The application of the hybrid copula-GARCH approach in the simulation of extreme discharge values

Mohammad Nazeri Tahroudi1,2 · Mirali Mohammadi3 · Keivan Khalili4

Received: 13 July 2022 / Accepted: 17 October 2022 / Published online: 2 November 2022 © The Author(s) 2022

Abstract

Statistical analysis and simulation of annual maximum discharge values, while considering the corresponding maximum daily rainfall, provide a comprehensive view of flood management. This research presents the application of copula functions for simulating and modeling two variables of annual maximum discharge and corresponding precipitation. In this research, the performance of copula-based models and ARCH-based models including VAR-GARCH, copula, and copula-GARCH models was then evaluated to simulate the annual maximum discharge values. The simulation results of all three models were evaluated using NSE and NRMSE statistics. According to the 95% confidence intervals, the accuracy of all three models was confirmed. The correlation results of the studied pair variables confirmed the possibility of using copula-based models. The results of simulations revealed that a higher accuracy of the copula-GARCH approach compared with two models copula and VAR-GARCH. Considering 76% efficiency (NSE = 0.76) of the copula-GARCH approach, the results indicated 20 and 2.7% improvements in the performance of the proposed approach compared to both VAR-GARCH and copula models. The results also illustrated that by combining nonlinear ARCH models with copula-based simulations, the reliability of simulation results increased. The results obtained in this study suggest that the proposed method is very effective for increasing the certainty of frequency analysis of two variables. Because the copula-GARCH approach simulates the average values, the first and third quarters, as well as the amplitude of changes of 5 and 95% of the data better than the other two models.

Graphical abstract

Violin plot of AMD series in copula scale

Keywords ARCH models · ARMA · Frequency analysis · Gumbel-Hougaard · Heteroscedasticity

Introduction

Lack of data and the existence of statistical gap in the data for various reasons, is one of the problems in the field of data analysis and uncertainty in its results (Kim et al. 2014). There are several methods in the field of simulation and production of artificial data such as genetic
programing, artificial neural network, and regression that one of the most widely used models in this field is time series linear models (Salas et al. 1982; Nazeri-Tahroudi and Ramezani 2020; Shahidi et al. 2020). Vector autoregressive models called VAR, which are able to model several time series with different lags, are one of linear time series models (Shahidi et al. 2020). Using the VAR models, it can be examined the effect of different variables on each other or the simultaneous effect of a particular variable on several sites. This model is one of the time series models that can be combined with nonlinear models and the residuals of VAR can be modeled.

Also due to the heteroskedasticity in the residual series of the VAR model and other linear time series models, and also in order to model the variance of the residual series, the autoregressive conditional heteroscedastic (ARCH) model was investigated. ARCH is the most important nonlinear models (Bollerslev et al. 1992). These models have a high ability to connect with the residual series of linear models. In most cases, while combining with linear models, they could increase the accuracy of simulation and modeling.

Recent developments in field of copula functions and their combination with nonlinear ARCH models, have led to new research that can consider the effect of various parameters in modeling and produce multivariate analysis. The class of generalized autoregressive conditional heteroscedastic models (GARCH) first became popular and development for application to economic data, especially financial analysis. These models were very useful for modeling and predicting movements in asset return volatilities over time (Brooks et al. 2001).

Copulas are functions that produce a multivariate distribution function by connecting univariate marginal distributions (Nelsen 2006; Sklar 1959). Separate analysis of marginal distributions and their dependency structure is one of the most important products of copula functions (Ramezani et al. 2019; Salvadori et al. 2007a, b; Sérialdi et al. 2009). In recent research studies, much attention has been paid to integration and copula-based models regarding simulations. Copula-based simulation was first discussed in a research given by Bedford and Cook (2001 and 2002), though no programmable algorithms were then provided. Aas et al. (2009) and also Kurowicka and Cook (2006) developed sampling algorithms for vine copulas family, C- and D-vine. Kurowicka et al. (2007) provided insights into how the R-vine process in general. Díımann (2010) demonstrated how to write a sampling algorithm for the R-vine using a matrix. Yoo et al. (2016) analyzed and simulated precipitation values at 12 rain gauges in Korea using the copula-GARCH approach. The results obtained in this research showed that the proposed method could be very effective to quantify the uncertainty of bivariate drought frequency curves. Fathian et al. (2019) modeled the daily flow data in the Zarrinehroud Dam, located in northwest of Iran, using VAR and VARGARCH methods. The results indicated that the use of nonlinear model would improve the daily flow modeling efficiency. Also, the evaluation criteria indicated that the VAR-DVECH approach outperforms the VAR model. Cucina et al. (2019) used period autoregressive models (PAR) and its integration with genetic algorithm to seasonally model the flow discharge of South Saskatchewan and Colorado rivers. Comparing the results of flow discharge simulations with other studies performed in river flow analysis confirmed the efficiency of their proposed method. Eslami et al. (2022) also used hybrid models based on the GARCH model in their research on simulating the groundwater level. They stated that the addition of the GARCH model increased the modeling accuracy.

Various studies such as (Ayantobo et al. 2018; Guo and Wang 2017; Kim et al. 2019; Li and Zheng 2016) can also be mentioned in this regard. Yuan et al. (2020) applied the copula-GARCH approach to study agricultural price fluctuations. Their results displayed that the proposed approach has a great power in showing the interaction of parameters with each other. Using the simulation approach based on the copula-GARCH hybrid model, the values obtained from bivariate frequency curve would be more reliable. The best strength of copula-GARCH hybrid model is that it integrates conditional variability interdependencies between variables in the simulation (Huang et al. 2009; Yoo et al. 2016; Yuan et al. 2020; Nazeri Tahroudi et al. 2022a and b).

Heteroscedasticity in hydrological time series including rainfall has been widely reported worldwide (Wang et al. 2005; Modarres and Ouarda 2013; Shahidi et al. 2020), so the copula-GARCH hybrid model is more appropriate to simulate a rainfall time series. Multivariate linear time series models, including vector autoregressive and several conditional variance methods, are commonly used in finance and economics. These models have not been widely used in hydrology. Regarding modeling and simulation of time series, there are different models such as intelligent models, time series models, vector models, etc. Meanwhile, research shows that the performance of hybrid models as well as multivariate models is higher than that of conventional models. Also, time series models, both univariate and multivariate, are mean-based models, while ARCH family models model the variance of the residual series of ARMA family models. The combination of these two models enhances the certainty of modeling and simulation.

This research indicates the confidence intervals based on the simulated time series of annual maximum discharge (AMD) using the copula-GARCH hybrid method. On the other hand, in this research, for the first time in the copula-GARCH hybrid approach, vector autoregressive models have been used as base
model. Since hydrological phenomena are not independent, their dependency on the other meteorological and hydrological variables requires the use of multivariate models. Combining linear and nonlinear models of the ARCH family, the accuracy of modeling will be affected. Overall, copula-based models, which have been used in modeling and simulation of the hydrological variables, have high performance than linear models based on criteria statistics (Huang et al. 2009; Yoo et al. 2016; Yuan et al. 2020).

By examining different models in this field, it was found that there are many data-driven models in the field of simulations and prediction of various variables, which all of them are data-driven and applicable to any data and only time series in consideration. With the development of copula-based models, in this study, we tried to use this model, which is based on marginal distribution of data, to achieve the main goal (simulating the maximum daily rainfall (MDR)). The superiority of the copula-based model on the existing researches and its relationships has been evaluated and confirmed in various studies. But in this study, in order to achieve the purpose of the research, we tried to improve the accuracy of the model by combining the copula-based model with time series models. For this purpose, the VAR model was considered to be combined with the copula model as base model. The VAR model has advantages over other models in this family, considering that it is a multivariate model from the time series models. This model considers consecutive lags in the time series structure and performs the simulation in proportion to the best lag. Also, due to the heteroskedasticity in the residual series, the GARCH model was added to this collection. In fact, in the first step, the initial purpose of the research (simulating the AMD values) was determined and in the next steps, different models were added to improve the accuracy of the calculations. The main innovation of this research is the use of VAR model as a base model in the copula-GARCH hybrid model and finally two-dimensional analysis and simulation based on copula and ARCH models. Since the existence of heteroskedasticity in hydrological and meteorological values is undeniable, it must be considered in simulation studies that a combination of linear and nonlinear models is obtained. Accordingly, in this research, while examining the performance of the VAR-GARCH model, the simulation of AMD values affected by the maximum daily rainfall (MDR) was also investigated using the copula-based simulation and copula-GARCH approach.

Material and methods

A case study

In this research, using a hybrid approach based on copula functions and ARCH family models, the time series of MDR and AMD series in annual scale were used. A data set used belongs to theDashband sub-basin in the southwest of Lake Urmia in a statistical period time of 1970–2018. Lake Urmia is located in northwest of Iran. Figure 1 shows the location of the Dashband sub-basin in the Lake Urmia basin and Iran. The statistical properties of the studied data are also presented in Table 1.

Correlation analysis

The first step in the analysis of copula functions is to discuss the existence of dependency between the pair variables. There are several tools for measuring the dependency between two random variables. The most common of these methods is the Kendall’s tau coefficient. Kendall’s tau, denoted by τ, is defined as the probability of adaptation minus the probability of non-adaptation between two random variables X1 and X2.

The Kendall’s tau value between the continuous random variables X1 and X2 is defined as Eq. (1) (Hollander et al. 2013; Nazeri Tahroudi et al. 2021):

\[
\tau(X_1, X_2) = P(X_{11} - X_{21})(X_{12} - X_{22}) > 0 - P(X_{11} - X_{21})(X_{12} - X_{22}) < 0
\]

where \((X_{11} - X_{12})\) and \((X_{21} - X_{22})\) are independent and evenly distributed forms of \((X_1, X_2)\).

Vector autoregressive models, VAR

VAR is one of the flexible models for analyzing multivariate time series. The VAR model was introduced to describe the dynamic behaviors of the economic and financial series and their prediction. In addition to describing and forecasting data, the VAR model is also used for structural inferences and analysis policies. In structural analysis, specific hypotheses are imposed on the structure of the data under investigation. Also, the effects of unexpected shocks or innovations are summed up with the variables specified on the model variables. These effects are usually summarized with impact reaction and predicted error variance analysis functions. Also, due to the fact that the studied data is extreme, the use of lags in data can increase the accuracy of the simulations. By creating lags in the data, the modeling dimensions increase and more lags are considered. These consecutive lags bring the simulation results closer to reality, given that the extreme values can depend on previous events. This model focuses on analyzing constant covariance multivariate. VAR models in economics analysis have been introduced by (Sims 1980). If \(Y_t = (y_{1t}, y_{2t}, ..., y_{nt})^T\) represents the vector \((n \times 1)\) of the time series variables, then the VAR (p) model with a p-year base delay is as follows:
where $\Pi_i$ is equal to the coefficient $(n \times n)$ of the matrix and $\epsilon_t$ represents the matrix $(n \times 1)$ of the white noise values with the mean value of zero (non-dependent or independent) with constant covariance matrix $\Sigma$. For example, the equation of the two-variable VAR model is as follows:

$$Y_t = c + \Pi_1 Y_{t-2} + \cdots + \Pi_p Y_{t-p} + \epsilon_t, \quad t = 1, \ldots, T$$  \hspace{1cm} (2)

where $Y_t$ is the vector of the variables at time $t$, and $\Pi_i$ represents the matrix $(n \times n)$ of the coefficients. The matrix $\Pi(L)$ is defined as:

$$\Pi(L) = I_n - \Pi_1 L - \cdots - \Pi_p L^p$$

The equation can be rewritten as:

$$Y_t = c + \Pi(L) Y_t \epsilon_t$$

where $c$ is a constant vector and $\epsilon_t$ is a vector of white noise terms.

The covariance of the white noise terms $\epsilon_{1t}$ and $\epsilon_{2t}$ is $\sigma_{12}$ for $t = s$; otherwise it is zero. Note that, each equation has a similar regression of the remainder of $y_{1t}$ and $y_{2t}$. Hence, the VAR (p) model is just an indirect regression model with remaining variables and definitive terms as common regressions. From a user's perspective, the VAR (p) model is written as Eq. (5) (Shahidi et al. 2020):

$$\Pi(L) Y_t = c + \epsilon_t$$  \hspace{1cm} (5)

where $\Pi(L)$ is the matrix defined above. If the determinant value of $(I_n - \Pi_1 z - \cdots - \Pi_p z^p)$ is zero, then the VAR (p) will be static.

If the eigenvalues of a composite matrix have a modulus of less than one, it is outside the complex unit loop (with a modulus greater than one), or equivalent, if the eigenvalues of the composite matrix have a modulus less than one. It is assumed that the process in the past has been initiated from an infinite value, then it is a stable process of VAR (p) with a constant mean variance and covariance. If $Y_t$ in (Eq. 3) is a constant covariance, then the mean is given by Eq. (7):
\[
F = \begin{pmatrix}
\Pi_1 & \Pi_2 & \ldots & \Pi_n \\
I_n & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 \\
0 & 0 & \ldots & I_n
\end{pmatrix}
\] (6)

\[
\mu = (I_n - \Pi_1 - \ldots - \Pi_p)^{-1}c
\] (7)

After adjusting the mean of the VAR (p) model:

\[
Y_t - \mu = \Pi_1(Y_{t-1} - \mu) + \Pi_2(Y_{t-2} - \mu) + \cdots + \Pi_p(Y_{t-p} - \mu) + \varepsilon_t
\] (8)

The basic VAR (p) model may be very limited to show the main characteristics of the data. Specifically, other conditions of determinism such as a linear time trend or seasonal variables may be used to display data correctly. Additionally, random variables may also be required. The general form of the VAR (p) model with definitive terms and external variables is as follows:

\[
Y_t = \Pi_1Y_{t-1} + \Pi_2Y_{t-2} + \cdots + \Pi_pY_{t-p} + \Phi D_t + G X_t + \varepsilon_t
\] (9)

where \(D_t\) is a matrix (1x1) of the definite components, \(X_t\) denotes the matrix (m\times1) of the external variables and, \(\Phi\) and \(G\) represent a matrix of the model parameters (Sims 1980; Shahidi et al. 2020).

ARCH models

The autoregressive conditional heteroscedasticity (ARCH) model introduced by Engle (1982) allows the variance of a regression to change over time. The variance in one period is allowed to depend upon variables known from previous periods including the disturbances. The GARCH model is actually a generalized model of the ARCH family. The GARCH model was developed as an alternative to time series models based on the assumption of linearity between variables at different time stages, which cannot take into account the conditional dependence of variance or heteroscedasticity. Because of its power in modeling variables where changes are significant, it has been widely used especially in finance and economics (Duan 1996; Floros et al. 2007; Tse and Tsui 2002; Watanabe 2012). The GARCH model has many applications in simulation of hydrological time series (Modarres and Ouarda 2013; Ramezani et al. 2019; Wang et al. 2005; Yusof and Kane 2013). In this paper, the ARCH family model is developed to detect heteroscedasticity in the AMD series. The parameters of the GARCH model were estimated using the maximum likelihood method. The main idea of the ARCH models is that (a) the modified average investment return is distinct but dependent, and (b) the model is dependent and can be described by a simple quadratic function of the values before it. In summary, the ARCH model is assumed to be:

\[
\varepsilon_t = \sigma_t z_t \quad \text{and} \quad \sigma_t^2 = a_0 + \sum_{i=1}^{m} b_i \varepsilon_{i-i}^2
\] (10)

where \(\sigma_t^2\) is the conditional variance, \(\varepsilon_t\) denotes the error term or the remainder of the model with mean value of zero and variance of 1, \(a_0 \geq 0, b_i \geq 0\) are the model parameters, \(m\) indicates the order of the model, and \(Z_t\) is the time series of the desired parameter (Engle 1982).

Although the ARCH model is simple, it often requires many parameters to obtain the proper modeling process. For this reason, we have to look for alternative models (Mofat et al. 2017; Engle 1982; Bollerslev, 1992). The ARCH model was developed by Bollerslev (1986) to the Generalized ARCH model by entering the lagged conditional variance. The \(\varepsilon_t\) from VAR model is a sequence of GARCH \((a_i, \beta_j)\) process given by (Bollerslev, 1986):

\[
\varepsilon_t = H_t^{1/2} z_t
\] (11)

\[
H_t = a_0 + \sum_{i=1}^{m} a_i H_{t-i} z_{t-i}^2 + \sum_{j=1}^{n} \beta_j H_{t-j}
\] (12)

where \(z_t\) is an independent identically distributed sequence.

Copula theory

The introduction and presentation of copula functions is attributed to Sklar (1959), which described in a theory how 1-D distribution functions can be combined in the form of multivariate distributions. For two-dimensional continuous random variables \(X_1, X_2\) with marginal distribution functions \(F(x_i) = P_{x_i}(X_i \leq x_i)\), the joint distribution of \(X\) variables can be defined as Eq. (13):

\[
H_{X_1, X_2}(x_1, x_2) = P[X_1 \leq x_1, X_2 \leq x_2]
\] (13)

Copula is a function that joins the univariate marginal distribution functions to form a bivariate or multivariate distribution function. Accordingly, Sklar (1959) showed that the probability multivariate distribution of \(H\) using the marginal distribution functions and dependence structure can be expressed by the copula function \(C\) (Tahroudi et al. 2020; Tabatabaei et al. 2022):

\[
C(F_{X_1}(x_1), F_{X_2}(x_2)) = H_{X_1, X_2}(x_1, x_2)
\] (14)

where \(F_{X_i}(x_i)\) is the continuous marginal distribution functions of \(i\)th, and \(H_{X_1, X_2}\) is joint cumulative distribution \(X_1, X_2\). Since for continuous random variables, the cumulative marginal distribution functions are non-decreasing.
from zero to one, the copula of C can be considered as a
transform $H_{x_1,x_2}$ from $(-\infty, +\infty)^N$ to $[0,1]^N$. This
transformation separates the marginal distribution functions
from each other, and hence, the copula function of C
only relates to the relationship between the variables
and a complete description of the dependency structure
obtained (Nelsen 2006). For 2D copula, Sklar's theory is
as follows:

Suppose H is the joint distribution of variables $X_1$ and
$X_2$ by cumulative distributions $u = F_{X_1}(x_1)$ and $v = F_{X_2}(x_2)$.
In this case, there exists a 2D copula function in the set
of real numbers and is expressed in a form of Eq. (15).

$$H(x_1, x_2) = C(u, v) = C(F_{X_1}(x_1), F_{X_2}(x_2))$$  (15)

The 2-D copulas have the following characteristics: A) For each $u$ and $v$ we have in $[0, 1]$:

$$C(u, 0) = C(0, v) = 0$$  (16)

$$C(u, 1) = C(1, v) = 1$$  (17)

The above equations are called the boundary condition
of the 2-D copula function. With respect to these boundary
conditions, it is concluded from Eq. (13) that if one
of the marginal distribution functions has a value equal
to zero, then the value of the copula function (joint distri-
bution) is equal to zero. The studied copula functions, as
well as the range of dependent parameter, are presented in
Table 2. The dependence parameter $\theta$ is used to measure
the interdependency between $u$ and $v$.

- **Copula-based simulation**

Copula-based simulations were first discussed in Bedford
and Cook (2001) as well as Bedford and Cook (2002).

To obtain the sample $u_j, v_j, \ldots, u_d$ from the d variable
of copula, the following steps are performed:

$$w_j^{i,d} \sim U[0;1] \quad j = 1, \ldots, d$$  (18)

Then,

$$u_1 = w_1$$
$$u_2 := C^{-1}_{2|1}(w_2 | u_1)$$
$$\vdots$$
$$u_d := C^{-1}_{d|d-1,\ldots,1}(w_d | u_{d-1}, \ldots, u_1)$$

To determine the conditional distribution functions
$C_{ij|1,\ldots,i-1}, j = 1, \ldots, d$ required for the pair copula structure,
a feedback relation for the conditional distribution function
with $h$ function is used. For a bivariate copula with param-
eter $\theta_{ij}$, the $h$ functions are defined as follows:

$$h_{\overline{i}j}(u_i | u_j; \theta_{ij}) := \frac{\partial}{\partial u_i} C_{ij}(u_i, u_j; \theta_{ij})$$  (20)

$$h_{\overline{j}i}(u_j | u_i; \theta_{ij}) := \frac{\partial}{\partial u_i} C_{ij}(u_i, u_j; \theta_{ij})$$  (21)

| Table 2 The copula functions used in present research work |
|-----------------------------------------------------------|
| copula family                | $C(u, v)$                              | $\theta$               |
|-------------------------------|----------------------------------------|------------------------|
| Ali-Mikhail-Haq (AMH)         | $\frac{w}{1-\theta(1-w)(1-v)}$        | $-1 \leq \theta \leq 1$ |
| Clayton                       | $\max \left[ (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, 0 \right]$ | $\theta \geq 0$        |
| Frank                         | $-\frac{1}{\theta} \ln \left[ 1 + \frac{(e^{-x-1})(e^{-y-1})}{e^{x-1}} \right]$ | $\theta \neq 0$       |
| Galambos                      | $av \exp \left\{ \left[ (-\ln u)^{\theta} + (-\ln v)^{\theta} \right]^{1/2} \right\}$ | $\theta \geq 0$       |
| Gumbel-Hougaard (GH)          | $\exp \left\{ -\left[ (-\ln u)^{\theta} + (-\ln v)^{\theta} \right]^{1/2} \right\}$ | $\theta \geq 1$       |
| Plackett                      | $\left\{ 1 + (\theta - 1)(u + v) \right\} \left[ (1 + \theta(\theta - 1)uv)^{1/2} - 4\theta(\theta - 1)uv \right]^{1/2}$ | $\theta \geq 0$       |
| Farlie-Gumbel-Morgenstern (FGM) | $\left\{ 1 + (\theta - 1)(u + v) - \frac{1}{2} \theta \left[ (1 + (\theta - 1)(u + v))^2 - 4\theta(\theta - 1)uv \right]^{1/2} \right\}$ | $-1 \leq \theta \leq 1$ |
Models performance evaluation

In this study, we used the root-mean-square error (RMSE), mean absolute error (MAE), Bias and Nash–Sutcliffe efficiency coefficient (NSE) statistical parameters to evaluate the models performance as below:

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{Q}_i - Q_i)^2}{n}}.
\]

(22)

\[
MAE = \frac{1}{N} \sum_{i=1}^{N} |Q_i - \hat{Q}_i|.
\]

(23)

\[
NSE = 1 - \frac{\sum_{i=1}^{N} (\hat{Q}_i - Q_i)^2}{\sum_{i=1}^{N} (Q_i - \bar{Q})^2}.
\]

(24)

\[
BIAS = \frac{1}{N} \sum_{i=1}^{N} (Q_i - \hat{Q}_i).
\]

(25)

Lower (higher) is the value of RMSE and SSE (NSE) higher is the accuracy of the model. In the above equations, \(Q_i, \hat{Q}_i\) and \(\bar{Q}\) are the measured, simulated and mean data, respectively, and \(n\) is the number of data (Nash and Sutcliffe 1970; Akbarpour et al. 2020).

Summary of studied models

In this study, three models of copula, VAR-GARCH and copula-GARCH were used to simulate the values of AMD given by MDR values in studied area. Initially, the VAR model was fitted to a time series with different lags. The residual series of VAR model was extracted and fitted using the GARCH model. Since a structural change affects all residual values at any time, the structural stability of the residual series was investigated using the ordinary least squares (OLS) residuals and Cusum tests before developing the hybrid model. This test calculates the process of empirical fluctuations using a specific method from a generalized statistical framework.

Finally, the residual series resulting from the GARCH model was added to the VAR model. Thus, the first model, the VAR-GARCH model, was produced. The second model is the copula-based model, which is based on the conditional density of the copula functions. The third model is the copula-GARCH model. This approach was produced as follows: First, the studied pair variables (MDR and AMD) were fitted by the VAR model. In the next step, the residual series of the VAR model were extracted and fitted to the GARCH model. The two new series obtained in the previous step were fitted with copula functions and according to the joint analysis and conditional density of copula functions, a new series was generated and added to the base model. Thus, the third model was produced.

Results and discussion

In this research work, in order to simulate the AMD Series given by the corresponding MDR, the performance of models based on copula simulation (2D copulas) and ARCH-based models (copula-GARCH and VAR-GARCH) are compared. First, a correlation between the pairs of the mentioned variables was investigated. The scatter plot of the studied data on copula scale is presented in Fig. 2. In this Figure, the empirical contour lines, histogram of the studied data, and Kendall’s tau coefficient can be seen. The results of correlation study between the studied values using Kendall’s tau statistic also shows number 0.58, which is acceptable for the further studies. In various studies in the field of frequency analysis and copula-based simulations, different correlations such as 0.4 and above were reported to be acceptable (De Michele et al. 2007; Salvadori et al. 2007a, b; Tahroudi et al. 2020; Tabatabaei et al. 2022). The correlation between these time series using the coefficient of determination is about 0.75.

Simulation of AMD values using VAR-GARCH models and considering MDR values

Using two parameters of AMD and MDR, the performance of VAR-GARCH model in simulating the AMD values affected by MDR series was investigated. In the first step, AMD modeling with MDR values was performed using VAR model. The VAR model is a vector model of the family of time series models where the condition of normal and stationary data requires the use of these models (Kendall 1938; Mann 1945; Tahroudi et al. 2019; Wilcoxon 1946). To combine the VAR model with the nonlinear GARCH...
model, the residual series of the model were extracted as shown in Fig. 3.

Finally, by modeling the residual series of the VAR model using GARCH models, the hybrid VAR-GARCH model was developed. The results of the study on the stability of the residual series presented as Fig. 4. According to Fig. 4 and the confidence intervals set, the results of this test were also confirmed. As can be seen, the OLS-based Cusum process has not exceeded the confidence intervals. Hence, there is no evidence for structural change. After confirming the studied data, the accuracy of simulations was evaluated using 20% of the data at the end of the time series. By 100 simulations,
the simulation results with 95% confidence interval are presented in Fig. 5. The 95% confidence interval indicates the acceptable accuracy of the VAR-GARCH model in simulating AMD values affected by MDR. It is noted that two delays were used in the modeling. The calibration results of VAR-GARCH model showed that all the predicted data are in 95% confidence range and the accuracy of model is also confirmed. Multivariate models will be more applicable to dependent time series such as AMD and MDR values as compared to univariate models. Also, by increasing these factors, results of the modeling are more reliable, while the present results being consistent with the findings of (Fathian et al. 2019).

Simulation of AMD values using two-dimensional copula and considering MDR values

The prerequisite for using copula functions in copula-based modeling and simulations is the existence of correlations between pairs of variables; in this research it was investigated and confirmed using Kendall’s tau test. After examining the existing correlation and considering the GEV marginal distribution for the pair variables, the copula functions presented in Table 2 are examined. Results of evaluating the accuracy and efficiency of different copulas in the combined analysis of AMD and MDR values are shown plotted in Fig. 6. According to Bias, MAE, NSE, and RMSE criteria statistics, the Galambos copula was selected and introduced as the best copula for the pair variables. These evaluation criteria have been used to compare the empirical copula with other studied copulas. The accuracy and error rate of the copulas used have been investigated by using the Bias, MAE, NSE, and RMSE criteria statistics in comparison with the empirical copula.

The results of examining the copula structure of the above pair variables showed that according to the studied criteria, the copulas of Galambos and Gumbel-Hougaard presented almost the same results. From 7 copula functions studied, the FGM copula has a lower accuracy than that of the others, while the rest of copulas have acceptable accuracies and error rates in modeling the studied pair values. Finally, considering the superior copula, copula-based simulation was
performed. The simulation results of AMD values affected by MDR are presented in Fig. 7. In this Figure, the red values represent the correlation coefficient of the observed variables while the black values indicate a correlation coefficient of the simulated variables. The simulation results of AMD values showed that the correlation coefficient of the simulated pair variable is higher than the observed pair variable. Kendall’s tau coefficient presented in this section is based on pair variable simulation, which has increased by about 7% compared to the observed pair variable. Recent studies in the field of simulation of different values using two-dimensional and multidimensional copula functions show that copula functions have high accuracy and ability in two-variable simulations of meteorological and hydrological values (Bezak et al. 2017; Tahroudi et al. 2020).

**Simulation of AMD values using the copula-GARCH hybrid approach and considering MDR values**

AMD values were simulated based on the copula-GARCH approach using the residual series of the VAR-GARCH model. In the VAR model, a delay was used to model the studied values. The use of copula theory in modeling and simulations requires a correlation between the pair variables under consideration. According to Kendall’s tau (0.58) and correlation coefficient ($R^2 = 0.75$), this correlation indicates the interdependence of the two parameters on each other. Increasing the independent variable (MDR) strongly affects the dependent variable (AMD). Finally, the simulation results of AMD values using the copula-GARCH approach are presented in Fig. 8. The 95% confidence intervals showed that the simulation accuracy was acceptable and the simulation data were within the 95% confidence range.

Finally, Nash–Sutcliffe and NRMSE statistics were used to compare the performance of the studied models. The results of error and efficiency of the studied models are presented in Table 3. Nash–Sutcliffe statistic showed 0.76 efficiency for the copula-GARCH approach. This number, compared to the Nash–Sutcliffe values of the other models, indicates the high ability of the proposed approach to simulate AMD values.

Using the copula-GARCH approach in simulating AMD values indicates 23.4% and 3.6% reductions in error rates compared to the VAR-GARCH and copula-GARCH models, respectively. Also, using the proposed copula-GARCH approach, the model efficiency in simulating AMD values compared to the two models, VAR-GARCH and copula-GARCH, has been improved by about 20 and 2.7%, respectively. Examining three used models, the results showed that by modeling the conditional variance of the values studied by GARCH model, the simulation results of the vector autoregressive model have improved, as also confirmed by (Tesfaye et al. 2006; Nazeri and Khalili 2015 and 2018). In addition, the results of this study showed that by using the copulas, the accuracy and efficiency of the hybrid VAR-GARCH model can be improved, whose results are consistent with the research of (Yoo et al. 2016). Yuan et al. (2020) also confirmed the acceptable accuracy of the copula-GARCH hybrid model in agricultural commodity price modeling by combining copulas with GARCH models. The simulation results based on copula-GARCH approach are also presented in Fig. 9. In this diagram, the white circle represents the average of the data, the lower limit of the black rectangle represents the first quartile of the data, and the upper limit represents the third quartile of the data. The similarity of the diagram in the violin plot indicates the high certainty of the simulator model. According to the results of violin plot, it can be seen that on a copula scale, the copula-GARCH approach has been able to cover the range of data changes, and quarters one and three. As can be seen from Fig. 9, the range of changes in AMD values, as well as 5 and 95% of the values simulated by the copula-GARCH approach, are closer to the observed values, which enhances the

---

**Table 3** Results of performance evaluation and error rate of the studied models using NSE and NRMSE statistics

| Model          | NSE   | NRMSE |
|----------------|-------|-------|
| VAR-GARCH      | 0.63  | 60.26 |
| Copula         | 0.74  | 50.56 |
| Copula-GARCH   | 0.76  | 48.82 |

---

**Fig. 8** Results of generating the AMD series in test step using copula-GARCH approach

**Fig. 9** Violin plot of AMD series in copula scale
reliability of the model. This increase in accuracy in simulating meteorological and hydrological values has also been confirmed in the research by (Kim et al. 2019).

Conclusions

In this research, to evaluate the efficiency of copula-GARCH approach in simulating AMD values given by MDR in Dashband station in northwest of Iran and southwest of Lake Urmia, the copula-based model and ARCH-based models including VAR-GARCH and copula-GARCH were used. Due to the influence of AMD time series on MDR values, vector autoregressive model was used as base model. The VAR model provides relatively acceptable results due to the involvement of the effective parameter in modeling. At first, the studied values fitted by VAR model and the residual series of this model were extracted. The results of OLS test also confirmed the stability of the residual series for combination with ARCH models. Considering the marginal distribution of GEV, the results of pair analysis of the studied variables revealed that the Galambos and Gumbel-Hougaard copula had the best fit with the studied data. The results showed that the accuracy of two copula-based models (copula and copula-GARCH) was almost similar. Although the accuracy of all three models was acceptable given the 95% confidence intervals for the simulated values, the two copula-based models outperformed the time series vector model. With the copula-GARCH proposed approach, the Nash–Sutcliffe efficiency coefficient was improved by about 20% compared to the VAR-GARCH model, but there was no significant increase compared to the copula-GARCH model. The results indicated that with modeling based on linear, nonlinear, and copula models, the accuracy and efficiency of the simulations in each step increased. Use of this approach would improve performance in the production of important variables related to meteorological and hydrological analyses, which will be very useful in the management of water resources in any basin. After evaluating the results of AMD simulations value, the accuracy of all three models was confirmed and the structure of the copula-GARCH approach was selected and introduced as the best approach. On the other hand, due to the computational complexity of the copula-GARCH approach, the efficiency and error rates of this model improved by about 2.7% and 3.6% compared to the copula model, respectively. One of the most important applications of the copula-GARCH and VAR-GARCH models, is the study of the effect of parameters dependent on independent series, simulation, correct and complete the statistical gap. This subject can be useful for extending the statistical period for frequency analysis and data management.

Funding The research was not funded by any organization.

Availability of data and material Data collected by the author and secondary data are cited.

Declarations

Conflict of interest No conflict of interest involved.

Ethical approval In compliance with ethical standards.

Consent to participate Consent by all authors.

Consent for publication Consent by all authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Aas K, Czado C, Frigessi A, Bakken H (2009) Pair-copula constructions of multiple dependence. Insur Math Econ 44(2):182–198
Akbarpour A, Zeynali MJ, Tahroudi MN (2020) Locating optimal position of pumping Wells in aquifer using meta-heuristic algorithms and finite element method. Water Resour Manage 34(1):21–34
Ayantobo OO, Li Y, Song S, Javed T, Yao N (2018) Probabilistic modelling of drought events in China via 2-dimensional joint copula. J Hydrol 559:373–391
Bedford TJ, Cooke R (2001) Monte Carlo simulation of vine dependent random variables for applications in uncertainty analysis. ESREL 2003
Bedford T, Cooke RM (2002) Vines—a new graphical model for dependent random variables. Ann Stat 30(4):1031–1068
Bezak N, Rusjan S, Kramar Fijavž M, Mikoš M, Šraj M (2017) Estimation of suspended sediment loads using copula functions. Water 9(8):628
Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. J Econ 31(3):307–327
Bollerslev T, Chou RY, Kroner KF (1992) ARCH modeling in finance: a review of the theory and empirical evidence. J Econ 52(1–2):5–59
Brooks C, Burke SP, Persand G (2001) Benchmarks and the accuracy of GARCH model estimation. Int J Forecast 17(1):45–56
Cucina D, Rizzo M, Ursu E (2019) Multiple changepoint detection for periodic autoregressive models with an application to river flow analysis. Stoch Env Res Risk Assess 33(4):1137–1157
De Michele C, Salvadori G, Passoni G, Vezzoli R (2007) A multivariate model of sea storms using copulas. Coast Eng 54(10):734–751
Düffmann JF (2010) Statistical inference for regular vines and application
Duan J.C. (1996) A unified theory of option pricing under stochastic volatility-from GARCH to diffusion. Hong Kong University of Science and Technology

Engle RF (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econ J Econ Soc. https://doi.org/10.2307/1912773

Eslami P, Nasirian A, Akbarpour A, Nazeri Tahroudi M (2022) Groundwater estimation of Ghayen plain with regression-based and hybrid time series models. Paddy Water Environ. https://doi.org/10.1007/s10333-022-00903-9

Fathian F, Fakheri-Farou A, Ouarda TB, Dinpashoh Y, Nadoushani SSM (2019) Multiple streamflow time series modeling using VAR-MGARCH approach. Stoch Env Res Risk Assess 33(2):407–425

Floros C, Jaffry S, Lima GV (2007) Long memory in the Portuguese stock market. Stud Econ Finance. https://doi.org/10.1108/1086730701087370

Gao, W. Y. (2017) Assessment of variability in the hydrological cycle of the Loess Plateau, China: examining dependence structures of hydrological processes. In: AGU fall meeting abstracts, Vol 2017, pp GC41C-1028

Hollander M, Wolfe DA, Chicken E (2013) Nonparametric statistical methods, vol 751. Wiley

Huang JJ, Lee KJ, Liang H, Lin WF (2009) Estimating value at risk volatility-from GARCH to diffusion. Hong Kong University of Science and Business Media

Kim JE, Yoo J, Kim D, Kim H, Kim TW (2016) Application of copula functions to construct confidence intervals of bivariate drought frequency analysis. Water 11(10):2052

Kurowicka D, Cooke RM (2006) Uncertainty analysis with high dimensional dependence modelling. Wiley

Kurowicka D, Cooke RM (2007) Sampling algorithms for generating joint uniform distributions using the vine-copula method. Comput Stat Data Anal 51(6):2889–2906

Li F, Zheng Q (2016) Probabilistic modelling of flood events using the entropy copula. Adv Water Resour 97:233–240

Mann HB (1945) Nonparametric tests against trend. Econ J Econ Soc. E 45(3):315–324

Nazeri-Tahroudi M, Ramezani Y, Ahmadi F (2020) Estimation of dew point temperature in different climates of Iran using support vector regression. Időjárás/quart J Hungar Meteorol Serv 124(4):521–539

Nelsen R (2006) An introduction to copulas. Springer, New York

Pasha MM, Nazeri Tahroudi M, Ahmadi F (2019) Analyzing the droughts in Iran and its eastern neighboring countries using copula functions. Időjárás/quart J Hungar Meteorol Serv 123(4):435–453

Salas JD, Boes DC, Smith RA (1982) Estimation of ARMA models with seasonal parameters. Water Resour Res 18(4):1006–1010

Salvadori G, De Michele C, Kotegoda NT, Rosso R (2007a) Extremes in nature: an approach using copulas, vol 56. Springer Science & Business Media

Salvadori G, De Michele C, Kotegoda NT, Rosso R (2007b) Extremes in nature: an approach using copulas (Vol 56). Springer Science & Business Media

Serrinaldi F, Bonaccorso B, Cancelleri A, Grimaldi S (2009) Probabilistic characterization of drought properties through copulas. Phys Chem Earth Parts a/b/c 34(10–12):596–605

Shahidi A, Ramezani Y, Nazeri-Tahroudi M, Mohammadi S (2020) Application of vector autoregressive models to estimate pan evaporation values at the Salt Lake Basin, Iran IDÖJÁRÁS/Quart J Hungar Meteorol Serv 124(4):463–482

Sims, C. A. (1980). Money, Income, and Causality. "American Economic Review 62, September 1972, 540–552.-. Macroeconomics and Reality." Econometrica, 48, 1–48

Sklar M (1959) Fonctions de repartition an dimensions et leurs marges. Publ Inst Statist Univ Paris 8:229–231

Tabataeabi SM, Dastourani M, Eslamian S, Nazeri Tahroudi M (2022) Ranking and optimizing the rain-gauge networks using the entropy–copula approach (Case study of the Siminehrood Basin, Iran). Appl Water Sci 12(9):1–13

Tahroudi MN, Ramezani Y, Ahmadi F (2019) Investigating the trend and time of precipitation and river flow rate changes in Lake Urmia basin, Iran. Arabian J Geosci 12(6):1–13

Tahroudi MN, Ramezani Y, De Michele C, Mirabbasi R (2020) Analyzing the conditional behavior of rainfall deficiency and groundwater level deficiency signatures by using copula functions. Hydrocl Res 51(6):1332–1348

Tesfaye YG, Meerschaert MM, Anderson PL (2006) Identification of periodic autoregressive moving average models and their application to the modeling of river flows. Water Resour Res. https://doi.org/10.1029/2004WR003772

Tse YK, Tsui AKC (2002) A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. J Bus Econ Stat 20(3):351–362

Wang W, Van Gelder PHAJ, Vrijling JK, Ma J (2005) Testing and modeling autoregressive conditional heteroskedasticity of streamflow processes. Nonlinear Process Geophys 12(1):55–66

Watanabe T (2012) Quantile forecasts of financial returns using realized GARCH models. Jpn Econ Rev 63(1):68–80

Wilcoxson F (1946) Individual comparisons of grouped data by ranking methods. J Econ Entomol 39(2):269–270

Yoo J, Kim D, Kim H, Kim TW (2016) Application of copula functions to construct confidence intervals of bivariate drought frequency curve. J Hydro-Environ Res 62, September 1972, 540–552.-. Macroeconomics and Reality." Econometrica, 48, 1–48

Yuan X, Tang J, Wong WK, Sriboonchitta S (2020) Modeling co-movement among different agricultural commodity markets: a copula-GARCH approach. Sustainability 12(1):393

Yusof F, Kane IL (2013) Volatility modeling of rainfall time series. Theor Appl Climatol 113(1):247–258

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Authors and Affiliations

Mohammad Nazeri Tahroudi\textsuperscript{1,2} · Mirali Mohammadi\textsuperscript{3} · Keivan Khalili\textsuperscript{4}

Mohammad Nazeri Tahroudi
m_nazeri2007@yahoo.com

Keivan Khalili
k.khalili@urmia.ac.ir

\textsuperscript{1} Department of Water Engineering, University of Birjand, Birjand, Iran

\textsuperscript{2} Department of Water Engineering, Shahrekord University, Shahrekord, Iran

\textsuperscript{3} Department of Civil Engineering, Faculty of Engineering, Urmia University, P O Box 165, Urmia, Iran

\textsuperscript{4} Department of Water Engineering, Urmia University, Urmia, Iran