012).The Gaussian and Student's t copulas are commonly used Elliptical copulas in these contexts. The bivariate and trivariate Gaussian copula functions are given by Equation (6) and Equation (7), respectively:

$$C(u_1, u_2; \Sigma) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{(2\pi)|\Sigma|^{(1/2)}} \exp\left(-\frac{1}{2}\mathbf{w}^T\Sigma^{-1}\mathbf{w}\right) d\mathbf{w} \quad (6)$$

$$\begin{aligned} &C(u_1, u_2, u_3; \Sigma) \\ &= \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \int_{-\infty}^{\Phi^{-1}(u_3)} \frac{1}{(2\pi)^{\frac{3}{2}}|\Sigma|^{(1/2)}} \exp\left(-\frac{1}{2}\mathbf{w}^T\Sigma^{-1}\mathbf{w}\right) d\mathbf{w} \end{aligned} \quad (7)$$

where $u_i (i = 1, 2, 3)$ is the marginal function of the variables; $\Sigma = \begin{bmatrix} 1 & \cdots & \rho_{1d} \\ \vdots & \cdots & \vdots \\ \rho_{d1} & \cdots & 1 \end{bmatrix}$, $\rho_{ij} = \begin{cases} 1; & i = j \\ \rho_{ji}; & i \neq j \end{cases}$, $-1 \leq \rho_{ij} \leq 1$, and ρ_{ij} is estimated using the maximum likelihood estimation method. d refers to the variable's dimensions; $\Phi^{-1}(.)$ is the inverse function of the standard normal distribution; w is the integrand variable matrix, and $\mathbf{w} = [w_1, w_2, \cdots, w_d]^T$.

The bivariate and trivariate Student's t-copula functions with v-degrees of freedom are expressed with Equation (8) and Equation (9) respectively:

$$\begin{aligned} C(u_1, u_2; \Sigma, \upsilon) = &\int_{-\infty}^{T_\upsilon^{-1}(u_1)} \int_{-\infty}^{T_\upsilon^{-1}(u_2)} \frac{\Gamma((\upsilon + d/2))}{\Gamma\left(\frac{\upsilon}{2}\right)} \frac{1}{(\pi\upsilon)^{(d/2)}|\Sigma|^{(1/2)}} \\ &\times \left(1 + \frac{\mathbf{w}^T\Sigma^{-1}\mathbf{w}}{\upsilon}\right)^{-(\upsilon + d/2)} d\mathbf{w} \end{aligned} \quad (8)$$

$$\begin{aligned} &C(u_1, u_2, u_3; \Sigma, \upsilon) \\ &= \int_{-\infty}^{T_\upsilon^{-1}(u_1)} \int_{-\infty}^{T_\upsilon^{-1}(u_2)} \int_{-\infty}^{T_\upsilon^{-1}(u_3)} \frac{\Gamma(\upsilon + d/2)}{\Gamma\left(\frac{\upsilon}{2}\right)} \frac{1}{(\pi\upsilon)^{(d/2)}|\Sigma|^{(1/2)}} \\ &\times \left(1 + \frac{\mathbf{w}^T\Sigma^{-1}\mathbf{w}}{\upsilon}\right)^{-(\upsilon + d/2)} d\mathbf{w} \end{aligned} \quad (9)$$

where υ is the degree of freedom of the Student's t copula, and $T_\upsilon^{-1}(\cdot)$ denotes the inverse function of Student's t distribution with v-degrees of freedom.

Identification and goodness-of-fit evaluation

For the bivariate variables (X, Y), the nonparametric Kolmogorov-Smirnov (K–S) test is employed to identify the copula, and the identified statistic D is written as

$$D = \max_{1 \leq k \leq n}\left\{\left|C_k - \frac{m_k}{n}\right|, \left|C_k - \frac{m_k - 1}{n}\right|\right\} \quad (10)$$

where n is the sample size, C_k is the value of the copula function of the observed (x_k, y_k) and m_k is the number of observed (x_k, y_k) such that $x \leq x_k$ and $y \leq y_k$.

The goodness-of-fit of the copula is estimated by employing Ordinary Least Squares (*OLS*), BIAS and the Akaike information criterion (*AIC*). *OLS* is expressed as

$$OLS = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Pe_i - P_i)^2} \quad (11)$$

where n is the sample size, Pe_i is the empirical frequency of the joint probability, and P_i is the frequency of the joint probability.

BIAS can be used as an indicator to understand if the selected model offers overall under- or overpredictions. It is expressed as (Mohammad *et al.* 2018; Mohammad & Saeed 2018):

$$BIAS = \frac{\sum_{i=1}^{n}(Pe_i - P_i)}{n} \quad (12)$$

AIC is given by:

$$AIC = n\ln\left(\frac{RSS}{n}\right) + 2k \quad (13)$$

where k is the number of parameters, n is the sample size, and *RSS* is the residual sum of squares:

$$RSS = \sum_{i=1}^{n}(Pe_i - P_i)^2 \quad (14)$$

RESULTS AND ANALYSIS

Data series

The Luhun irrigation district (112°5′-113°5′E, 34°10′-34°45′N) is located in the hilly area of western Henan Province, China. It comprises a total irrigation area of 1,838.48 km^2, as shown in Figure 1 (Zhang *et al.* 2018). Its annual average rainfall is only 611.02 mm but the annual evaporation reaches 1,034.32 mm, making the area prone to meteorological droughts. The volume of annual irrigation water for the Luhun irrigation district is $1.82 \times 10^8\ m^3$. The water is sourced from the Luhun reservoir built on the upper reaches of the Yihe River, which is the secondary tributary of the Yellow River.

The researched data series for the Luhun irrigation district comprise the daily meteorological data (including rainfall, wind speed, maximum temperature, minimum temperature, relative humidity, net radiation, etc.) and irrigation water data from 1970 to 2013. The Penman–Monteith formula (Luo *et al.* 2008) is used to calculate ET_0. The annual data series of rainfall, ET_0, and irrigation water in the Luhun irrigation district appear in Table 1, while their statistical characteristics are provided in Table 2.

It can be seen that: (1) the inter-annual variation of rainfall, ET_0, and irrigation water are all uneven; (2) the volume for ET_0 exceeds that for rainfall, which means more irrigation water is needed in the irrigation district; and (3) typically, ET_0 has a negative correlation with rainfall but a positive correlation exists with the irrigation water. The statistical tests on stationarity, consistency and randomness were conducted as described by Li (2016).

Marginal distributions of rainfall, ET_0, and irrigation water

The marginal distributions of rainfall, ET_0, and irrigation water are fitted by the aforementioned eight distributions. Using the maximum likelihood method, the parameters of the eight distributions are estimated, and the results are exhibited in Table 3.

The K–S test is used to identify the eight marginal distributions. Assuming a significance level of 95%, the corresponding fractile value is 0.20056. Table 4 gives the goodness-of-fit evaluations of the eight marginal distributions for rainfall, ET_0, and irrigation water. Irrespective of the variable (rainfall, ET_0 or irrigation water), the values of the K–S statistics D of the exponential and GP distributions exceed 0.20056, and hence the exponential and GP distributions are inappropriate for the situation at hand.

The best-fitted marginal distributions are verified and determined according to the value of the *OLS* and K–S test (as shown in Table 4). From Table 4, it is evident that the normal distribution presents the lowest value of *OLS* for rainfall, and same is true for the Weibull distribution for ET_0 and irrigation water. The result indicates that the normal distribution is the best-fitted marginal distribution

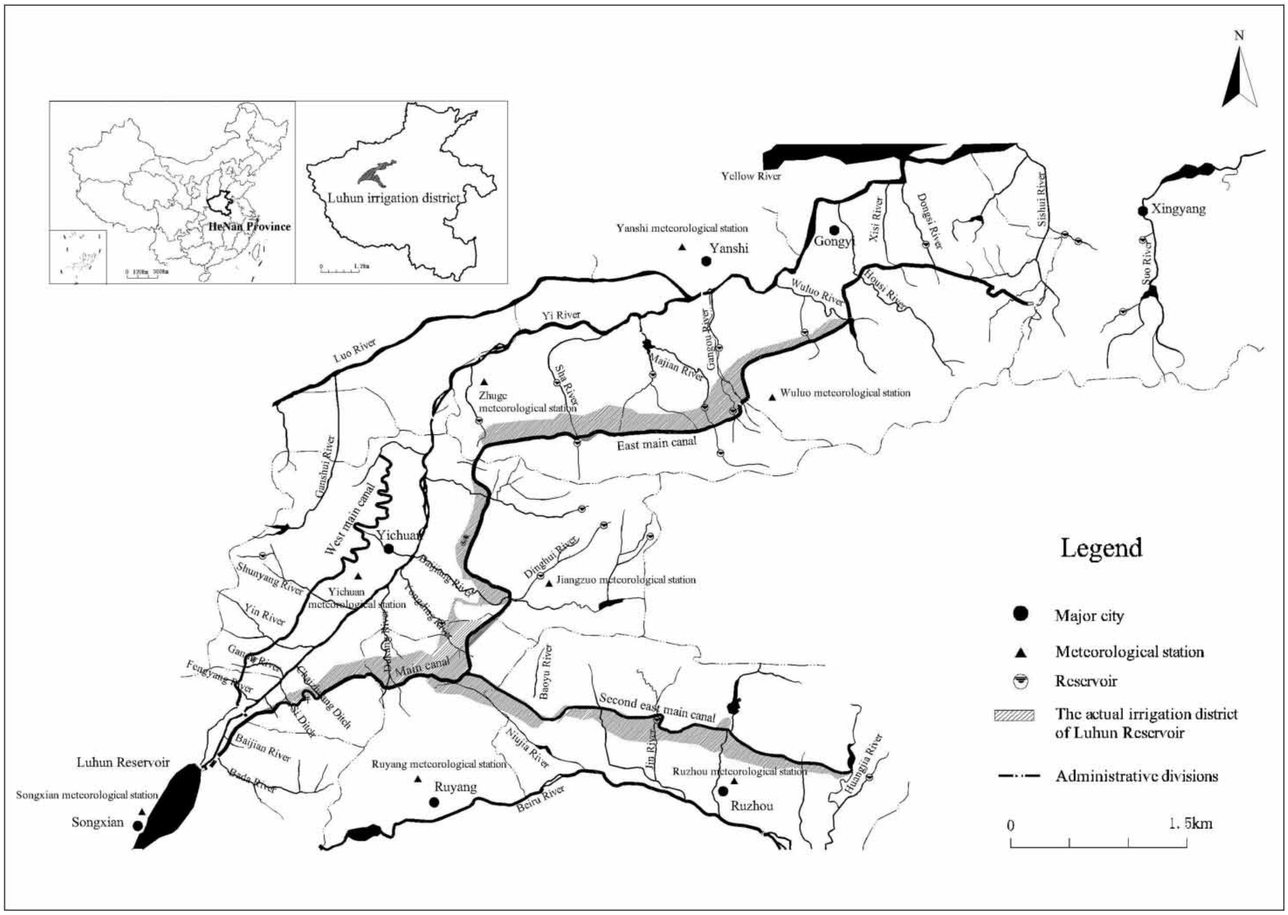

Figure 1 | Location map of the Luhun irrigation district.

for rainfall, while the Weibull distribution is the best one for ET_0 and irrigation water. Let $F(x)$, $F(y)$ and $F(z)$ be the cumulative distribution functions of rainfall, ET_0, and irrigation water, respectively, and the selected marginal distribution functions for rainfall, ET_0, and irrigation water are expressed as follows:

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{(x - 611.0227)^2}{37669.4156}\right) dx, -\infty < x < \infty \tag{15}$$

$$F(y) = 1 - \exp\left(-\left(\frac{y}{1066.4}\right)^{20.9}\right) \tag{16}$$

$$F(z) = 1 - \exp\left(-\left(\frac{z}{2.0357}\right)^{2.9975}\right) \tag{17}$$

The comparisons of empirical probability with calculated probability based on the selected marginal distributions for rainfall, ET_0, and irrigation water are displayed in Figure 2. All the correlation coefficients of the empirical probability distributions and calculated probability distributions exceed 0.97, indicating that the selected marginal distributions are suitable for application in the given context.

Correlation between rainfall, ET_0, and irrigation water

The correlation coefficients between rainfall, ET_0, and irrigation water in the Luhun irrigation district are presented in Table 5. It can be seen that a negative correlation exists not only between rainfall and ET_0 but also between ET_0 and irrigation water, while a positive

Table 1 | Data series of rainfall, ET_0 and irrigation water in Luhun irrigation district

Year	Rainfall (mm)	ET_0 (mm)	Irrigation water (10^8 m^3)	Year	Rainfall (mm)	ET_0 (mm)	Irrigation water (10^8 m^3)
1970	557.850	1,067.814	2.012	1992	679.100	1,009.013	1.699
1971	557.650	1,103.150	1.996	1993	559.000	972.071	1.225
1972	521.750	1,081.929	2.108	1994	718.900	1,052.835	1.596
1973	748.050	1,065.220	1.962	1995	548.600	1,072.172	2.489
1974	648.850	1,102.339	2.121	1996	628.300	981.695	1.375
1975	579.050	1,066.823	0.705	1997	380.600	1,077.800	2.886
1976	579.200	1,057.709	1.219	1998	781.800	993.259	1.203
1977	575.400	1,087.760	1.590	1999	616.900	1,043.432	1.679
1978	471.700	1,116.644	2.189	2000	637.100	1,035.435	1.404
1979	568.250	1,073.022	2.117	2001	401.800	1,109.791	2.430
1980	663.050	995.547	2.976	2002	599.300	1074.932	1.964
1981	456.350	1,073.982	2.900	2003	953.900	903.766	0.720
1982	611.050	1,013.299	2.821	2004	767.400	1,045.761	2.020
1983	890.000	990.806	1.000	2005	728.800	1,060.298	2.450
1984	836.900	942.560	1.102	2006	692.600	1,059.791	1.240
1985	666.950	971.262	1.710	2007	596.400	1,071.822	1.733
1986	382.750	1,059.945	2.538	2008	658.200	1,080.310	2.550
1987	571.000	1,008.485	0.702	2009	762.500	1,086.360	1.650
1988	580.400	989.792	1.239	2010	600.300	1,092.342	1.820
1989	551.400	920.345	1.669	2011	706.500	1,101.749	1.780
1990	702.550	911.865	0.603	2012	498.700	1,168.252	1.997
1991	353.750	944.680	3.290	2013	294.400	1,017.433	2.372

correlation exists only between ET_0 and irrigation water. Moreover, a strong correlation exists between every pair of variables. Thus, the copula method can be applied to the given situation.

Table 2 | Statistical characteristics of rainfall, ET_0, and irrigation water

Statistical characteristics \ Variable	Rainfall (mm)	ET_0 (mm)	Irrigation water (10^8 m^3)
Mean	611.020	1,039.890	1.815
Maximum	953.900	1,168.250	3.290
Minimum	294.400	903.770	0.603
Standard deviation	135.668	58.996	0.651
Skewness	0.024	−0.574	0.084
Kurtosis	0.435	−0.110	−0.469

The JPD of WSD under different water supply conditions

Under the natural water supply condition

The copula method is applied to attain the JPDs of rainfall and ET_0 for the Luhun irrigation district. Because of the negative correlation between rainfall and ET_0, the Frank, Gaussian and Student's t copulas can be used as the copula functions. The parameter θ, Kendall's rank correlation coefficient τ, and the identified statistics of the Frank, Gaussian and Student's t copulas are estimated in Table 6. R^2 stands for the correlation coefficients of the empirical probability distributions and calculated probability distributions.

These three copulas are identified using K–S test statistic D. Taking the significance level as $\alpha = 0.05$, when $n = 44$, the corresponding fractile value of the K–S statistic is

Table 3 | Parameters of the eight marginal distributions of the three variables

Distribution and parameters (Variable)		Rainfall (mm)	ET_0 (mm)	Irrigation water (10^8 m^3)
Normal	μ	611.023	1,039.900	1.815
	σ	137.237	59.678	0.676
Exponential	α	611.023	1,039.900	1.815
Gamma	β	18.823	303.095	6.451
	α	32.462	3.431	0.281
P-III	C_v	0.240	0.060	0.350
	C_s	0.030	0.010	0.090
GEV	μ	561.679	1,021.700	1.576
	k	−0.259	−0.400	−0.273
	α			
GP	k	135.935	62.500	0.656
	α	−1.200	−1.700	−1.023
Weibull	a	1,148.300	2,019.200	3.366
	b	664.660	1,066.400	2.036
Lognormal	μ	4.914	20.900	2.998
	σ			

$D_0 = 0.20056$. Evidently, all three copulas pass the K–S test (Table 3). Although the values of *OLS* are quite similar, the value of *AIC* of the Student's *t* copula is the least, and so the Student's *t* copula is the best option to describe the dependence structure of rainfall and ET_0. Besides, Student's *t* copula also has the highest accuracy in terms of BIAS = –0.0164 compared to the other models.

Additionally, it can be seen that the Student's *t* copula has higher correlation between the computed copula and the empirical copulas compared with the other two copulas, which confirms that the Student's *t* copula is the most suitable option in the context. Using the Student's *t* copula, the JPD of rainfall and ET_0 is obtained as a surface plot in Figure 3(a1) and a contour plot in Figure 3(b1).

Under the man-made water supply condition

In order to construct the JPD of ET_0 and irrigation water, five bivariate copulas– that is, the Frank, Clayton, Gumbel, Gaussian, and Student's *t* copulas– are chosen to depict the dependence structure between the two variables.

Using the previous formulas, the parameters and K–S test statistic D are estimated, and the goodness-of-fit is assessed by using *OLS*, *BIAS* and *AIC* (shown in Table 7). It can be seen that the statistic D is lower than the fractile value $D_0 = 0.20056$, which indicates that the chosen five copulas pass the K–S test and thus, they can all be accepted. However, the Clayton copula is considered as the most appropriate one because it provides the lowest values of *OLS* and *AIC*. Besides, the Clayton copula also has the highest accuracy in terms of BIAS = −0.0126 compared to other models. In addition, it can be observed that the Clayton copula performs better, as evidenced by the higher correlation coefficient (R^2) between the empirical copula and the calculated copula. Thus, it is reasonable that the Clayton copula be selected to describe the JPD of ET_0 and irrigation water as a surface plot, displayed in Figure 3(a2), and a contour plot in Figure 3(b2).

Under the natural–man-made WSD condition

The JPDs of rainfall, ET_0, and irrigation water are established using the Frank, Clayton, Gumbel, Gaussian, and Student's *t* copulas. Next, the corresponding parameters are estimated using the maximum likelihood method. Then, the best-fitted joint CPDs are identified as follows. Firstly, the K–S test is used to obtain the available copula functions. Secondly, the minimum deviation

Table 4 | Statistical evaluations for the marginal distribution functions of the three variables

Variable (Distribution)	Statistical evaluation	Gamma	Lognormal	GP	Exponential	P-III	Normal	GEV	Weibull
Rainfall (mm)	K–S	0.147	0.164	0.304	0.418	0.132	0.120	0.128	0.118
	OLS	0.045	0.053	–	–	0.045	0.039	0.042	0.045
ET_0 (mm)	K–S	0.169	0.171	0.569	0.581	0.158	0.163	0.140	0.115
	OLS	0.067	0.068	-	-	0.063	0.064	0.055	0.045
Irrigation water (10^8 m^3)	K–S	0.107	0.133	0.201	0.326	0.088	0.075	0.076	0.070
	OLS	0.082	0.049	0.100	-	0.034	0.030	0.031	0.029

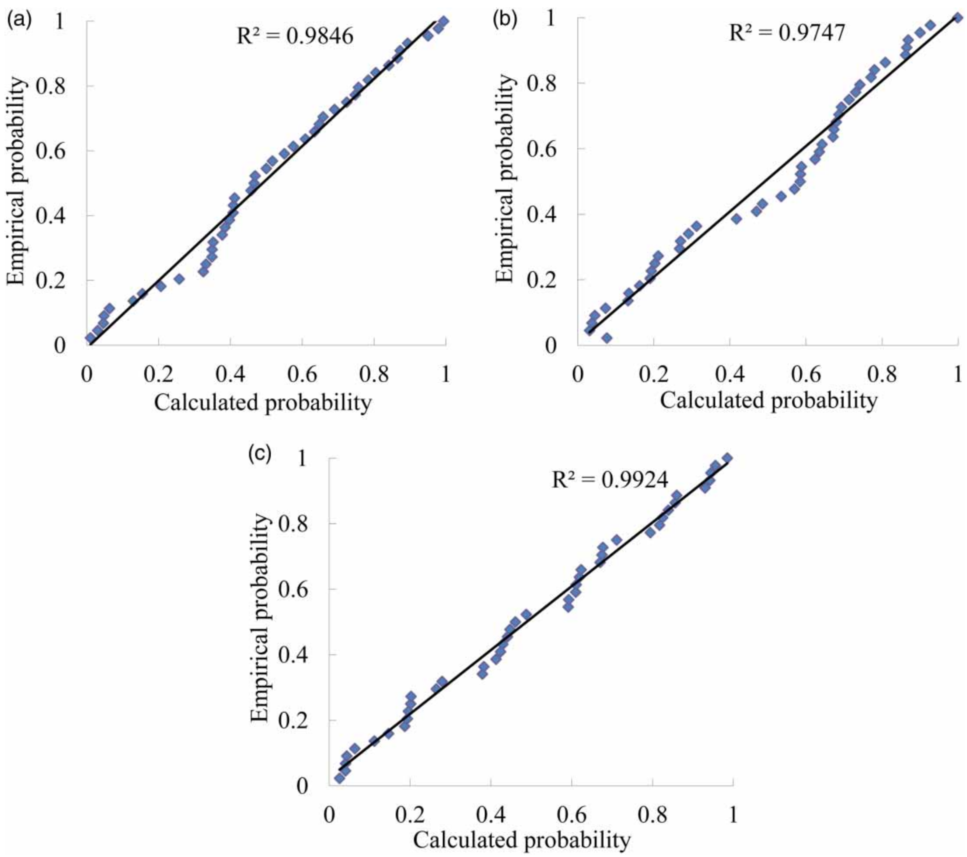

Figure 2 | Comparison of empirical probability with calculated probability based on the selected marginal distribution for rainfall (a), ET_0 (b), and irrigation water (c).

squares and AIC are employed to determine the best-fitted joint distribution. Table 8 shows the results of the parameters θ, K–S test statistic *D*, *OLS* and *AIC*. The findings reveal that: (1) all the values of K–S test statistic D are less than D_0 (0.20056), which illustrates that all the tested copula functions are appropriate to establish the dependence structure among rainfall, ET_0, and irrigation water; (2) for the three-dimensional joint distribution, the Student's *t* copula is the most suitable as it provides the lowest values of *OLS*, *AIC*, from among all copula functions. Besides, Student's *t* copula also has the highest accuracy in terms of BIAS = −0.0131 compared to other models.

Furthermore, the Student's *t* copula has a higher correlation coefficient (R^2), as exhibited in Table 5. Therefore, the Student's *t* copula can best characterize the dependence

Table 5 | Correlation coefficients between rainfall, ET_0, and irrigation water

Correlation coefficient	Rainfall and ET_0	Rainfall and irrigation	ET_0 and irrigation water
r	−0.3244	−0.5364	0.3225
τ	−0.2389	−0.3383	0.2622
ρ	−0.3353	−0.4698	0.3875

Table 6 | Parameter θ, Kendall's rank correlation coefficient τ, and identified statistics of the Frank, Gaussian, and Student's *t* copulas for rainfall and ET_0

Kendall's τ	Parameters or identified statistics	Copula functions		
		Frank	Gaussian	Student's *t*
− 0.2389	θ	−2.2559	−0.3217	$\rho = -0.3322$ $\upsilon = 9.5928$
	D	0.0969	0.1068	0.1042
	OLS	0.0424	0.0425	0.0424
	BIAS	−0.0174	−0.0168	−0.0164
	AIC	−276.0700	−275.9600	−276.1000
	R^2	0.9435	0.9418	0.9428

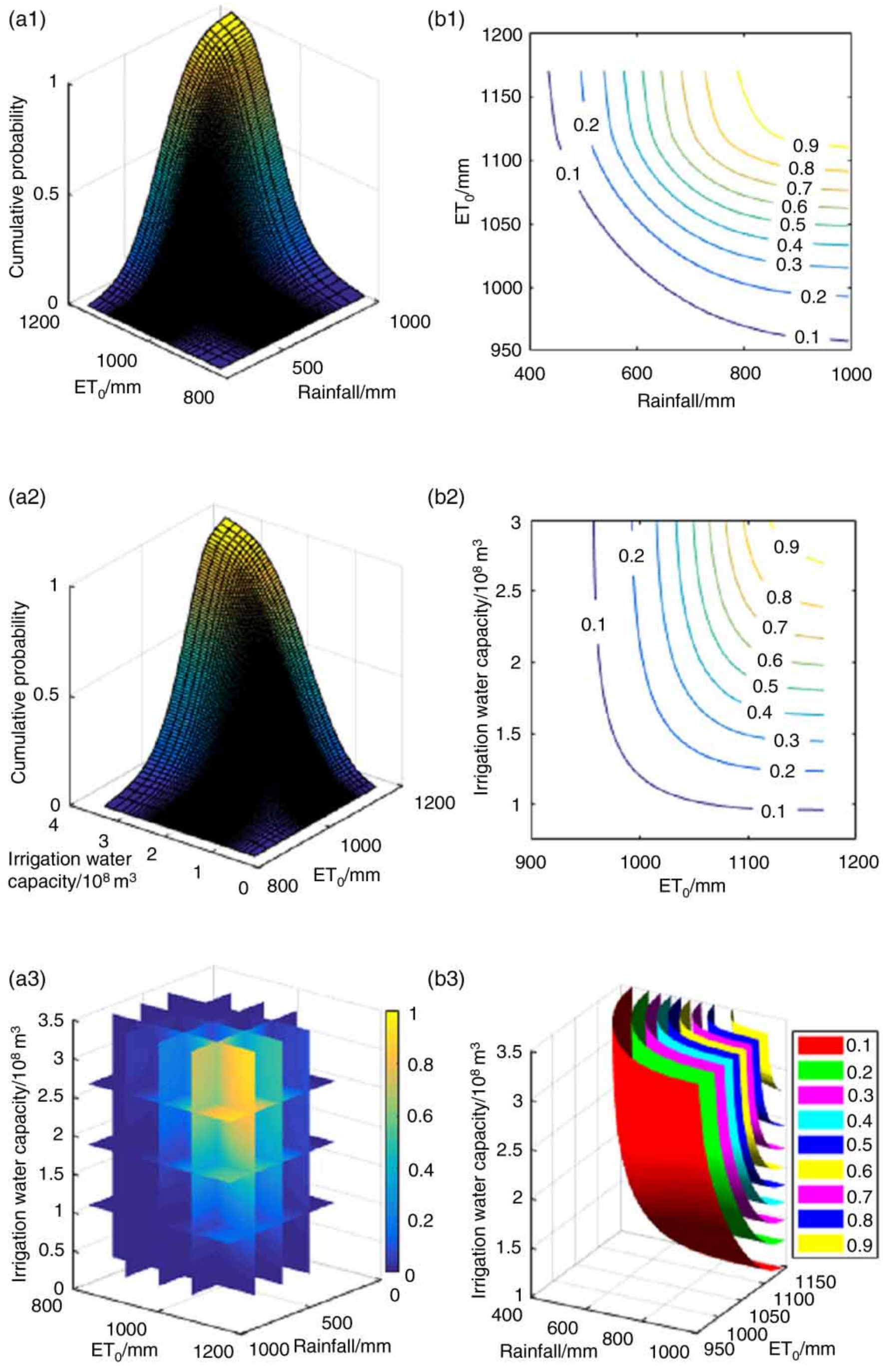

Figure 3 | Surface plot (a) and contour plot (b) of the joint cumulative probability of rainfall, ET_0 and irrigation water in the Luhun irrigation district.

structure and should be used to construct the joint distribution slices plot, as exhibited in Figure 3(a3) and surface plot, as seen in Figure 3(b3).

For IR representing irrigation water, compared with the bivariate JPD of (rainfall, ET_0) (Figure 3(a1) and 3(b1)) and (ET_0, IR) (Figure 3(a2) and 3(b2)), the trivariate JPD

Table 7 | Kendall coefficient and parameters of the copula functions for ET_0 and irrigation water

		Copula functions				
Kendall's τ	**Parameter**	**Frank**	**Clayton**	**Gumbel**	**Gaussian**	**Student's *t***
0.2622	θ	2.5015	0.7108	1.3554	0.3048	$\rho = 0.3832$ $\upsilon = 4.5056$
	D	0.1053	0.0937	0.1129	0.0972	0.1047
	OLS	0.0383	0.0349	0.0402	0.0413	0.0380
	BIAS	−0.0158	−0.0126	−0.0161	−0.0162	−0.0158
	AIC	−285.1108	−293.2372	−280.8594	−278.4447	−285.7734
	R^2	0.9811	0.9798	0.9798	0.9806	0.9805

Table 8 | Parameter θ, OLS and AIC of the three-dimensional copula distributions for rainfall, ET_0 and irrigation water

	Copula functions				
Parameters	**Frank**	**Gumbel**	**Clayton**	**Gaussian**	**Student's *t***
θ	−0.7929	0.9938	−0.0438	$\rho_{12} = -0.3217 \rho_{13} = -0.5320$ $\rho_{23} = 0.3048$	$\rho_{12} = -0.3404 \rho_{13} = -0.5209$ $\rho_{23} = 0.3391$ $\upsilon = 17.9123$
D	0.1361	0.1604	0.1560	0.1177	0.1137
OLS	0.0452	0.0500	0.0489	0.0410	0.0404
BIAS	−0.0135	−0.0130	−0.0152	−0.0135	−0.0131
AIC	−270.4200	−261.6200	−263.5200	−275.0900	−276.3100
R^2	0.8668	0.8687	0.8191	0.8084	0.8108

of (rainfall, ET_0, IR) (Figure 3(a3) and 3(b3)) is quite different. The superposition from the bivariate JPD of (rainfall, ET_0) and (ET_0, IR) to the trivariate JPD of (rainfall, ET_0, IR), is not simple. For example, it can be seen from Figure 3(b1) that when the value of JPD is between 0.1 and 0.3, the contour plot is relatively sparse and (rainfall, ET_0) change more obviously compared with that between 0.3 and 0.7; as for (ET_0, IR) in Figure 3(b2), the value of JPD is relatively sparse from 0.1 to 0.4 and (ET_0,IR) change more obviously compared with that between 0.4 and 0.7. For the trivariate JPD of pair (rainfall, ET_0, IR), when the value of JPD is between 0.1 and 0.2 or 0.8 and 0.9, the contour surface has larger spacing than that for JPD between 0.2 and 0.8. Thus, we conclude that considering the trivariate JPD can lead to a more accurate result.

The joint return period can be obtained with the constructed JPD of rainfall, ET_0 and irrigation water. Moreover, the encountered events of (rainfall,ET_0, IR) can be obtained under the same return period. Thus, the exceeded risk of any given planning indicator in the encountered events can be calculated, and the risk scope of a planning indicator corresponding to a return period can also be provided. For example, if an appropriate distribution of (rainfall, ET_0, IR) in a typical year is determined, different groups of designed rainfall, ET_0 and irrigation water surfaces can be obtained simultaneously using the same frequency amplification method. This understanding will provide an added guidance for irrigation planning.

CONCLUSIONS

The study describes the joint behaviour of three hydrological variables (rainfall, ET_0 and irrigation water) as a 44-year (1970–2013) time series of WSD in the Luhun irrigation district. In reality, the length of the time series of the three hydrological variables (rainfall, ET_0, and irrigation water) is only 44 years, and the results would be more accurate if the time series were expanded. However, the results with the 44-year time series are still feasible and can be used to explain the statistical characteristics of rainfall, ET_0, and irrigation water. The joint behaviours of these hydrological variables were investigated using the copula method, estimated marginal distributions, and their correlations. The findings imply that the trivariate JPD outperforms the bivariate distribution in terms of reflecting the water shortage risk in reality. Among the eight popular CPD functions that were tested, the marginal distribution of rainfall can be best conveyed by a normal distribution, while for ET_0 and irrigation water, the Weibull distribution is the best choice. Moreover, ET_0 exhibits a negative correlation with rainfall, as well as irrigation water. However, a positive correlation exists between ET_0 and irrigation water.

After selection of the Symmetry Archimedean copulas and Elliptical copulas, the joint distributions of the natural WSD (involving rainfall and ET_0) and the natural–man-made WSD (involving rainfall, irrigation water and ET_0) can be adequately described by the Student's *t* copula. However, the dependence structure of ET_0 and irrigation water, representing the man-made water demand and supply, can be well presented by the Clayton copula. This study attempted to identify the best general method for modelling the dependence structure among trivariates and showed that the copula is a robust method for water shortage risk assessment in the irrigation district.

ACKNOWLEDGEMENTS

This research is supported by the National Key R&D Program of China (Grant No. 2018YFC0406501), Outstanding Young Talent Research Fund of Zhengzhou University (Grant No. 1521323002), Program for Innovative Talents (in Science and Technology) at University of Henan Province (Grant No. 18HASTIT014), State Key Laboratory of Hydraulic Engineering Simulation and Safety, Tianjin University (Grant No. HESS-1717), and Foundation for University Key Teacher by Henan Province of China in 2017. We are thankful for the work of editorial office and suggestions from anonymous reviewers.

REFERENCES

Chen, L., Vijay, P. S., Guo, S. L., Zhou, J. Z. & Zhang, J. H. 2015 Copula-based method for multisite monthly and daily streamflow simulation. *Journal of Hydrology* **528**, 369–384.

Corbella, S. & Stretch, D. D. 2013 Simulating a multivariate sea storm using Archimedean copulas. *Coastal Engineering* **76**, 68–78.

Dalezios, N. R., Loukas, A., Vasiliades, L. & Liakopoulos, E. 2000 Severity-duration-frequency analysis of droughts and wet periods in Greece. *Hydrological Sciences Journal* **45** (5), 751–770.

Ding, Z. H. H., He, H. M. & Wang, H. 2011 Joint distribution model of precipitation and reference crop evapotranspiration in irrigation area. *Water Resources and Hydropower Engineering* **7**, 15–18.

Fan, Y. R., Huang, W. W., Huang, G. H., Li, Y. P., Huang, K. & Li, Z. 2016 Hydrologic risk analysis in the Yangtze River basin through coupling Gaussian mixtures into copulas. *Advances in Water Resources* **88**, 170–185.

Fu, G. T. & Butler, D. 2014 Copula-based frequency analysis of overflow and flooding in urban drainage systems. *Journal of Hydrology* **10**, 49–58.

Genest, C., Favre, A.-C., Béliveau, J. & Jacques, C. 2007 Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resources Research* **43** (9), W09401(1–12).

Guttman, N. B. 1998 Comparing the Palmer drought index and the standardized precipitation index. *Journal of the American Water Resources Association* **34** (1), 113–121.

Huang, S. Z., Li, P., Huang, Q. & Leng, G. Y. 2017 Copula-based identification of the non-stationarity of the relation between runoff and sediment load. *International Journal of Sediment Research* **32** (2), 221–230.

Jeong, C. & Lee, T. 2015 Copula-based modeling and stochastic simulation of seasonal intermittent streamflows for arid regions. *Journal of Hydro-Environment Research* **9** (4), 604–613.

Kao, S. & Govindaraju, R. S. 2010 A copula-based joint deficit index for droughts. *Journal of Hydrology* **380** (1–2), 121–134.

Li, Y. Y. 2016 *Research on the Relationship Between Water Supply and Demand of Irrigation Area with Multi-Time Scales*. Master Thesis, Zhengzhou University, Zhengzhou, Henan, China.

Liu, D. F., Wang, D., Singh, V. P., Wang, Y. K., Wu, J. C., Wang, L. C., Zou, X. Q., Chen, X. F. & Chen, X. 2017 Optimal moment determination in POME-Copula based hydrometeorological dependence modelling. *Advances in Water Resources* **105**, 39–50.

Luo, Y. F., Jiao, X. Y. & Peng, S. Z. 2008 Long-term trend of reference evapotranspiration in Liuyuankou irrigation system and its impacts on irrigation water requirements. *Chinese Agricultural Science Bulletin* **24** (9), 480–484.

Masina, M., Lamberti, A. & Archetti, R. 2015 Coastal flooding: a copula based approach for estimating the joint probability of water levels and waves. *Coastal Engineering* **97**, 37–52.

Michele, C. D., Salvadori, G., Passoni, G. & Vezzoli, R. 2007 A multivariate model of sea storms using copulas. *Coastal Engineering* **54** (10), 734–751.

Mohammad, N. & Saeed, S. 2018 Extraction of optimal equations for evaluation of pipeline scour depth due to currents. *Maritime Engineering* **171** (1), 1–34.

Mohammad, N., Mohammad, R. B. & Ali, T. 2018 Prediction of riprap stone size under overtopping flow using data-driven models. *International Journal of River Basin Management* **1571–5124**, 1814–2060.

Montes-Iturrizaga, R. & Heredia-Zavoni, E. 2015 Environmental contours using copulas. *Applied Ocean Research* **52**, 125–139.

Nelsen, R. B. 2006 *An Introduction to Copulas*. Springer, New York, pp. 1–30.

Shiau, J. T. 2006 Fitting drought duration and severity with two-dimensional copulas. *Water Resources Management* **20**, 795–815.

Song, S. B., Cai, H. J., Jin, J. L. & Kang, Y. 2012 Copulas function and its application in hydrology. Science Press, Beijing, China, pp. 55–116.

Wang, X., Zang, N., Liang, P. Y., Cai, Y. P., Li, C. H. & Yang, Z. F. 2017 Identifying priority management intervals of discharge and TN/TP concentration with copula analysis for Miyun Reservoir inflows, North China. *Science of The Total Environment* **609**, 1258–1269.

Xiong, L. H., Yu, K. X. & Lars, G. 2014 Estimation of the distribution of annual runoff from climatic variables using copulas. *Water Resources Research* **50** (9), 7134–7152.

Xu, K., Yang, D. W., Xu, X. Y. & Lei, H. M. 2015 Copula based drought frequency analysis considering the spatio-temporal variability in Southwest China. *Journal of Hydrology* **527**, 630–640.

Yang, P., Chen, K., Yan, H. X., Yu, X. L., Yan, H. X. & Yu, X. L. 2017 Improving flood-risk analysis for confluence flooding control downstream using copula Monte Carlo method. *Journal of Hydrologic Engineering* **22** (8).

Zhang, L. & Singh, V. P. 2007 Bivariate rainfall frequency distributions using Archimedean copulas. *Journal of Hydrology* **332**, 93–109.

Zhang, J. P., Ding, Z. H. H., Yuan, W. L. & Zuo, Q. T. 2013a Research on the relationship between rainfall and reference crop evapotranspiration with multi-time scales. *Paddy and Water Environment* **11** (1–4), 473–482.

Zhang, Q., Xiao, M. Z., Singh, V. P. & Chen, X. H. 2013b Copula-based risk evaluation of hydrological droughts in the East River basin, China. *Stochastic Environmental Research and Risk Assessment* **27** (6), 1397–1406.

Zhang, J. P., Lin, X. M. & Guo, B. T. 2016 Multivariate copula-based joint probability distribution of water supply and demand in irrigation district. *Water Resources Management* **30** (7), 2361–2375.

Zhang, J. P., Lin, X. M. & Zhao, Y. 2018 The joint distribution of reference crop evapotranspiration and irrigation water in the irrigation district. *Water Science & Technology: Water Supply* **18** (2), 567–576.

Zhou, N. Q., Zhao, L. & Shen, X. P. 2014 Copula-based probability evaluation of rich-poor runoff and sediment encounter in Dongting lake basin. *Scientia Geographica Sinica* **34** (2), 242–248.

First received 13 March 2018; accepted in revised form 2 August 2018. Available online 14 August 2018