JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION
Vol. 58, No. 3 AMERICAN WATER RESOURCES ASSOCIATION June 2022

A Hybrid of Copula Prediction and Time Series Computation to Estimate Stream Discharge Based on Precipitation Data

Ying Ouyang

Research Impact Statement: Developed a novel approach to predict stream discharge based on precipitation using Copula method and time series computational algorithm.

ABSTRACT: Stream discharge is a key hydrological factor for water supply planning, wetland loss investigation, ecological service assessment, and climate change impact estimation. Conceptually, stream discharge is expected to be highly and positively related to precipitation. In reality, however, such a relationship may be weaker because precipitation characteristics are affected by local climate of watersheds. For many watersheds around the world, a vast amount of precipitation data are readily available but the stream discharge data are very limited or unavailable. It would be time-saving and cost-effective to predict stream discharge based on precipitation data. Unfortunately, this task is very difficult to achieve using the traditional methods. Although the copula method is able to establish a good relationship (or a good dependence structure) between discharges and precipitations, this relationship does not include the time series process, and thus is impractical for applications. Therefore, a hybrid of copula prediciton and time series computation was developed (with detailed procedures) here to estimate stream discharge based on precipitation data. The method was validated using the measured daily discharges with the good statistical measures, that is, the Kandell's τ (0.42–0.44), normalized root mean square error (2.19–2.28 m^3/s), and R^2 (0.66–0.84). This study suggests that the hybrid method is a useful tool to predict stream discharges based on precipitation data.

(KEYWORDS: algorithm; copula; discharge; method; precipitation; watershed.)

INTRODUCTION

Water resource sustainability due to increasing water demand is a critical concern worldwide (Scanlon et al. 2012; Famiglietti 2014). Many parts of the world, including North Africa, Middle East, South and Central Asia, north China, North America, and Australia, are now experiencing water resource depletion and/or shortage (Garduno and Foster 2010; Doll et al. 2014; Dalin et al. 2017; Ouyang et al. 2019, 2020). Stream discharge is a crucial hydrological factor for water resource management. More specifically, stream discharge is used to determine the minimum streamflow and level, assess wetland loss, reconstruct ecological service, and estimate climate change impact (Ouyang et al. 2013, 2019). Conceptually, stream discharges are expected to be highly and positively related to the amounts of precipitations. In reality, however, such a relationship may not be strong enough because the precipitation characteristics, including the direction, amount, and frequency, are also governed by the biophysical features of local watersheds (Clement and Djebou 2017). In other words, although precipitation is a major source of water for stream discharge, the correlation between discharge and precipitation depends on a wide range of biophysical features and social activities that have

Paper No. JAWR-20-0170-P of the *Journal of the American Water Resources Association* (JAWR). Received November 23, 2020; accepted February 23, 2022. Published 2022. This article is a U.S. Government work and is in the public domain in the USA. **Discussions are open until six months from issue publication**.

Center for Bottomland Hardwoods Research, Southern Research Station, USDA Forest Service, Mississippi State, Mississippi, USA (Correspondence to Ouyang: ying.ouyang@usda.gov).

Citation: Ouyang, Y. 2022. "A Hybrid of Copula Prediction and Time Series Computation to Estimate Stream Discharge Based on Precipitation Data." *Journal of the American Water Resources Association* 58 (3): 471–484. https://doi.org/10.1111/1752-1688.13003.

temporal and spatial interactions (Tidwell et al. 2004). Because of this complexity, methods to estimate stream discharges using precipitation data are still not well developed. Dawdy and Bergmann (1969) studied the effect of rainfall variability on stream discharge in a 25.12 km^2 basin in southern California. They found that the use of a single rain gage can predict peak discharge with a standard error of 20%. Nandagiri and Shetty (2003) analyzed the relationship between the daily stream discharge and the daily rainfall for the Yennehole catchment (327 km^2) in India using the linear and nonlinear regression equations and artificial neural network (ANN) models. These authors compared the prediction accuracies of regression and ANN approaches and concluded that the ANN approach has a better prediction of accuracy. Moon et al. (2004) estimated the stream discharge using spatially distributed rainfall in the Trinity River Basin, Texas using the Soil and Water Assessment Tool (SWAT) model and NEXRAD data. They reported that the SWAT-NEXRAD simulations, in general, overpredict the high flow events and underpredict the low flow events. Supriya et al. (2015) estimated stream discharges for flood forecasting in the Vellar River Basin, southern India using the annual maximum daily rainfall and multiple regression analysis and found that the lower Vellar River Basin is the most vulnerable catchment needed for the flood control. Although the above studies have provided some useful insights into estimating stream discharges from precipitations, a thorough literature search reveals that no reliable method has been developed for such a purpose due to the complex relationship between discharges and precipitations associated with the effects of a wide range of biophysical and social systems. However, there are vast amounts of precipitation data readily available for most watersheds around the world. It would be time-savings, cost-effective, and a great breakthrough if one could estimate the steam discharges in watersheds using the precipitation data. To this end, a new method is developed here to resolve this issue using the copula-based approach associated with a novel time series computational algorithm.

Copula method is a multivariate probability approach used to identify the relationships among many random variables, which is otherwise very difficult (if not impossible) by using traditional methods (Sklar 1959; Schweizer and Wolff 1981; Dall'Aglio, et al. 1991; Nelson 2006). For example, we normally quantify the dependence of two or more variables by the linear or nonlinear multiple regressions. When we obtain poor statistical measurements such as lower coefficient of determination (R^2) and larger probability value (p-value), we may naturally conclude that very poor correlations exist among the variables. However, this conclusion may not be always true if we apply the copula method for the analysis. The Latin word "copula" means "a link, tie or bond" and was first employed in a mathematical analysis by Sklar (1959). The copula theory and its applications have been around for a while, especially in actuarial science and finance applications (Frees and Valdez 1998; Sun 2013) and its uses to hydrological processes and water quality have just begun since last decade (Salvadori and Michele 2004; Wang et al. 2012; Alizadeh et al. 2018). Salvadori and Michele (2004) applied the copula method to study the return periods of hydrological events such as flooding peak and storm intensity and concluded that the calculations on the return periods of hydrological events are greatly simplified using the copula method. Shiau et al. (2007) estimated the relationship between drought duration and drought severity in Yellow River, China using the copula approach. Their results showed that the return period of the drought in the late 1920s to early 1930s is 105 years, whereas the return period of the drought that occurred from 1997 to 1998 is only 4.4 years. Madadgar et al. (2013) performed a drought analysis under climate change at the upper Klamath River Basin in Oregon, USA with copula method. They argued that the duration severity has the strongest correlation with drought, whereas the duration intensity shows the least correlation with the drought. Although more intense extreme events are projected to occur in most parts of the world in the future, their results showed that the upper Klamath River Basin will experience fewer intense droughts as affected by climate change. Chen et al. (2013) applied the copula method to analyze the drought characteristics such as drought duration, severity, and time interval in Han River, China. These authors demonstrated that the normal copula fits every state of the drought periods well and is selected for computing probability and return period analysis of the drought. Alizadeh et al. (2018) developed a copula-based hydro-economic optimization model for optimal design of reservoir-irrigation district systems under multiple interdependent sources of uncertainty. Their results showed that the smaller sizes of reservoir, irrigation district, and stress-avoidance irrigation policies are better than the deficit-irrigation policies. All the above studies have provided useful information on the applications of copula method to determine the correlations among the hydrological processes. However, no effort has been devoted to establishing the relationship between discharges and precipitations using the copula approach.

Several studies have applied the copula regressions to predict one variable using the other variable (Parsa and Klugman 2008; Masarotto and Varin

2017; Cote et al. 2019). The major advantage of the copula regression is that no restriction on probability distribution is required as compared to the ordinary least square and generalized linear regression methods (Parsa and Klugman 2008). Masarotto and Varin (2017) developed a Gaussian copula regression model in the R platform to fit the bivariate data using the maximum likelihood inference function. The model is applied to the malaria data with a very good linear correlation. Cote et al. (2019) applied the ranked-based tools using the copula regression for property and casualty insurance analysis. All of these copula-based regression studies have provided good insights into the copula regression analysis. However, for some natural processes, the linear and/or nonlinear correlations between two or more variables such as discharges and precipitations may not exist even after the copula transformation. These make the application of the copula-based regressions difficult. Additionally, the copula bivariate distribution function is normally used to randomly generate the paired values of two variables simultaneously, which do not include the time series process. In real-world practices, however, we are sometimes required to predict one variable using the known values of the other variable at a given time. For example, we need to predict the future stream discharges when the future precipitations at the given dates are known. Therefore, the time series computational algorithm is required to associate with the copula method.

The goal of this study was to develop a copula-based method in conjunction with a novel time series computational algorithm to predict stream discharges using precipitation data. Specific objectives were to: (1) select a better copula model based on Kendall's statistics for randomly generating the discharge and precipitation data; (2) validate the selected copula model using the long-term measured discharge and precipitation data; (3) develop a novel computational algorithm connected to the copula-based approach to predict the time series stream discharges using the precipitation data; (4) verify the computational algorithm; and (5) apply the method to predict the future stream discharges using the future precipitation data that were obtained from a climate change scenario.

MATERIALS AND METHODS

Copula Model

A copula is a function that couples a multivariate distribution function to its marginal distribution function (Sklar 1959) and is used to define the nonparametric measures of dependence among random variables. An elaboration of the copula theory can be found elsewhere (Genest and Mackay 1986; Frees and Valdez 1998; Dupuis 2007; Zhang and Singh 2007). There exist several methods to derive copula functions and the widely used ones are the inversion, generation, and extreme value methods (Nelsen 2006). The copulas derived from the inversion method are defined as elliptical copulas; the copulas built by generator functions are termed as Archimedean copulas; and the copulas represented the dependence structure between extreme values or exceptional events are named as extreme value copulas (Nelson 2006). Copulas derived by these methods are determined by a small number of parameters that are normally not flexible in developing dependent structure, especially with multi-variables. This weakness can be circumvented by the Vine copulas (Bedford and Cooke 2002).

Vine copulas are highly flexible in applications that are based on a decomposition of the joint copula density into bivariate building blocks (Bedford and Cooke 2002; Aas et al. 2009). A vast variety of copula families are available in the Vine copulas for selecting the dependence structure, which makes the choice of appropriate copulas somewhat difficult. In this study, the commonly used copulas such as Clayton, Frank, Gumbel, and Normal copulas, which are available in the Vine copula package of the R-Statistics, were employed to determine the correlations between discharge and precipitation. The resulted correlations were then validated with the long-term field-observed data. The copula that has the best goodness-of-fit was selected for applications in this study.

Although an elaborate discussion of the copula models is beyond the scope of this study, a brief description of each copula model used in this study is given below for readers' convenience.

Clayton copula, first introduced by Clayton (1978), is an asymmetric Archimedean copula and is defined as:

$$C_{\text{clayton}}(u_1, u_2;\theta) = \left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-1/\theta}, \qquad (1)$$

where C is the copula, u is the distribution function and θ is the copula parameter at the interval $(0, \infty)$. When $\theta = 0$, the marginal distributions become independent. This suggests that the Clayton copula cannot be used to approximate the negative dependence. It should be noted that copula is a tool to assess the dependence structure (or correlation) of random variables. In Clayton copula, the Kendall's tau (τ) is normally used to measure the dependence of random variables (or the correlation of random variables) and is given as:

$$\tau = \frac{\theta}{\theta + 2}. \tag{2}$$

Frank copula is defined as (Frank 1979):

$$C_{\text{Frank}}(u_1, u_2; \theta) = -\theta^{-1} \log\left\{1 + \frac{\left(e^{-\theta u_1} - 1\right)\left(e^{-\theta u_2} - 1\right)}{e^{-\theta} - 1}\right\}, \tag{3}$$

with θ at the interval (-∞, +∞). Frank copula allows the approximation of positive and negative dependence in the data. When θ approaches -∞, the Fréchet–Hoeffding lower bound is attained; while θ approaches +∞, the Fréchet–Hoeffding upper bound is reached. When θ is equal to 0, the independence case is approached. Frank copula is suitable for modeling data characterized by weak tail dependence. In Frank copula, the Kendall's tau (τ) is calculated as:

$$\tau = 1 + 4[D(\theta) - 1]/\theta. \tag{4}$$

where D is the Debye function.

Gumbel copula is used to model asymmetric dependence in the data and is capable to capture strong upper tail dependence and weak lower tail dependence. For those variables with strong correlations at high values but weak correlations at low values, Gumbel copula is a good choice. The bivariate Gumbel copula can be defined as (Gumbel 1960):

$$C_{\text{gumbel}}(u_1, u_2, \theta) = \exp\left(-\left[(-\log u_1)^{\theta} + (-\log u_2)^{\theta}\right]^{1/\theta}\right), \tag{5}$$

with θ at the interval of [1, ∞). If θ approaches 1, the marginal distributions are independent, and if θ goes to infinity, the Gumbel copula approaches the Fréchet–Hoeffding upper bound. Analogous to the Clayton copula, the Gumbel copula cannot be used to approximate the negative dependence.

The correlation between Kendall's τ and Gumbel copula parameter θ for measuring the dependence of random variables is given by the following equation:

$$\tau = 1 - \theta^{-1}. \tag{6}$$

The Normal (or Gaussian) copula belongs to the elliptical copulas family and is derived from the multivariate Gaussian or Normal distribution (Renard and Lang 2007) as:

$$C_{\text{Gaussian}}(u_1, \ldots, u_n) = \Phi_n\left[\phi^{-1}(u_1), \ldots, \phi^{-1}(u_n)\right], \tag{7}$$

where Φ is the cumulative distribution function (cdf) of a multivariate normal distribution with zero mean and covariance matrix, and ϕ is the cdf of the standard normal distribution at θ (0, 1). In practice, Normal copula is popular because it allows modeling dependence in arbitrarily high dimensions with only one parameter, governing the strength of dependence. The correlation between Kendall's τ and Normal copula for measuring the dependence of random variables is given by:

$$\tau = \frac{2}{\pi} \arcsin\theta. \tag{8}$$

Study Site and Data Acquisition

The long-term measured discharge and precipitation data from 1900 to 2018 in forest watersheds of the lower Mississippi River Alluvial Valley (LMRAV) were used in this study (Figure 1). The LMRAV is situated in the floodplain of the Mississippi River (MR) beginning from the north at Illinois, continuing through Missouri, Kentucky, Arkansas, Tennessee, and Mississippi, and ending at the Gulf of Mexico (GOM) in South Louisiana. Clearcuttings of bottomland hardwood forests, conversions from forests to agricultural lands

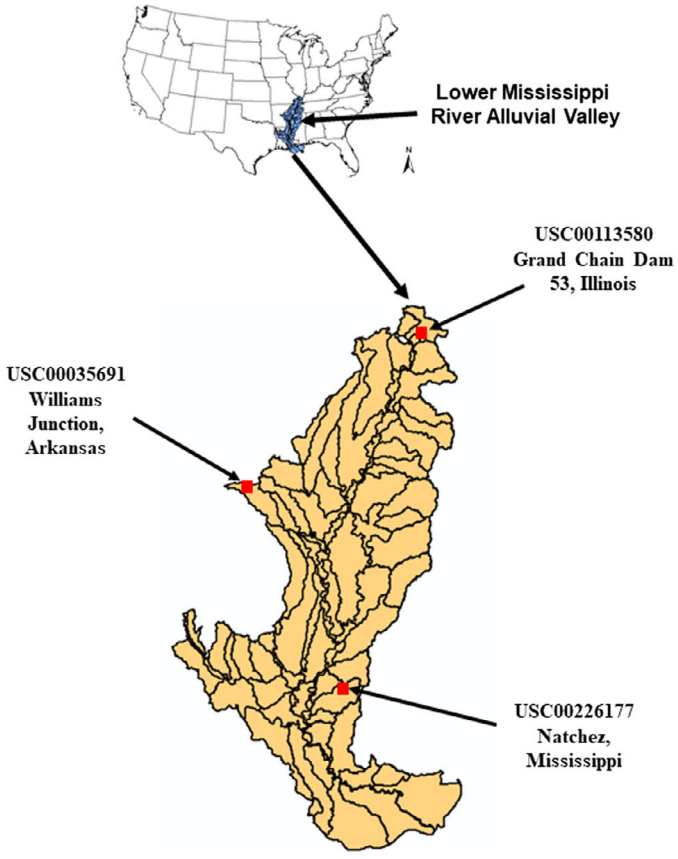


FIGURE 1. Location map for the three study sites used in this study.

with intensive crop production, and intensified and extreme precipitations are the major factors affect river flooding, wetland loss, and water quality degradation in the MR and the adjacent GOM (Munoz and Dee 2017; Ouyang et al. 2013, 2018, 2020).

Three NOAA (National Oceanic and Atmospheric Administration) weather stations, namely the USC00226177 in Natchez, Mississippi; USC00113580 in Grand Chain Dam 53, Illinois; and USC035691 in Williams Junction, Arkansas, were selected to download the daily precipitation data (Figure 1). These NOAA weather stations (https://www.ncdc.noaa.gov/cdo-web/datasets#GHCND) are located near the headwater areas of the forest lands in the LMRAV and have more than 100 years of data records. Meanwhile, three USGS (United States Geological Survey) gauge stations, namely the #07291000 in Homochitto; MS; #03621000 in Forman, Illinois; and #07362100 in Smackover, Arkansas, were selected to download the daily discharge data. These USGS gauge stations are located at the same or nearby their corresponding NOAA weather stations in each state and have daily stream discharges data for the periods of records ranged from 60 to 90 years. These NOAA weather and USGS gauge stations were selected partially because they have long-term data records and partially because the forest watersheds in the area have less land-use disturbances, which provide a better condition for analyzing how the future climate change affects the stream discharges.

The future daily precipitation data for the three study sites were downloaded from the Hydrologic and Water Quality System (HAWQS) model (https://hawqs.tamu.edu/#/). More specifically, the daily future precipitation data for the USGS #07291000, USGS #03621000, and USGS #07362100 were downloaded from HAWQS, respectively, using the HUC 8 (hydrologic unit code 8) numbers of 08060205, 05140206, and 08040201 from the CCSM4 (Community Climate System Model 4) with the RCP45 (Representative Concentration Pathway 45) scenario. HAWQS is a national watershed and water quality assessment tool distributed by the United States Environmental Protection Agency, which is a customized version of the SWAT model.

Method Development

The detailed steps used to develop the method are given below for readers' convenience. The copula analysis (Figure 2) was performed with an R script that was modified from https://www.r-bloggers.com/how-to-fit-a-copula-model-in-r-heavily-revised-part-2-fitting-the-copula/. Steps 1–5 (below) are similar to those reported at https://www.r-bloggers.com/how-to-fit-a-copula-model-in-r-heavily-revised-part-2-fitting-the-copula/, whereas Steps 6 and 7 were developed in this study.

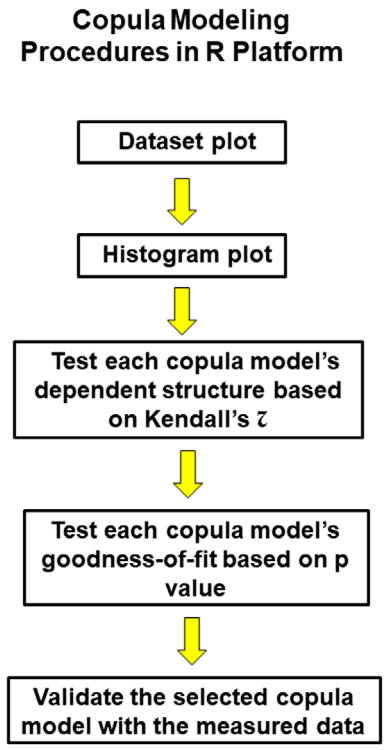


FIGURE 2. Steps used in copula analysis.

1. Dataset plot. The first step is to visually inspect the correlation between the measured discharges and the precipitations (Figure 3). It is apparent from the figure that very poor or no correlations exist between the two variables for all the three study sites used in this study. Therefore, the copula method was employed to determine their dependences or correlations.
2. Histogram plot. A histogram plot can provide a good estimate on the marginal distributions of the two variables, which can help select the suitable copula marginal function for discharges and precipitations. Comparisons of the histograms between the field measured and Gamma simulated precipitations and discharges for the three study sites are given in Figure 4. These histograms showed a Gamma type of distribution. Therefore, the Gamma distribution was used as the marginal distribution function when building the bivariate distribution from the selected copula model. In this study, the bivariate distribution tells the probability that a certain event will occur for each possible choice of the two variables, that is, the discharge and precipitation.
3. Dependence of discharge on precipitation. As shown in Figure 3, no dependence of the

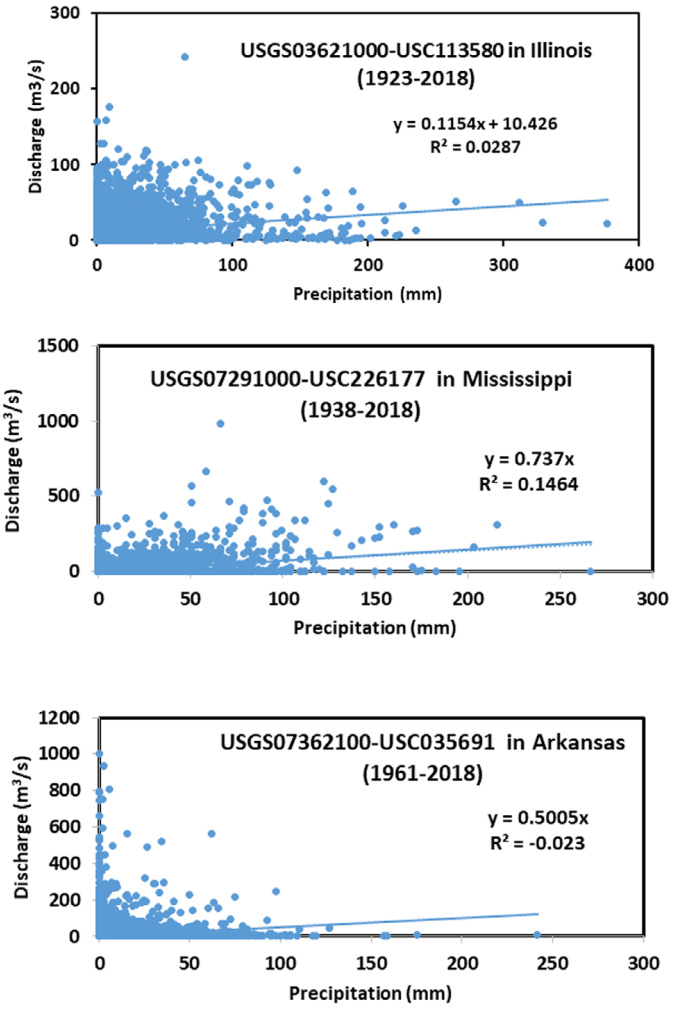


FIGURE 3. Correlations between the measured precipitations and discharges for the three study sites.

measured discharges on the measured precipitations existed for the three study sites since the R^2 values were too low. In the copula analysis, the correlation (or dependence) of two or more variables is normally measured (or estimated) by the Kendall's τ and Spearman's Rho methods. In this study, the Kendall's τ values were used to measure the correlations (or dependences) between the discharges and the precipitations (Table 1). It should also be noted that the measured discharges were not used for copula analysis if the measured precipitations were zero or near zero. If either the precipitation or discharge data were missing on certain dates, the data for those were not used for copula analysis either.

4. Goodness-of-fit. After the copula models have been fitted, the "gofCopula ( )" function from the Vine Copula package in the R platform was applied to test the goodness-of-fit for each copula model, which was measured by the p value (Table 1). The best copula model was then selected for further study.
5. Comparison of copula generations (or predictions) and field measurements. After the copula model was selected, the copula bivariate distribution function was used to randomly generate the paired discharges and precipitations. These generated data were then compared with the measured data using the Kendall's τ values (Figure 5). It should be caution that the copula-generated discharges and precipitations are not the time series data, that is, they do not tell when (e.g., the date) an event occurs; whereas the measured discharges and precipitations are the time series data.
6. Development of the time series computational algorithm. As stated in Step 5, the copula bivariate distribution function can randomly generate the paired discharge and precipitation data. This data, however, cannot be directly used because it does not have the time concept on when the discharges and precipitations occur. In this study, we need to predict the future time series stream discharges when the future precipitations at given dates are known. Therefore, a time series computational algorithm was developed to circumvent the obstacle. It is assumed that the copula-generated discharge and precipitation dataset is proportional to the predicted discharge and precipitation dataset, which can be characterized as:

$$\frac{R_c}{D_c} = \frac{R_m}{D_p}, \tag{9}$$

or

$$D_p = \frac{D_c R_m}{R_c}. \tag{10}$$

where D_p is the predicted discharge (m^3/s), D_c is the copula-generated discharge (m^3/s), R_m (>0) is the measured or known precipitation (mm), and R_c (> 0) is the copula-generated precipitation (mm). Equation (10) is used to predict the stream discharges based on both the measured (or known) precipitations as well as the copula-generated discharges and precipitations. Figure 6 shows the following procedures on how to implement Equation (10) in Microsoft Excel for including time sereis process: (a) sort the copula-generated discharge and precipitation data (Figure 6a) from smallest to largest based on the precipitation data; (b) sort the future date and precipitation data (Figure 6b) from smallest to largest also based on the precipitation data; (c) add the two sorted datasets together side-by-side (Figure 6c); and (d) sort the data in Figure 6c based on

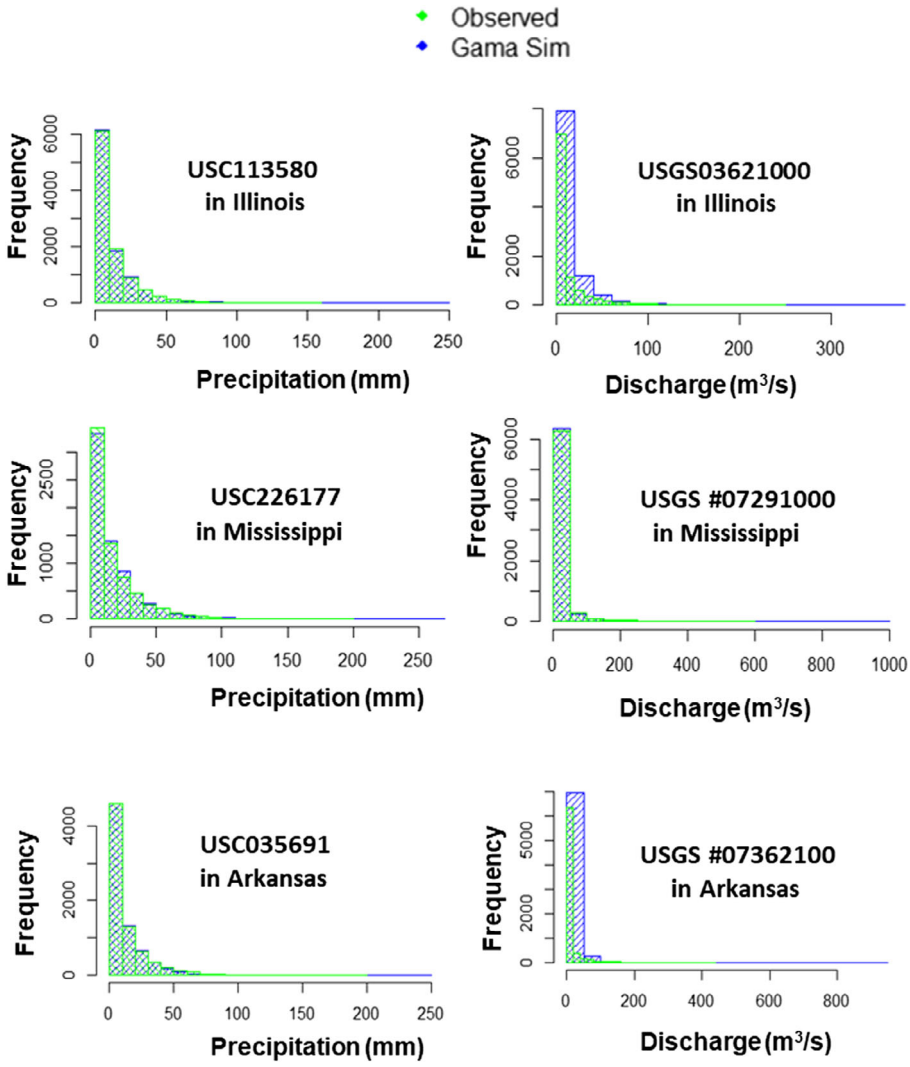


FIGURE 4. Histogram plots of the precipitations and discharges for the three study sites.

TABLE 1. Statistical measures of the copula functions for identifying the dependence structures between discharges and precipitations at the three study sites used in this study.

Copula function	Copula fit *t*	Goodness of fit *p*-value
USGS03621000-USC113580 in Illinois		
Clayton copula	0.44	0.0098
Frank copula	0.16	0.0098
Gumbel copula	0.17	0.0098
Normal copula	0.33	0.0098
USGS07291000-USC226177 in Mississippi		
Clayton copula	0.43	0.0098
Frank copula	0.16	0.0098
Gumbel copula	0.17	0.0098
Normal copula	0.33	0.0098
USGS07362100-USC035691 in Arkansas		
Clayton copula	0.42	0.0290
Frank copula	0.17	0.3235
Gumbel copula	0.16	0.0098
Normal copula	0.33	0.0098

date in chronological order and then predict the future discharge using Equation (10) (Figure 6d). The reason for sorting the copula-generated and future precipitation datasets from smallest to largest was to make them into the same rank order when adding them together. The reason for sorting the final dataset based on dates was for chronological display of the predicted future discharge (Figure 6d). The predicted discharges from Equation (10) were verified against the copula-generated discharges to see if these two datasets are proportional to each other (i.e, a linear correlation) using the statistical measures such as coefficient of determination (R^2) and normalized root mean square error (nRMSE) (Figure 7).

7. Application of the method. The method was then applied to predict the future stream discharge for given watersheds based on the future precipitation

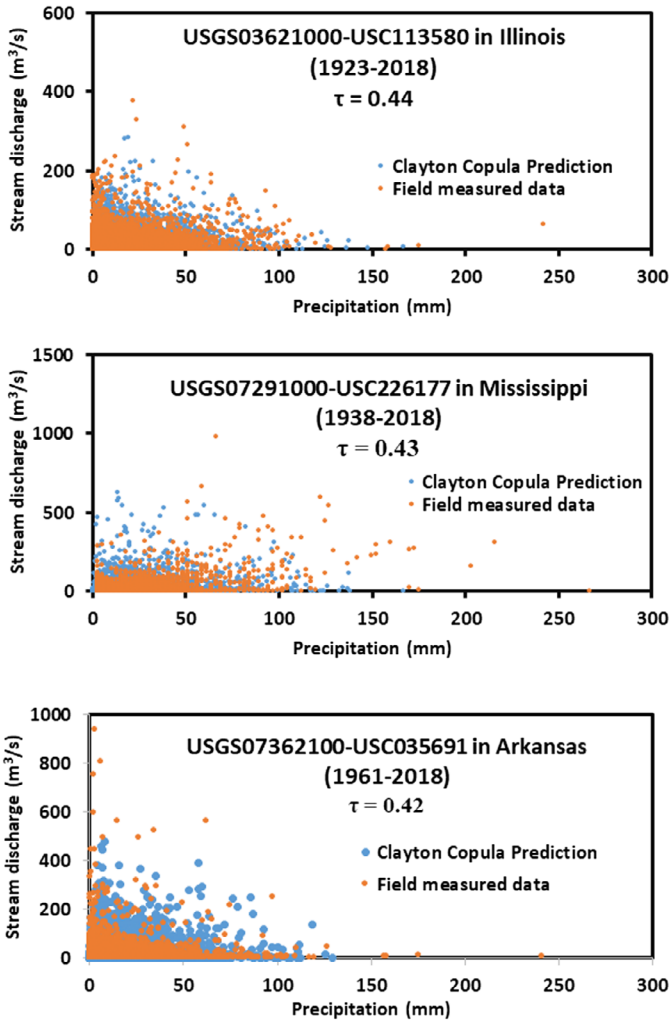


FIGURE 5. Comparison of the measured and copula-generated precipitations and discharges for the three study sites.

data with an assumption that the future geophysical settings such as river channels and land uses for the watersheds of interest remain unchanged.

RESULTS AND DISCUSSIONS

Copula Model Selection

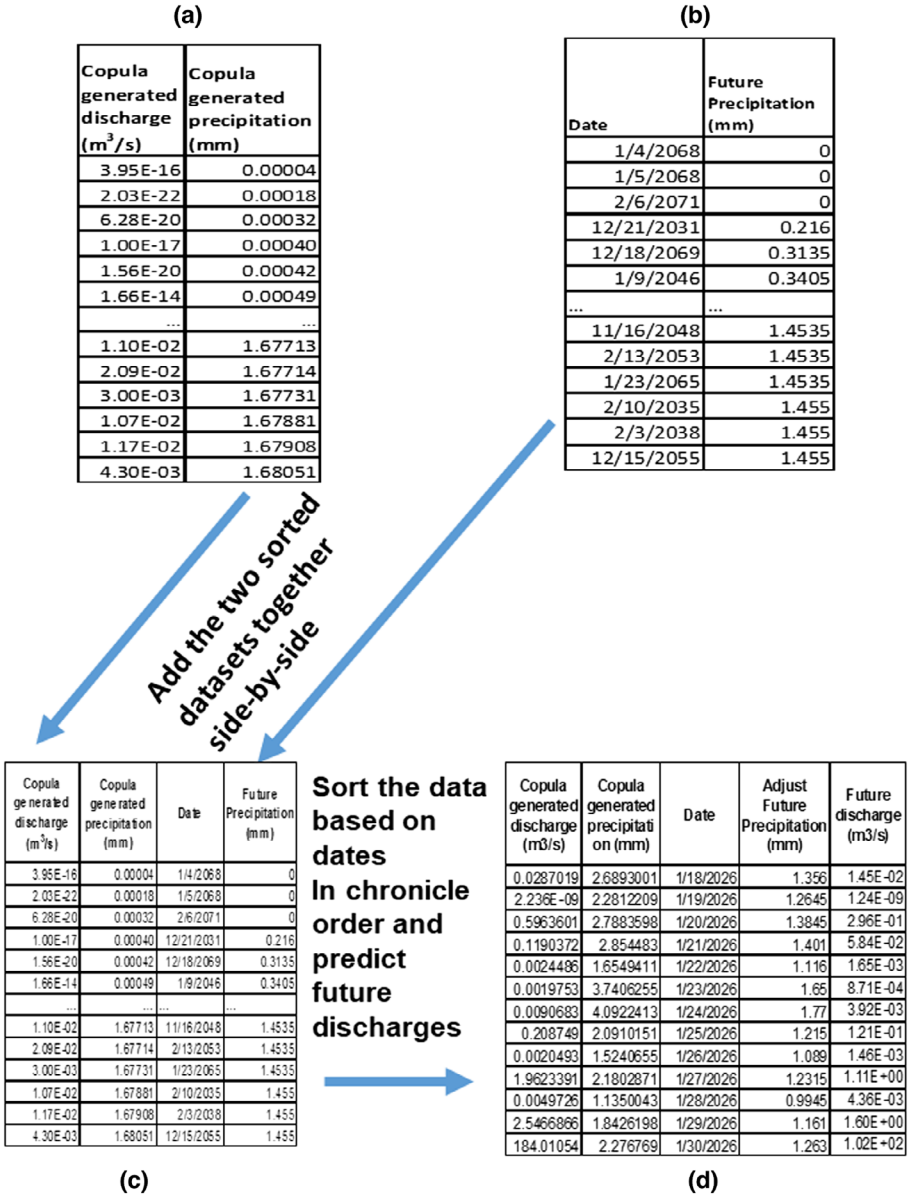
The dependence measures (τ values) of the discharges on the precipitations from the Clayton, Frank, Gumbel, and Normal copulas for the three study sites are given in Table 1. Compared with the Frank, Gumbel and Normal copulas, the Clayton copula was selected for the three study sites because of its highest τ values (Table 1). In Kendall statistic, the value τ ranges from -1 to 1 and is a measure of relationships between variables, where 0 is no relationship and 1 (or -1) is a perfect relationship (with positive τ for increasing trend and negative τ for decreasing trend) (Mangiafico 2016). In other words, the Clayton copula was the best in generating a good relationship between the discharges and the precipitations for the three study sites.

The goodness-of-fit of the Clayton copula was further investigated by comparing the copula-generated and the field-measured discharges and precipitations. In the goodness-of-fit analysis, the p value was used to measure a trend. If the p value was ≤ 0.05, there was a monotonic trend (Mangiafico 2016). With the low p values (Table 1), we confirmed that the Clayton copula model is suitable for the purpose of this study. It should be noted that the p value of the Clayton (0.029) copula was larger than those of the Gumbel (<0.01) and Normal (<0.01) copulas for the study site in Arkansas, but this p value was acceptable ($0.029 < 0.05$). Since the τ value of the Clayton (0.42) copula was larger than those of the Gumbel (0.16) and Normal (0.33) copulas for the study site in Arkansas (Table 1), the Clayton copula was selected.

Plots of the discharges against the precipitations between the copula predictions (or generations) and the field measurements for the three study sites are shown in Figure 5. The copula predictions were generated from the Clayton copula bivariate distribution function. Overall, the copula predictions matched the field measurements reasonably well for the three study sites because the τ values were >0.3 (Mangiafico 2016). Mangiafico (2016) stated that a τ value between 0.1 and 0.3 indicates a small impact and a τ value >0.3 denotes a significant trend. In particular, the copula predictions for the study site in Illinois were slightly better than that in Arkansas, and the study site in Mississippi was in between based on the τ values (Figure 5).

Computational Algorithm Verification

As stated in Step 6 of Method Development section, the Clayton copula-generated discharges and precipitations did not have the chronological times (e.g., dates) on when the discharges and precipitations occurred. In other words, the Clayton copula bivariate function can only provide a possible stream discharge for each possible precipitation event regardless of the time course. In practices, water resource managers and researchers normally want to predict the stream discharges based on the given precipitations at given times (e.g., dates and years). To circumvent the obstacle, a computational algorithm was developed (Equation 10 and Step 6) and verified in this study.



(a)

Copula generated discharge (m^3/s)	Copula generated precipitation (mm)
3.95E-16	0.00004
2.03E-22	0.00018
6.28E-20	0.00032
1.00E-17	0.00040
1.56E-20	0.00042
1.66E-14	0.00049
...	...
1.10E-02	1.67713
2.09E-02	1.67714
3.00E-03	1.67731
1.07E-02	1.67881
1.17E-02	1.67908
4.30E-03	1.68051

(b)

Date	Future Precipitation (mm)
1/4/2068	0
1/5/2068	0
2/6/2071	0
12/21/2031	0.216
12/18/2069	0.3135
1/9/2046	0.3405
...	...
11/16/2048	1.4535
2/13/2053	1.4535
1/23/2065	1.4535
2/10/2035	1.455
2/3/2038	1.455
12/15/2055	1.455

(c)

Copula generated discharge (m^3/s)	Copula generated precipitation (mm)	Date	Future Precipitation (mm)
3.95E-16	0.00004	1/4/2068	0
2.03E-22	0.00018	1/5/2068	0
6.28E-20	0.00032	2/6/2071	0
1.00E-17	0.00040	12/21/2031	0.216
1.56E-20	0.00042	12/18/2069	0.3135
1.66E-14	0.00049	1/9/2046	0.3405
	...	...	...
1.10E-02	1.67713	11/16/2048	1.4535
2.09E-02	1.67714	2/13/2053	1.4535
3.00E-03	1.67731	1/23/2065	1.4535
1.07E-02	1.67881	2/10/2035	1.455
1.17E-02	1.67908	2/3/2038	1.455
4.30E-03	1.68051	12/15/2055	1.455

(d)

Copula generated discharge (m3/s)	Copula generated precipitation (mm)	Date	Adjust Future Precipitation (mm)	Future discharge (m3/s)
0.0287019	2.6893001	1/18/2026	1.356	1.45E-02
2.236E-09	2.2812209	1/19/2026	1.2645	1.24E-09
0.5963601	2.7883598	1/20/2026	1.3845	2.96E-01
0.1190372	2.854483	1/21/2026	1.401	5.84E-02
0.0024486	1.6549411	1/22/2026	1.116	1.65E-03
0.0019753	3.7406255	1/23/2026	1.65	8.71E-04
0.0090683	4.0922413	1/24/2026	1.77	3.92E-03
0.208749	2.0910151	1/25/2026	1.215	1.21E-01
0.0020493	1.5240655	1/26/2026	1.089	1.46E-03
1.9623391	2.1802871	1/27/2026	1.2315	1.11E+00
0.0049726	1.1350043	1/28/2026	0.9945	4.36E-03
2.5466866	1.8426198	1/29/2026	1.161	1.60E+00
184.01054	2.276769	1/30/2026	1.263	1.02E+02

FIGURE 6. Procedures in developing the time series predictive algorithm. (a) sort the copula-generated discharge and precipitation data from smallest to largest based on the precipitation data, (b) sort the future date and precipitation data from smallest to largest also based on the precipitation data, (c) add the two sorted datasets together side-by-side, and (d) sort the data in (c) based on date in chronological order and then predict the future discharge using Equation (10).

Comparisons of the Clayton copula-generated discharges with the Equation (10) calculated discharges for the three study sites are shown in Figure 7. The values of R^2 and nRMSE were, respectively, 0.838 and 2.22 for the study site in Illinois; 0.662 and 2.194 in Mississippi; and 0.704 and 2.283 in Arkansas. These statistical measures confirmed the assumption that the Clayton copula-generated discharges were proportion to the Equation (10) calculated discharges. Therefore, Equation (10) is feasible to predict the time series discharges from the given (or measured) precipitations along with the copula-generated discharges and precipitations.

Method Application

The predicted future daily discharges in response to future daily precipitations for the three study sites over the 50-year simulation period from 2026 to 2074 are shown in Figure 8. The future daily precipitation data were obtained from the climate change scenario as described in Study Site and Data Acquisition section, whereas the future daily discharges were calculated by Equation (10) with the data generated from the Clayton copula bivariate distribution function. In general, the future daily discharges varied with years

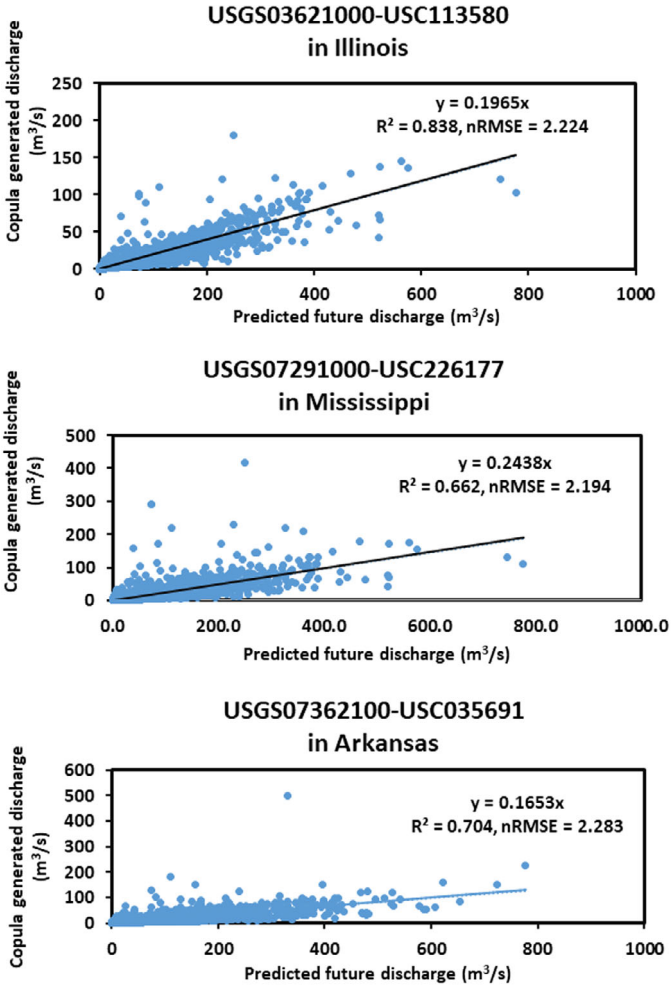


FIGURE 7. Correlations of the predicted and copula-generated discharges.

and locations, and they did not correspond well to the future daily precipitations. For example, the maximum daily discharge was 179.58 m^3/s on January 10, 2055 for the study site in Illinois (Figure 8a) but the amount of the daily precipitation was only 0.22 mm on the same date at the same location (Figure 8b). The similar results were also found for the other two study sites (Figure 8). These occurred because the daily discharges had a very poor relationship with those of the daily precipitations as the daily discharges depended not only on the precipitation rates but also on the watershed hydrogeological conditions as well as the time-lag in stream discharges after precipitations. Such poor relationships were also confirmed by the field-measured data (Figure 3). Overall, the steeper slope, narrower stream channel, larger drainage area, and lesser tree and grass covered land would result in the higher stream discharges (Ouyang et al. 2019). In addition, the antecedent stream flow conditions (i.e., wet or dry) also play an important role.

Comparisons of the past and future daily discharges for the three study sites are shown in Figure 9. The past discharges were obtained from the field measurements, while the future discharges were predicted with the method developed in this study. In general, the future daily discharges were lower than those of the past daily discharges for the three study sites, especially for the study site in Arkansas (Figure 9c). For instance, the past average daily discharge was 11.82 m^3/s but the future average daily discharge was 3.94 m^3/s for the study sites in Arkansas. I attributed this discrepancy to the difference in average daily precipitation between the past and the future. The past average daily precipitation was 11.9 mm but the future average daily precipitation was 2.25 mm at the study site in Arkansas. A 5.3-fold decrease in the future average daily precipitation decreased the future average daily discharge by about three times. It should be caution that the future average daily precipitations obtained from the RCP45 scenario for the three study sites were three to five times lower than those of the past measured daily precipitations. These scenario datasets may underestimate the future precipitation conditions although it is beyond the scope of this study.

Variations in future annual discharges for the three study sites over the 50-year period are shown in Figure 10a–c. Analogous to the case of the future daily discharges, the future annual discharges also varied with years and locations, which occurred because of the different biological, geological, and hydrological watershed conditions. Compared to the future average annual precipitation, the variations in the future average annual discharge with locations were dramatic. For example, the future average annual precipitations for the study sites in Arkansas and Mississippi were, respectively, 822.36 and 847.43 mm, whereas the future average annual discharges for the same locations were, respectively, 354.14 and 1226.77 m^3/s. The difference between the two sites was about 3% for the future average annual precipitation but was about 246% for the future average annual discharge. Results further confirmed that although the precipitation was a major source of water for stream discharges, the rate of stream discharge depended on a wide range of watershed and social conditions.

Finally, it is worth mentioning that the method developed in this study can also be used to quantify the relationships of other variables if their copula dependence structures follow. For many variables, such dependence structures or relationships may

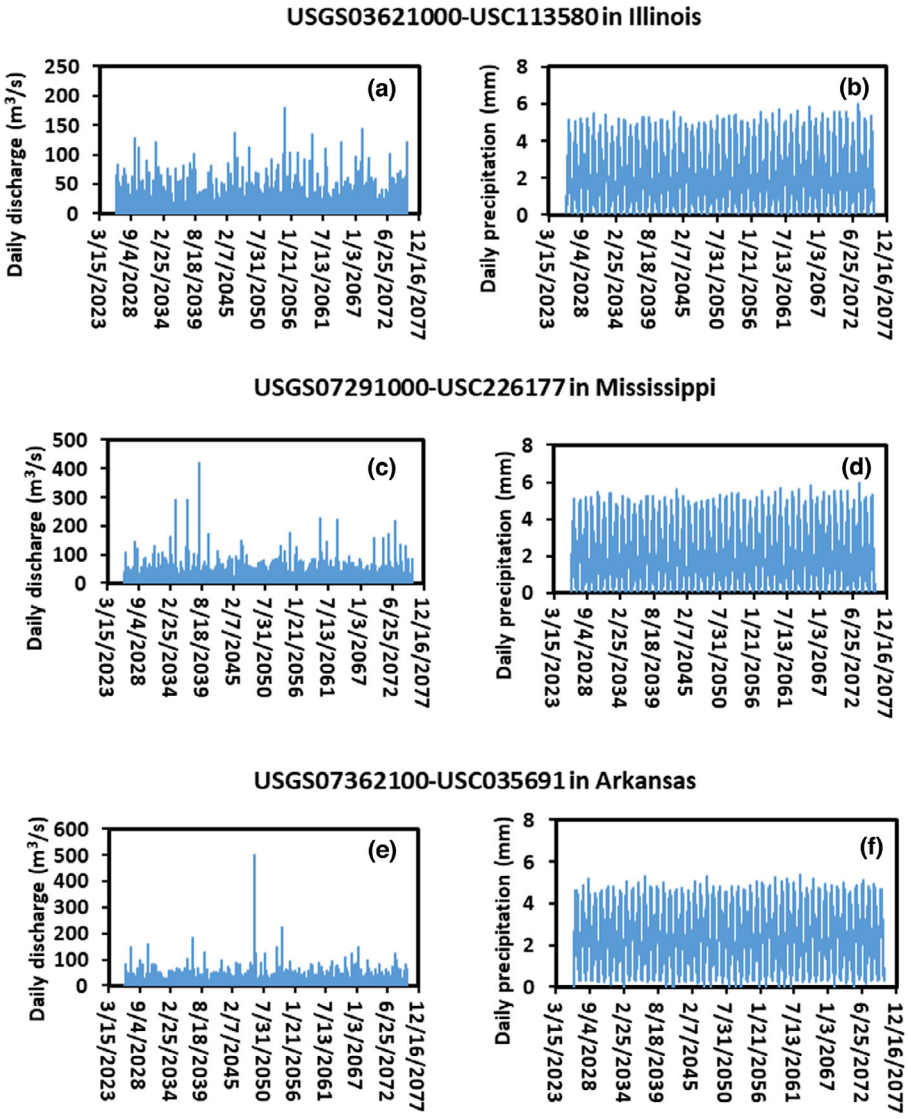


FIGURE 8. Future daily discharges and precipitations for the three study sites.

exist although it may not be possible to identify them using the traditional methods.

CONCLUSIONS

A time-savings, cost-effective, and novel method was developed to predict steam discharges using precipitation data, which is otherwise very difficult (if not impossible) by using the traditional methods.

This copula-based method was validated using the long-term (>60 years) field-measured data from three USGS stream gage stations and three NOAA weather monitoring stations based on the Kendall's τ and p value. A novel computational algorithm (Equation 10) was formulated to calculate the time series discharges in conjunction with copula-generated discharges and precipitations, which was verified using R^2 and nRMSE. Based on the very good statistical measures, the method developed here is capable to predict the stream discharges using the precipitation data.

The method was applied to project the future 50 years stream discharges at the three study sites using the future precipitations obtained from the climate change scenario with very promising predictions. It should also be noted that no future discharge data are available to valid the copula

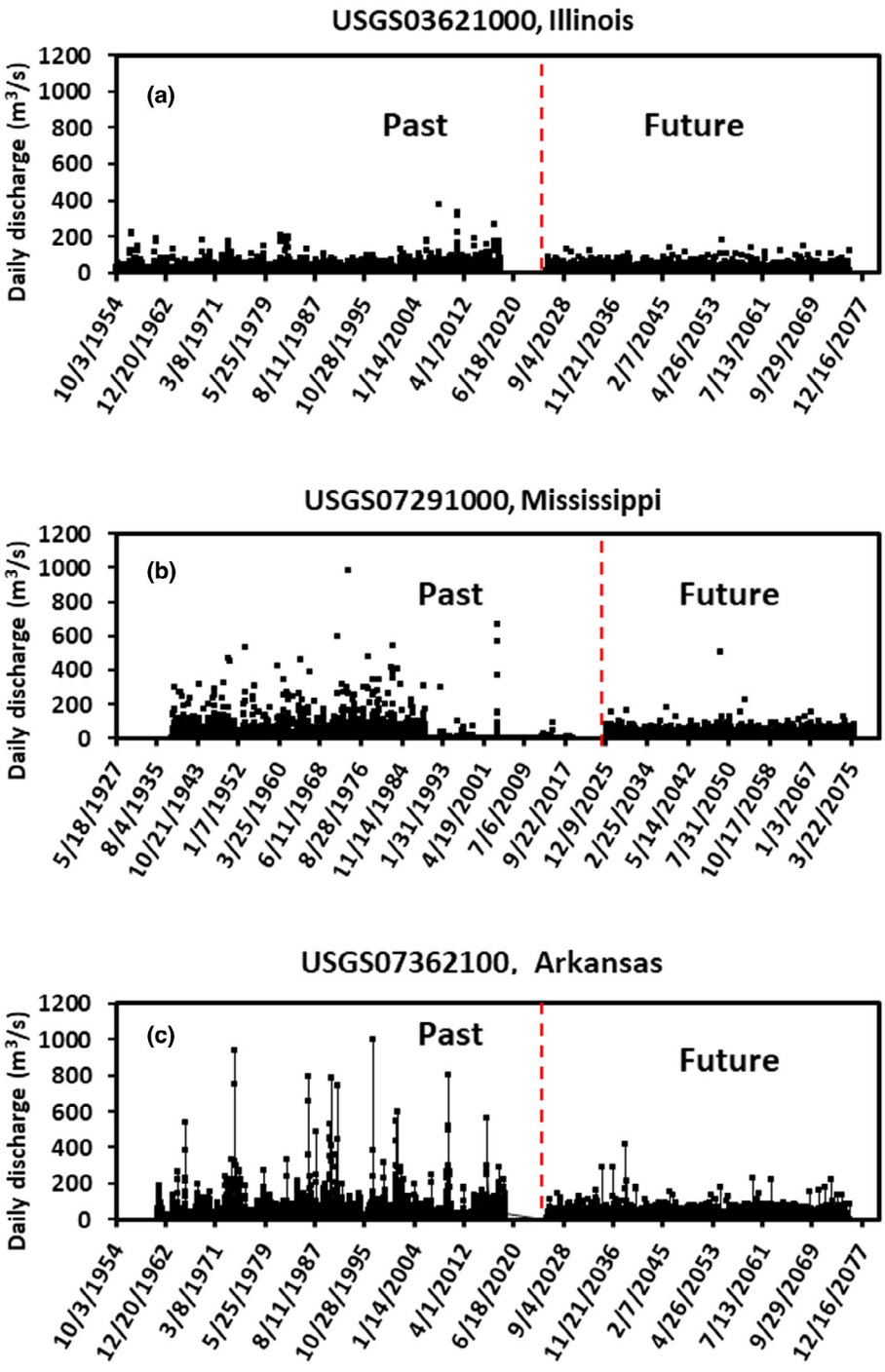


FIGURE 9. Comparison of the past measured and the future predicted daily discharges.

dependence structure of future precipitation and discharge. It was assumed that the watershed properties such as land use and topography will not be changed in the future and the only changes are discharge and precipitation for watersheds of interest in this study. It was further assumed that the future precipitation and discharge have a similar relationship as that of the past precipitation and discharge.

There is a very good potential to employ the method for quantifying the relationships of other variables if their copula dependence structures exist. Further study is therefore warranted to investigate the issue.

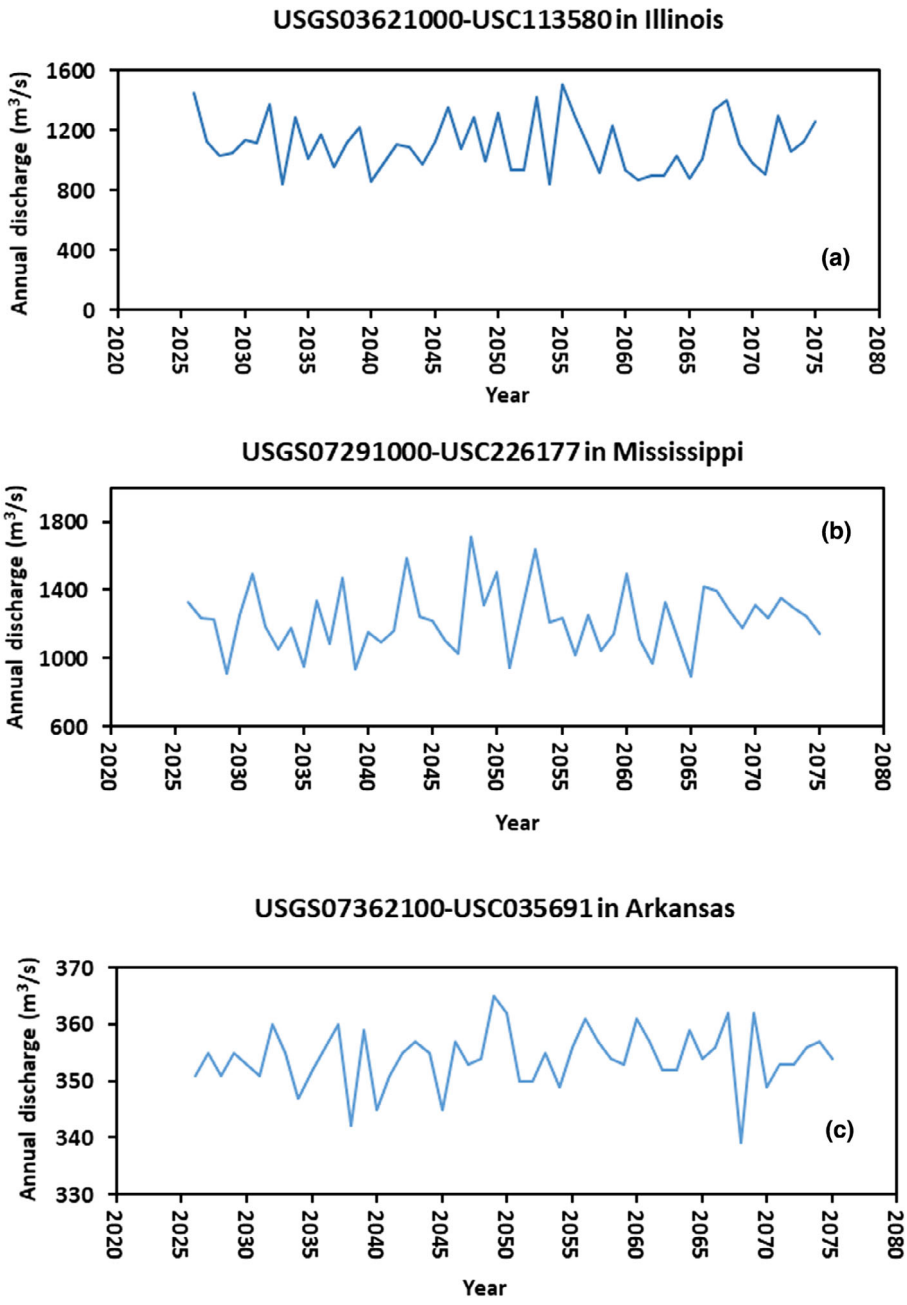


FIGURE 10. Predicted future annual discharges over the 50-year period for the three study sites.

DATA AVAILABILITY STATEMENT

All the data are available to readers upon request.

AUTHOR CONTRIBUTIONS

Ying Ouyang: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Software; Validation; Visualization; Writing – original draft; Writing – review & editing.

LITERATURE CITED

Aas, K., C. Czadob, A. Frigessic, and H. Bakkend. 2009. "Pair-Copula Constructions of Multiple Dependence." *Insurance: Mathematics and Economics* 44: 182–98.

Alizadeh, H.S., J. Mousavi, and K. Ponnambalam. 2018. "Copula-Based Chance-Constrained Hydro-Economic Optimization Model for Optimal Design of Reservoir-Irrigation District Systems under Multiple Interdependent Sources of Uncertainty." *Water Resources Research* 53: 5763–84.

Bedford, T., and R.M. Cooke. 2002. "Vines–A New Graphical Model for Dependent Random Variables." *Annals of Statistics* 4: 1031–68.

Chen, L., V.P. Singh, S. Guo, A.K. Mishra, and J. Guo. 2013. "Drought Analysis Using Copulas." *Journal of Hydrologic Engineering* 18: 797–808.

Clayton, D.G. 1978. "Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence." *Biometrika* 65: 141–51.

Clement, D., and S. Djebou. 2017. "Integrated Approach to Assessing Stream Discharges and Precipitation Alterations under Environmental Change: Application in the Niger River Basin." *Journal of Hydrology: Regional Studies* 4: 571–82.

Cote, M.P., C. Genest, and K. Omelka. 2019. "Rank-Based Inference Tools for Copula Regression, with Property and Casualty Insurance Applications." *Insurance: Mathematics and Economics* 89: 1–15.

Dalin, C., Y. Wada, T. Kastner, and M.J. Puma. 2017. "Groundwater Depletion Embedded in International Food Trade." *Nature* 543: 700–04.

Dall'Aglio, G., S. Kotz, and G. Salinetti, eds. 1991. *Advances in Probability Distribution Functions with Given Marginals: Beyond the Copulas*. Dordrecht, The Netherlands: Kluwer.

Dawdy, D.R., and J.M. Bergmann. 1969. "Effect of Rainfall Variability on Stream Discharges Simulation." *Water Resources Research* 5: 958–66.

Doll, P., S.H. Müller, C. Schuh, F.T. Portmann, and A. Eicker. 2014. "Global-Scale Assessment of Groundwater Depletion and Related Groundwater Abstractions: Combining Hydrological Modeling with Information from Well Observations and GRACE Satellites." *Water Resources Research* 50: 5698–720.

Dupuis, D.J. 2007. "Using Copulas in Hydrology: Benefits, Cautions, and Issues." *Journal of Hydrologic Engineering* 12: 381–93.

Famiglietti, J.S. 2014. "The Global Groundwater Crisis." *Nature Climate Change* 4: 945–48.

Frank, M.J. 1979. "On the Simultaneous Associativity of F(x, y) and x + y – F(x, y)." *Aequationes Mathematicae* 19: 194–226.

Frees, E.W., and E.A. Valdez. 1998. "Understanding Relationships Using Copulas." *North American Actuarial Journal* 2: 1–25.

Garduno, H., and S. Foster. 2010. "Sustainable Groundwater Irrigation Approaches to Reconciling Demand with Resources." GW-MATE Strat. Overv., Ser. 4.

Genest, C., and L. Mackay. 1986. "The Joy of Copulas: Bivariate Distributions with Uniform Marginals." *American Statistician* 40: 280–83.

Gumbel, E.J. 1960. "Bivariate Exponential Distributions." *Journal of American Statistical Association* 55: 698–707.

Madadgar, S., H. Moradkhani, S.M. Madadgar, and H. Moradkhani. 2013. Drought analysis under climate change using copula. *Journal of Hydrologic Engineering* 18: 746–59.

Mangiafico, S.S. 2016. *Summary and Analysis of Extension Program Evaluation in R, Version 1.11.1*. New Brunswick, NJ: Rutgers Cooperative Extension.

Masarotto, G., and C. Varin. 2017. "Gaussian Copula Regression in R." *Journal of Statistical Software* 77: 8. https://doi.org/10.18637/jss.v077.i08.

Moon, J., R. Srinivasan, and J.H. Jacobs. 2004. "Stream Flow Estimation Using Spatially Distributed Rainfall in the Trinity River Basin, Texas." *Transactions of the American Society of Agricultural Engineers* 47 (5): 1445–51.

Munoz, S.E., and S.G. Dee. 2017. "El Niño Increases the Risk of Lower Mississippi River flooding." *Scientific Reports* 7: 1772.

Nandagiri, L., and A. Shetty. 2003. "Stream Flow Estimation from Rainfall-A Comparison of Statistical and Artificial Neural Network Approaches." In *Advance in Hydrology*, edited by V.P. Singh and R.N. Yadava, 253–225. India: Allied Publishers Pvt. Ltd.

Nelson, R.B. 2006. *An Introduction to Copulas* (Second Edition). New York: Springer-Verlag.

Ouyang, Y., W. Jin, J. Grace, S.E. Obalum, W.C. Zipperer, and X. Huang. 2019. "Estimating Impact of Forest Land on Groundwater Recharge in a Humid Subtropical Watershed of the Lower Mississippi River Alluvial Valley." *Journal of Hydrology: Regional Studies* 26: 100631.

Ouyang, Y., T.D. Leininger, and M. Moran. 2013. "Impacts of Reforestation upon Sediment Load and Water Outflow in the Lower Yazoo River Watershed, Mississippi." *Ecological Engineering* 61: 394–406.

Ouyang, Y., P.B. Parajuli, G. Feng, T.D. Leininger, Y. Wan, and P. Dash. 2018. "Application of Climate Assessment Tool (CAT) to Estimate Climate Variability Impacts on Nutrient Loading from Local Watersheds." *Journal of Hydrology* 563: 363–71.

Ouyang, Y., J. Zhang, G. Feng, Y. Wan, and T.D. Leininger. 2020. "A Century of Precipitation Trends in Forest Lands of the Lower Mississippi River Alluvial Valley." *Scientific Reports* 10: 1–16.

Parsa, R.A., and S.A. Klugman. 2008. "Copula Regression." *Variance Advancing the Science of Risk* 5: 45–54.

Renard, B., and M. Lang. 2007. "Use of a Gaussian Copula for Multivariate Extreme Value Analysis: Some Case Studies in Hydrology." *Advances in Water Resources* 30: 897–912.

Salvadori, G., and C. De Michele. 2004. "Frequency Analysis via Copulas: Theoretical Aspects and Applications to Hydrological Events." *Water Resources Research* 40: 1–17.

Scanlon, B.R., C.C. Faunt, L. Longuevergne, R.C. Reedy, W.M. Alley, V.L. McGuire, and P.V. McMahon. 2012. "Groundwater Depletion and Sustainability of Irrigation in the US High Plains and Central Valley." *Proceedings of the National Academy of Sciences of the United States of America* 109: 9320–25.

Schweizer, B., and E. Wolff. 1981. "On Nonparametric Measures of Dependence for Random Variables." *The Annals of Statistics* 9: 879–85.

Shiau, J.-T., S. Feng, and S. Nadarajah. 2007. "Assessment of Hydrological Droughts for the Yellow River, China, Using Copulas." *Hydrological Processes: an International Journal* 21: 2157–63.

Sklar, A. 1959. "Fonctions de répartition à n dimensions et leursmarges." *Publications de l'Institut de statistique de l'Université de Paris* 8: 229–31.

Sun, C. 2013. "Bivariate Extreme Value Modeling of Wildland Fire Area and Duration." *Forest Science* 59: 649–60.

Supriya, P., M. Krishnavenia, and M. Subbulakshmia. 2015. "Regression Analysis of Annual Maximum Daily Rainfall and Stream Flow for Flood Forecasting in Vellar River Basin." *Aquatic Procedia* 4: 957–63.

Tidwell, V.C., H.D. Passell, S.H. Conrad, and R.P. Thomas. 2004. "System Dynamics Modeling for Community-Based Water Planning: Application to the Middle Rio Grande." *Aquatic Sciences* 66: 357–72.

Wang, Y., H. Ma, D. Sheng, and D. Wang. 2012. "Assessing the Interactions between Chlorophyll a and Environmental Variables Using Copula Method." *Journal of Hydrologic Engineering* 17: 495–506.

Zhang, L., and V.P. Singh. 2007. "Gumbel-Hougaard Copula for Trivariate Rainfall Frequency Analysis." *Journal of Hydrologic Engineering* 12: 409–19.