

Application of vine copulas to estimate dew point temperature

Yousef RAMEZANI^{1*}, Mohammad NAZERI TAHROUDI² and Matina PRONOOS SEDIGHI³

¹ Associate Professor, Department of Water Engineering, University of Birjand, Birjand, Iran.

² Ph.D. Graduate, Department of Water Engineering, University of Birjand, Birjand, Iran.

³ M.Sc. Graduate, Department of Water Engineering, University of Birjand, Birjand, Iran.

*Corresponding author; email: y.ramezani@birjand.ac.ir

Received: June 3, 2022; accepted: November 22, 2022

RESUMEN

En este estudio se investigó la precisión del modelo basado en cópulas para simular la temperatura del punto de rocío en varios climas de Irán, utilizando para ello simulaciones basadas en las cópulas de vid C, D y R. Al examinar diversas cópulas de vid y sus secuencias de árboles, se realizó la selección de la mejor cópula y secuencia de árboles según los criterios AIC, BIC y logaritmo de la verosimilitud. Los resultados mostraron que entre las cópulas de vid C, D y R, las cópulas de vid C y D se ajustan bien a las temperaturas mínima y máxima del aire y la temperatura del punto de rocío. Los resultados de la simulación se analizaron utilizando la raíz del error cuadrático medio (RMSE, por su sigla en inglés), coeficiente de eficiencia de Nash-Sutcliffe (NSE) y diagramas de violín. Los resultados mostraron que el modelo basado en cópula tiene una alta precisión en todas las estaciones. El RMSE mínimo (máximo) está relacionado con la estación Kerman (Ahvaz) con RMSE = 0.396 °C (0.617 °C). Además, el NSE mínimo (máximo) está relacionado con la estación Ahvaz (Urmia) con NSE = 0.925 (0.955). También, de acuerdo con la trama del violín, es posible ver la certeza aceptable del modelo basado en la cópula. Debido a la diversidad de las secuencias de árboles de las cópulas de vid y al uso de los estados rotacionales de las cópulas de vid internas, así como a la posibilidad de interferir con los parámetros efectivos en dimensiones altas, los resultados de la simulación son confiables y no tienen restricciones. Este modelo es el mejor para estimar la temperatura del punto de rocío debido a la cobertura total del rango de cambios en los datos.

ABSTRACT

In this study, the accuracy of the copula-based model in the simulation of the dew point temperature in various climates of Iran was investigated, using simulations based on vine copulas such as C-, D-, and R-vine copulas. By examining the various vine copulas and their tree sequences, the best copula and best tree sequence based on AIC, BIC, and log-likelihood were selected. The results show that based on the complete similarity in our case between C-, D- and R-vine copulas, the selected best C-vine copulas fit well the dependence between the minimum and maximum air temperatures and dew point temperature. The simulation results were analyzed using root mean square error (RMSE), Nash-Sutcliffe efficiency (NSE) coefficient, and violin plots. The results show that the copula-based model has high accuracy at all stations. The min (max) RMSE is related to Kerman (Ahvaz) station with RMSE = 0.396 °C (0.617 °C). Also, the min (max) NSE is related to Ahvaz (Urmia) station with NSE = 0.925 (0.955). Also, according to the violin plot, it is possible to appreciate the acceptable certainty of the copula-based model. Due to the diversity of the tree sequences of vine copulas and the use of the rotated states of the internal vine copulas, as well as the possibility of interfering with the effective parameters in high dimensions, the simulation results are reliable and have no restrictions. This model can be used as the best model to estimate dew point temperature due to the full coverage of the range of changes in data.

Keywords: conditional density, dew point temperature, rotated copulas, simulation, vine copulas.

1. Introduction

Access to accurate dew point temperature (DPT) data is important in different fields such as climate, agriculture, and hydrology. DPT is the temperature where air reaches its lowest pressure point to saturate. Radiation exchange between the atmosphere and the Earth's surface, water vapor pressure, and turbulent heat are the main factors influencing the formation of dew (Nazeri-Tahroudi and Ramezani, 2020). Shank et al. (2008) used artificial neural networks (ANN) to predict DPT for improvement in previous research, which included optimizing stop criteria and comparing seasonal models with year-round models by developing ANN to combine the results of seasonal models. They concluded that the methods used in their study had been effective in more accurately predicting the DPT during the year. Zounemat-Kermani (2012) evaluated the multiple linear regression and Levenberg-Marquardt (LM) models to estimate DPT in Ontario, Canada. By examining the mentioned models, they established that the LM model provides more accurate results than the other one. Nadig et al. (2013) used ANN models to predict air temperature and hourly DPT for 12 horizons. They improved the forecast accuracy of ANN models by combining two climate variables into a single ANN model for each forecast horizon. Combined models generated air temperature reduction for 10 of the 12 forecast horizons with an average MAE reduction of 1.93%. The hybrid models showed a significant reduction in the prediction anomalies for each of the 12 predicted horizons with an average reduction of 34.1%. Shiri et al. (2014) evaluated the accuracy of the ANN and Gaussian process (GP) models for estimating DPT in Korea. The results indicated that the performance of the ANN model is worse than the GP model. Kim et al. (2014) estimated DPT data on a daily scale in California (USA) using two soft computing techniques. By using the conventional regression model, the results indicated that soft computing techniques were more flexible and accurate to estimate daily DPT. Mohammadi et al. (2016) resorted to the ANFIS model to predict DPT data, using daily minimum air temperature, average air temperature, maximum air temperature, atmospheric pressure, relative humidity, horizontal global solar radiation, water vapor pressure, and sunshine hour

at two stations in Iran. They reported that the use of two series of water vapor pressure and daily minimum temperature increases the prediction accuracy of DPT. Baghban et al. (2016) used two statistical learning models, namely the least square support vector machine and the ANFIS model to predict DPT. A genetic algorithm was applied to optimize the parameters of the models. In this regard, a set of available data including 1300 data points were collected from the dew point of humid air in the temperature range between 0-50 °C. They stated that the present tools could be of great practical value to engineers and researchers, as precision instruments for simulating DPT. Mehdizadeh et al. (2017) used gene expression programming to estimate the daily FPT in two stations in northwestern Iran. They used different combinations of inputs. Their results showed that actual vapor pressure (e_a) is the most effective parameter in dew point temperature estimation. Nazeri-Tahroudi and Ramezani (2020) simulated DPT in various climates of Iran using support vector regression (SVR) and the ant colony optimization algorithm. To implement the optimized SVR model, different patterns including 1, 2, 3 and 7 inputs were used. Finally, the Pattern III (with two inputs including maximum and minimum air temperatures) was selected as the best one. Mehdizadeh et al. (2022) used nature-inspired optimization algorithms to estimate daily DPT at Rasht and Urmia stations, Iran. Optimization algorithms including the bee dragonfly algorithm and colony optimization were used in combination with the ANFIS model. The results showed that combining the dragonfly algorithm with ANFIS provided the most accurate results for both selected stations. Dong et al. (2022) simulated DPT using the optimized grasshopper algorithm (hybrid extreme gradient boosting with the grasshopper optimization algorithm [GOA-XGBoost]). The results showed that the GOA-XGBoost model had the best performance when compared to the random forest and XGBoost models. On a daily time scale, the random forest model overestimated results in the validation step. They also suggested that in future studies, the GOA-XGBoost model with e_a as input should be examined. Zhang et al. (2022) used the ANFIS method as a data-based technique to estimate the DPT. The results showed that this method is able

to identify the data pattern with high accuracy. In addition, this study fully compares ANFIS with two layer neural network models at different scales in which the ANFIS model provided high accuracy.

Various studies in the field of simulation and modeling of different meteorological and hydrological parameters show that the copula function has recently been considered by several researchers. Copula-based simulation has the best acceptable certainty due to the involvement of proportional data distribution in modeling and simulation as well as the use of conditional density of copula functions (de Michele and Salvadori, 2003; Salvadori and de Michele, 2004, 2007, 2010; de Michele et al., 2005; Salvadori et al., 2007, 2011; Tahroudi et al., 2020a,b; Khashei-Siuki et al., 2021). Copula functions increase the accuracy of simulations and improve their certainty by considering the marginal distribution of data in simulations. In addition, by increasing the dimension of simulations using the vine copulas, the effect of various parameters can be involved in the simulations (Nazeri-Tahroudi et al., 2022). Vine copulas with a wide variety of tree sequences allow the selection of different structures that improve the results. Khashei-Siuki et al. (2021) used vine copulas to simulate the potential of evapotranspiration at Birjand meteorological station, Iran. The results of the simulation of the potential evapotranspiration given by precipitation, air temperature and relative humidity using vine copulas showed a Nash-Sutcliffe efficiency (NSE) coefficient of 92%. The efficiency of the C-vine copula in the analysis of dependence and the results of the potential evapotranspiration simulation indicate the ability of vines in multivariate analysis.

Given the above, it will be difficult to provide a model for all regions of Iran due to different climates. Therefore, different models should be examined and validated in different climates, which is negligible in the case of copula-based models for this limitation. Due to the use of marginal distribution appropriate to the input data in copula-based models, the data conditions regardless of the required correlation will not interfere in the simulations. Therefore, in this study, by examining different vine copulas (C-, D- and R-vine), the performance of the copula-based model to predict DPT using minimum and maximum air temperatures (according to the Pattern III with

two inputs in the study of the Nazeri-Tahroudi and Ramezani [2020]) in five different climates of Iran was studied and compared.

2. Materials and methods

2.1 Study area

Iran is located in Asia between 25°-40° N and 44°-64° E, comprising an area of more than 1648 000 km² (Khalili et al., 2016). The climate of Iran includes the four seasons of the year in most parts. Minimum and maximum air temperature data for the period 1951-2014 were used in Urmia, Ahvaz, Babolsar, Kerman, Gorgan and Rasht stations (located in different climates of Iran), according to the optimized support vector regression (SVR) model in pattern III of the Nazeri-Tahroudi and Ramezani (2020) study with the two mentioned inputs. The studied climates were selected based on this study. Using actual and saturation vapor pressure, and average, minimum and maximum air temperature, as well as the FAO Penman-Monteith method, DPT values of the studied stations were calculated (Nazeri-Tahroudi and Ramezani, 2020). Finally, the DPT data extracted from the FAO Penman-Monteith method are simulated with the studied stations' minimum and maximum air temperatures using a copula-based model and three-dimensional analysis. Figure 1 shows the study area and the location of the studied stations. The characteristics of the studied meteorological stations are also presented in Table I (Nazeri-Tahroudi and Ramezani, 2020). Figure 2 shows the studied data in the selected stations.

2.2 R-vine Structure

R-vine copula is a general form of vine copulas, more flexible due to its high range of tree sequences and structures. This type of copula has a high flexibility, being able to select different structures with different tree sequences in each level (Morales-Nápoles, 2010; Nazeri Tahroudi et al., 2021).

According to Dißmann et al. (2013):

$$f(x) = \prod_{k=1}^d f_k(x_k) \times \prod_{i=1}^{d-1} \prod_{e \in E_i} c_{C_{e,a}, C_{e,b} | D_e} (F_{C_{e,a} | D_e}(x_{C_{e,a}} | x_{D_e}), (F_{C_{e,b} | D_e}(x_{C_{e,b}} | x_{D_e})) \quad (1)$$



Fig. 1. Location map of the selected stations in Iran (source: Nazeri-Tahroudi and Ramezani, 2020).

Table I. Annual statistics of stations used in the study.

Station	T _{min} (°C)	T _{avg} (°C)	T _{max} (°C)	Climate
Urmia	0.01	11.28	23.74	Moderately dry
Ahvaz	13.41	26.10	37.91	Dry
Babolsar	8.71	16.80	26.85	Wet
Kerman	0.55	17.03	29.89	Dry
Gorgan	7.56	17.75	31.49	Mediterranean
Rasht	6.04	16.22	23.97	Extremely wet

Source: (Nazeri-Tahroudi and Ramezani, 2020).

where $e = a, b$, $x_{D_e} = D_e$, and $x_{D_e} = x_i \mid i \in D_e$.

The log-likelihood function of the R -vine with parameter θ_{RV} and $E_1, E_2, \dots, E_{d-1}$ is calculated based on Eq. (2):

$$l_{RV}(\theta_{RV} \mid \mathbf{v}) = \sum_{k=1}^N \sum_{i=1}^{d-1} \sum_{e \in E_i} \log [c_{j(e),k(e),D(e)}(F(u_{i,j(e)} \mid u_{i,D(e)}) \mid \theta_{j(e),k(e),D(e)})] \quad (2)$$

where $u_i = (u_{i,1}, \dots, u_{i,d})' \in [0,1]^d$, $1, \dots, N \cdot c_{j(e),k(e),D(e)}$ equal to the bivariate copula density with edge e and parameter $\theta_{j(e),k(e),D(e)}$.

2.3 C- and D-vine structures

The drawable vine (D-vine) and canonical vine (C-vine) are two common structures of vines (Li et al., 2021). It is necessary to mention that the tree sequence of the vine copula in three dimensions is the same in C-, D-, and R-vine. An example of D-vine (right) and C-vine (left) are presented in Fig. 3 in three dimensions, according to which only the order of the nodes is different.

The main difference between C- and D-vine copulas is their tree sequence and the choice of roots and nodes in more than three dimensions. The C-vine copula has a star shape in the tree sequence, while the D-vine copula has a straight structure (Li et al., 2021). Similarly, in tree T_2 of the C-vine copula (Fig. 3, left), $e = 2,3 \mid 1$ is the edge; 1,3 and 1,2 are called node and root, respectively, while for the D-vine copula (Fig. 3, right), $e = 3,2 \mid 1$ is the edge, and 3,1 and 1,2 are called the node and root, respectively. Based on Aas and Berg (2009), the multivariate density of C-vine

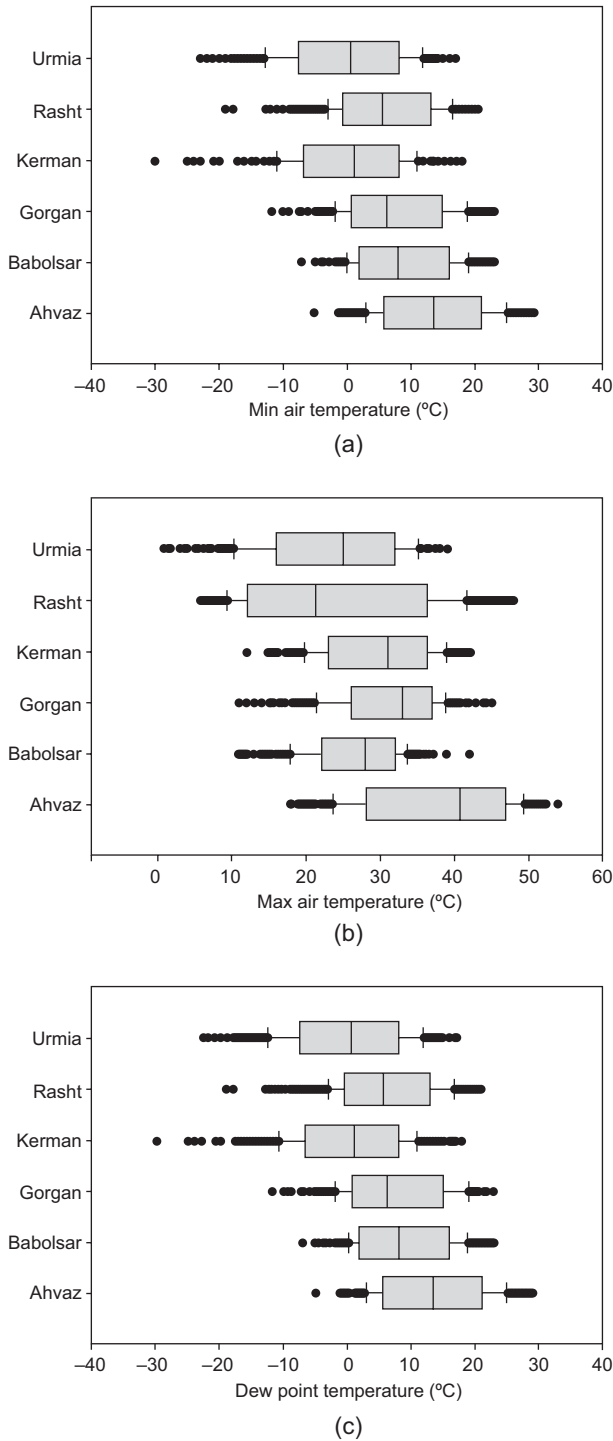


Fig. 2. Initial changes of the studied series at selected stations. (a) Minimum air temperature, (b) maximum air temperature, (c) dew point temperature

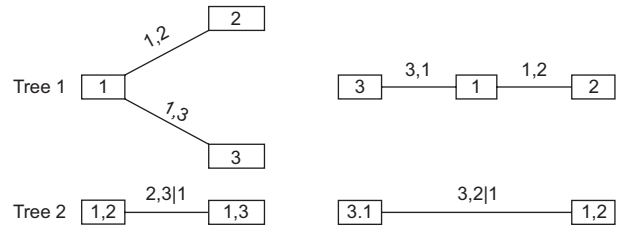


Fig. 3. Tree sequence of the 3-D copula (left: C-vine; right: D-vine).

and D-vine copulas are similar to Eqs. (3) and (5), respectively:

$$f(x) = \prod_{k=1}^d f_k(x_k) \times \prod_{i=1}^{d-1} \prod_{j=1}^{d-i} c_{i,i+j|1:(i-1)} \quad (3)$$

$$\{F(x_i | x_1, \dots, x_{i-1}), F(x_{i+j} | x_1, \dots, x_{i-1})\} | \theta_{i,i+j|1:(i-1)}$$

where $c_{i,i+j|1:(i-1)}$ is the density of the bivariate copula with the parameter $\theta_{i,i+j|1:(i-1)}$ (i_k, i_m means: $i_k, \dots, i_m$). The performance of the log-likelihood function for the C-vine copula with the parameter θ_{CV} is:

$$l_{CV}(\theta_{CV} | u) = \sum_{k=1}^N \sum_{i=1}^{d-1} \sum_{j=1}^{d-i} \log [c_{i,i+j|1:(i-1)}(F_{i|1:(i-1)}(F_{i+j|1:(i-1)} | \theta_{i,i+j|1:(i-1)})) \quad (4)$$

where $F_{j|1:i_m} = F(u_{kj} | u_{k,i_1}, \dots, u_{k,i_m})$.

The density of a D-vine copula is as follows:

$$f(x) = \prod_{k=1}^d f_k(x_k) \times \prod_{i=1}^{d-1} \prod_{j=1}^{d-i} c_{j,i+i|(j+1):(j+i-1)}(F(x_j | x_{j+1}, \dots, x_{j+i-1}), \quad (5)$$

$$(F(x_{j+i} | x_{j+1}, \dots, x_{j+i-1}) | \theta_{j,j+i|x_{j+1}, \dots, x_{j+i-1}})$$

where the output has more than d-1 tree and the pair-variables are determined by the input in each tree. The common definition of both C- and D-vine structures in three dimensions is:

$$f(x_1, x_2, x_3) = f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot c_{12}[F_1(x_1), F_2(x_2)] \cdot c_{23}[F_2(x_2), F_3(x_3)] \cdot c_{13|2}[F_1(x_1 | x_2), F_2(x_3 | x_2)] \quad (6)$$

2.4 Simulation based on vine copulas

Simulations based on R-vine copulas were presented originally in Bedford and Cooke (2001). Kurowicka

and Cooke (2005), Aas et al. (2009), and Czado, 2019 developed sampling algorithms for the C-vine and D-vine copulas. The following steps were presented to achieve a sample $u_1, \dots, u_d$ from a d variate copula:

$$\begin{aligned} \text{First : Sample } w_j &\sim U[0;1], j = 1, \dots, d \\ \text{Then : } u_1 &:= w_1 \\ u_2 &:= C_{2|1}^{-1}(w_2 | u_1) \\ &\vdots \\ u_d &:= C_{d|d-1, \dots, 1}^{-1}(w_d | u_{d-1}, \dots, u_1) \end{aligned} \quad (7)$$

Determining the $C_{j|j-1, \dots, 1}, j = 1, \dots, d$ conditional functions is needed to construct the pair-copula. For the desired conditional distribution function, this gives an iterative expression using the h-functions, which can be easily inverted recursively (Czado, 2019). To determine the conditional distribution functions $C_{j|j-1, \dots, 1}, j = 1, \dots, d$ that are required for the pair copula structure, the equation for the conditional distribution function is used together with the h-function. For the bivariate copula $C_{ij}(u_i, u_j; \theta_{ij})$ with parameter θ_{ij} , h, functions are defined as follows (Aas et al., 2009):

$$h_{i|j}(u_i | u_j; \theta_{ij}) := \frac{\partial}{\partial u_j} C_{ij}(u_i, u_j; \theta_{ij}) \quad (8)$$

$$h_{j|i}(u_j | u_i; \theta_{ij}) := \frac{\partial}{\partial u_i} C_{ij}(u_i, u_j; \theta_{ij}) \quad (9)$$

The above equations can be calculated using BiCop() function in R (Aas et al., 2009; Bevacqua et al., 2017).

2.5. Evaluation criteria

The Bayesian information criterion (BIC), root mean square error (RMSE), Akaike information criterion (AIC) and log likelihood commonly applied to select the best copula (Nash and Sutcliffe, 1970; Harville, 1974; Zhang and Singh, 2006; Ma and Sun, 2011; Khozaymehnezhad and Nazeri-Tahroudi, 2020; Nazeri Tahroudi et al., 2021; Raji et al., 2022).

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (p_i - \bar{p}_i)^2} \quad (10)$$

$$AIC = 2m - 2\ln(L) \quad (11)$$

$$BIC = N \ln \left(\frac{1}{N} \sum_{i=1}^N (p_i - \bar{p}_i)^2 \right) + m \ln(N) \quad (12)$$

$$NSE = 1 - \left(\frac{\sum_{i=1}^N (p_i - \bar{p}_i)^2}{\sum_{i=1}^N (p_i - p_{ave})^2} \right) \quad (13)$$

where $\bar{p}_i$ and p_i are equal to the simulated and observed values, respectively. N is the number of data, m is the number of parameters and L is the maximum value of the probability function for the model. p_{ave} is the mean of observed values. The calculations were done in the R environment using packages VineCopula and CDVineCopulaConditional (Brechmann and Schepsmeier, 2013; Bevacqua et al., 2017). The flowchart of the proposed methodology is presented in Figure 4.

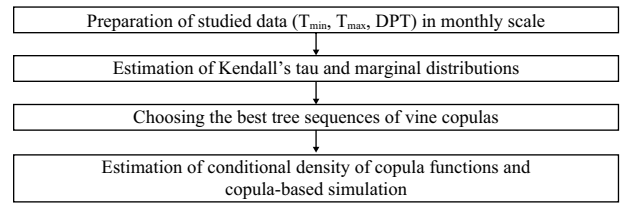


Fig. 4. Flowchart of the proposed methodology.

3. Results and discussion

The first step is to examine the correlation of the studied variables using Kendall's tau, which is the basis of copula research. As mentioned, the correlation between minimum and maximum air temperatures of the studied stations, was investigated using Kendall's tau. The results (Table II) in the study area show that there is an acceptable correlation between minimum and maximum air temperatures, and dew point temperature at all stations. Therefore, the main and initial condition of simulation and study of copula functions is satisfied.

3.1 Investigation of tree sequences of the studied variables

In this study, common marginal distributions in hydrology and water resources were used to check the marginal distributions according to the studied data. The best marginal distributions were selected

Table II. Results of the correlation between minimum ($T_{\min}$) and maximum ($T_{\max}$) air temperature, and dew point temperature (DPT) using Kendall's tau.

Station	$T_{\min}$ - $T_{\max}$	$T_{\min}$ - DPT	$T_{\max}$ - DPT
Ahvaz	0.874	0.823	0.803
Babolsar	0.742	0.839	0.680
Gorgan	0.715	0.839	0.643
Kerman	0.836	0.806	0.770
Rasht	0.845	0.789	0.789
Urmia	0.874	0.871	0.797

using two statistics RMSE and NSE and presented in Table III. According to Table III, GEV is the best distribution on the studied data.

In this study, after confirming the correlation between the studied variables and selecting the best marginal distribution based on Table III, the vine copulas and their tree sequences were investigated. In this regard, the types of copula functions related to this family, including C-, D- and R-vine copulas, and their independent and dependent states, were investigated. Table IV presents the results of the best tree sequences of the mentioned copulas. It should be noted that AIC, BIC and log-likelihood statistics were used to select the best tree sequence and the best copula. Since the tree sequence of the vine copula in three dimensions is the same in C-, D-, and R- vine, in this study, C-vine was chosen as the best vine copula. By examining vine copulas in the multivariate simulation of dew point temperature given by $T_{\min}$ and $T_{\max}$, the best tree sequence of C-vine was introduced as Table IV, where it can be seen that the selected tree sequences were well able to maintain correlation to the last tree. Also, rotated copulas were used to cover the correlation in all directions. Rotated copulas examine the correlation in different directions such as 90, 180 and 270 degrees. In the first and second trees, at most of the studied stations, the Frank copula was selected as the best copula function.

3.2 Copula-based simulation

Finally, by confirming the tree sequence of the studied vine copulas, the copula-based simulation of dew point temperature given by the minimum and maximum air temperatures was obtained at all studied

Table III. Best marginal distributions of the studied data.

Station	Parameter	Marginal distribution	RMSE	NSE
Ahvaz	$T_{\min}$	GEV	6.47	0.95
	$T_{\max}$	GEV	5.66	0.96
	DPT	Normal	5.51	0.96
Babolsar	$T_{\min}$	GEV	5.21	0.97
	$T_{\max}$	GEV	4.27	0.98
	DPT	Normal	6.11	0.96
Gorgan	$T_{\min}$	Rayleigh	5.63	0.96
	$T_{\max}$	GEV	3.38	0.99
	DPT	GEV	5.75	0.96
Kerman	$T_{\min}$	GEV	5.22	0.97
	$T_{\max}$	GEV	4.79	0.97
	DPT	GEV	3.52	0.99
Rasht	$T_{\min}$	Rayleigh	5.42	0.96
	$T_{\max}$	Rayleigh	6.49	0.95
	DPT	Normal	5.15	0.97
Urmia	$T_{\min}$	Normal	4.94	0.97
	$T_{\max}$	GEV	4.29	0.98
	DPT	GEV	3.24	0.99

RMSE: root mean square error; NSE: Nash-Sutcliffe efficiency; GEV: generalized extreme value; DPT: dew-point temperature.

stations. Figure 5 shows the simulation results of DPT at the Urmia station.

Figure 6 shows RMSE and NSE resulting from the simulation of DPT using the copula-based model. The lowest RMSE (0.396 °C) is related to Kerman station and the highest (0.617 °C) that is related to Ahvaz station. The NSE of the copula-based model in the simulation of DPT at all studied stations is higher than 90%, being the lowest (highest) related to Ahvaz station (Urmia station) with NSE of 925% (955%). In the optimized SVR model obtained from the study of the Nazeri-Tahroudi and Ramezani (2020), Pattern I (with seven inputs) and Pattern III (with two inputs) were introduced as the best models. Based on the number of inputs used in Pattern I, these researchers introduced Pattern III with two inputs (minimum and maximum air temperatures) as the best model (recommended as more user-friendly); however, Pattern I has yielded better results. The results of the copula-based model showed that compared to

Table IV. Results of the best tree sequences of vine copulas at the studied stations.

Station	Tree	Edge	Family	Copula	Parameter	Tau
Ahvaz	1	1,2	5	Frank copula	27.90	0.87
		3,1	5	Frank copula	19.30	0.81
	2	3,2:1	16	Joe 180	1.20	0.10
	C-vine	log likelihood: 1495		AIC: -2984	BIC: -2971	
Babolsar	1	1,2	5	Frank copula	12.98	0.73
		3,1	5	Frank copula	21.77	0.83
	2	3,2:1	1	Gaussian copula	-0.11	-0.07
	C-vine	log likelihood: 1563		AIC: -3120	BIC: -3106	
Gorgan	1	1,2	14	Gumbel 180	2.92	0.66
		3,1	5	Frank copula	21.77	0.83
	2	3,2:1	2	Student t	-0.19	-0.12
	C-vine	log likelihood: 1768		AIC: -3430	BIC: -3326	
Kerman	1	1,2	10.231	Frank copula	21.26	0.83
		3,1	5	Frank copula	18.15	0.80
	2	3,2:1	16	Joe 180	1.20	0.10
	C-vine	log likelihood: 1773		AIC: -3540	BIC: -3256	
Rasht	1	2,1	5	Frank copula	23.10	0.84
		3,2	6	Joe copula	7.16	0.76
	2	3,1:2	2	Student t	-0.20	-0.13
	C-vine	log likelihood: 1933		AIC: -3857	BIC: -3839	
Urmia	1	1,2	5	Frank copula	27.59	0.86
		3,1	5	Frank copula	27.35	0.86
	2	3,2:1	34	Gumbel 270	-1.20	-0.17
	C-vine	log likelihood: 2234		AIC: -4461	BIC: -4447	

1: $T_{\min}$; 2: $T_{\max}$; 3: DPT; BIC Bayesian information criterion; AIC: Akaike information criterion.

Pattern I and Pattern III (based on the optimized SVR model), the copula-based model has a higher accuracy according to the RMSE. The percentage of improvement of the RMSE in the copula-based model compared the Pattern I and Pattern III in the study of Nazeri-Tahrudi and Ramezani (2020) is presented in Table V.

In Table V it can be seen that the copula-based model provides more RMSE than Pattern I at two stations (Gorgan and Ahvaz stations). However, compared to the proposed method of Nazeri-Tahrudi and Ramezani (2020), the results show that the copula-based model is more accurate at all stations and has improved the average accuracy by about 38%. In addition to the RMSE, the NSE of the model was evaluated and the percentage improvement of the

NSE of the copula-based model compared to patterns I and III is presented in Table VI.

Using the results of Table VI, it can be seen that similarly to RMSE, NSE of the copula-based model in the simulation of DPT compared to the Pattern I at Urmia, Kerman, Babolsar and Rasht stations increased by 0.95, 1.27, 0.107 and 2.71%, respectively. This increase in NSE values is also present in the simulation of DPT using the copula-based model compared to Pattern III at all stations. On average, the copula-based model was able to increase the NSE of the copula-based model by 6% at all studied stations. In order to evaluate the certainty of the studied models, the violin plot is also presented in Figure 7, according to which there is a good agreement between the observed and simulated DPT by

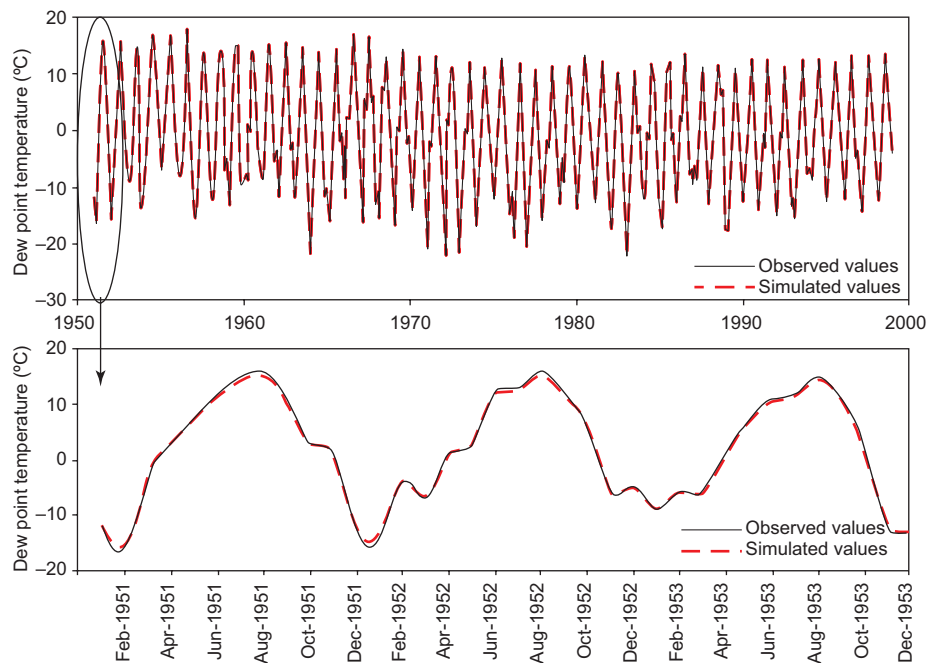


Fig. 5. Simulation results of dew point temperature using the copula-based model at the Urmia station.

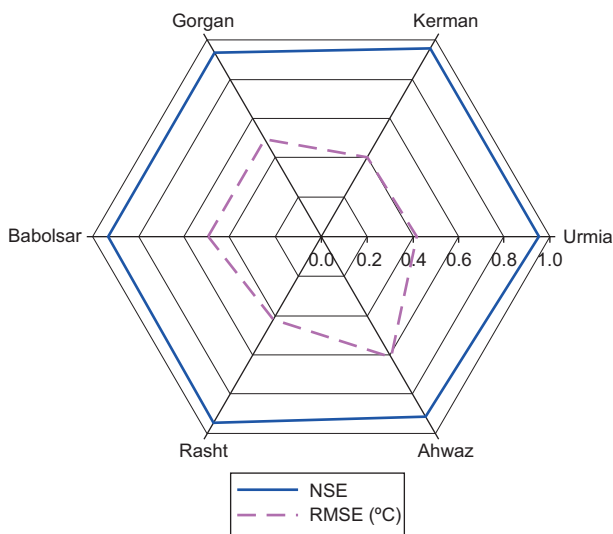


Fig. 6. Results of the Nash-Sutcliffe efficiency (NSE) and root mean square error (RMSE) of the copula-based model in simulation of the dew point temperature.

the copula-based model at all stations. Also, the results show acceptable accuracy and certainty of the copula-based model in simulation of DPT.

A comparison is made between the results of the different models and the number of inputs in Table VII. At Kerman station, the results showed that the copula-based model was able to provide better results than the adaptive neuro fuzzy inference system (ANFIS) mentioned in Mohammadi et al. (2016). According to Table VII, results show that the copula-based model has more accuracy than other studied models. In the violin plots in Figure 7, the white circle is the average of the data, the black rectangle represents the major changes in the data. The upper and lower limits of the black rectangle indicate the third and first quartiles, respectively.

4. Conclusion

One of the problems in estimating DPT is the number of parameters required. Therefore, using a model that can adequately estimate DPT with fewer parameters is important. Recently, with the development of strong processors, there are numerous methods and different software for estimating non-existent data, forecasting and data generation. With the development of copula-based models and their multivariate modeling and

Table V. Percentage improvement of the RMSE (°C) of the copula-based model compared to Pattern I and Pattern III in the study of the Nazeri-Tahroudi and Ramezani (2020).

Station	Optimized SVR (Pattern I)	Optimized SVR (Pattern III)	Copula-based model	Percentage improvement of the RMSE of the copula-based model compared to Pattern I	Percentage improvement of the RMSE of the copula-based model compared to Pattern III
Urmia	0.485	0.838	0.421	13.196	49.761
Kerman	0.495	2.412	0.396	20.000	83.582
Gorgan	0.428	0.611	0.496	-15.888	18.822
Babolsar	0.497	0.69	0.495	0.402	28.261
Rasht	0.594	0.801	0.422	28.956	47.316
Ahvaz	0.435	0.642	0.617	-41.839	3.894

SVR: support vector regression; RMSE: root mean square error.

Table VI. Percentage improvement of the copula-based model NSE compared to patterns I and III in the study of the Nazeri-Tahroudi and Ramezani (2020).

Station	Optimized SVR (Pattern I)	Optimized SVR (Pattern III)	Copula-based model	Percentage improvement of the NSE of the copula-based model compared to the Pattern I	Percentage improvement of the NSE of the copula-based model compared to the Pattern III
Urmia	0.946	0.888	0.955	0.951	7.545
Kerman	0.942	0.795	0.954	1.271	20.000
Gorgan	0.948	0.926	0.936	-1.26	1.080
Babolsar	0.932	0.912	0.933	0.107	2.303
Rasht	0.921	0.905	0.946	2.714	4.530
Ahvaz	0.947	0.923	0.925	-3.285	0.217

SVR: support vector regression; NSE: Nash-Sutcliffe efficiency.

simulation, simulation of meteorological and hydrological parameters has also been developed. The type of input data and the number of these inputs to the models are different. In addition to the above, a data selection must be made before simulation and modeling to select appropriate inputs with acceptable correlation as inputs. The meteorological data from different climates of Iran used in the study of Nazeri-Tahroudi and Ramezani (2020), were used to estimate the dew point temperature. Based on de martonne classification, Kerman and Ahvaz meteorological stations with dry climate, Babolsar station with wet climate, Gorgan station with Mediterranean climate, Rasht station with extremely

wet climate and Urmia station with moderately dry climate were selected. In order to simulate the dew point temperature of the studied stations, after confirming the correlation between the variables by the Kendall's tau, different vine copulas were examined. Based on AIC, BIC, Loglike criteria and the complete similarity between C-, D-, and R-vine copulas, the results showed that C-vine copulas with selected tree sequences fit well with the studied data. The results of the simulations were analyzed using RMSE, NSE and violin plot. The simulation results of the dew point temperature given by the minimum and maximum air temperatures using the copula-based model showed that the accuracy of

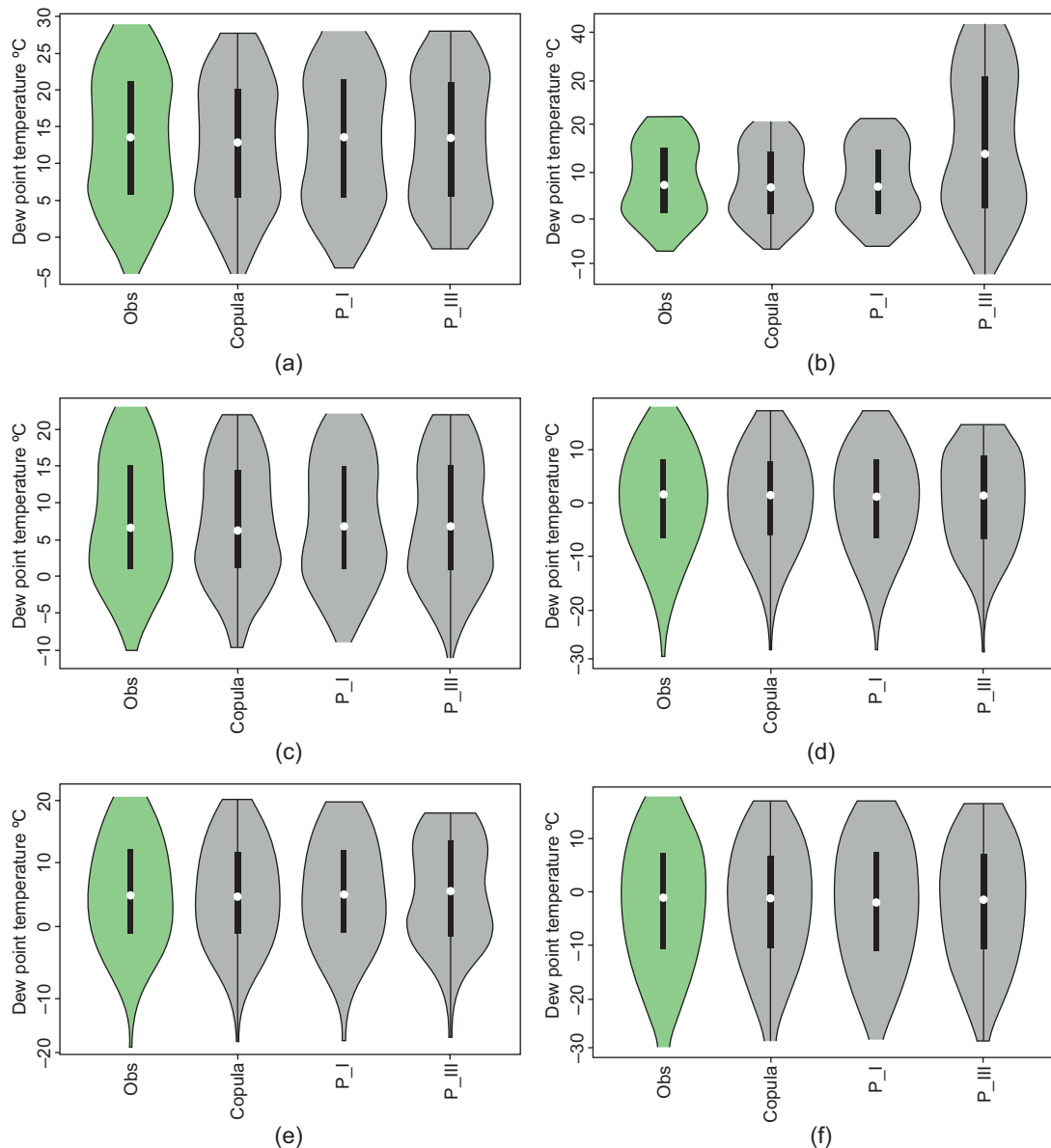


Fig. 7. Violin plots of the studied models in the simulation of dew point temperature at the following stations: (a) Ahvaz, (b) Babolsar, (c) Gorgan, (d) Kerman, (e) Rasht, and (f) Urmia.

the copula-based model was higher than the Pattern I and Pattern III in the study of Nazeri-Tahroudi and Ramezani (2020). The certainty of the proposed model was also confirmed through violin plots. Since the proposed model is based on the tree sequence according to the studied data, there is no geographical and climatic limitation regarding its implementation.

Acknowledgments

The authors would like to acknowledge the financial support of University of Birjand for this research under contract number 1399/D/18294. Also, the authors would like to thank Iran's Meteorological Organization (IRIMO) for providing the data.

Table VII. Comparison between the results of the present study and previous studies.

Reference	Studied model	Number of inputs	Region	RMSE ($^{\circ}$ C)
Mohammadi et al. (2016)	ANFIS	2	Iran (Kerman)	0.833
Mohammadi et al. (2016)	ANFIS	2	Iran (Tabas)	0.544
Hubbard et al. (2003)	Regression based	2	USA (six stations)	3.22
Zounemat-Kermani (2012)	MLR	4	Canada (Geraldton)	0.931
Zounemat-Kermani (2012)	ANN	4	Canada (Geraldton)	0.904
Kim et al. (2014)	GRNN	2	USA (U.C. Riverside)	1.20
Kim et al. (2014)	GRNN	2	USA (Durham)	1.84
Kim et al. (2014)	MLP	2	USA (U.C. Riverside)	1.29
Kim et al. (2014)	MLP	2	USA (Durham)	1.89
Nazeri-Tahroudi and Ramezani (2020)	Optimized SVR (Pattern I)	7	Iran (Urmia)	0.485
			Iran (Kerman)	0.495
			Iran (Gorgan)	0.428
			Iran (Bobolsar)	0.497
			Iran (Rashat)	0.594
			Iran (Ahvaz)	0.435
Nazeri-Tahroudi and Ramezani (2020)	Optimized SVR (Pattern II)	3	Iran (Urmia)	0.805
			Iran (Kerman)	2.971
			Iran (Gorgan)	0.682
			Iran (Bobolsar)	0.696
			Iran (Rashat)	0.878
			Iran (Ahvaz)	0.660
Nazeri-Tahroudi and Ramezani (2020)	Optimized SVR (Pattern III)	2	Iran (Urmia)	0.838
			Iran (Kerman)	2.412
			Iran (Gorgan)	0.611
			Iran (Bobolsar)	0.690
			Iran (Rashat)	0.801
			Iran (Ahvaz)	0.642
Nazeri-Tahroudi and Ramezani (2020)	Optimized SVR (Pattern IV)	1	Iran (Urmia)	1.362
			Iran (Kerman)	2.734
			Iran (Gorgan)	1.394
			Iran (Bobolsar)	1.268
			Iran (Rashat)	0.757
			Iran (Ahvaz)	1.369
Present study	Copula-based model	3 ($T_{\min}$, $T_{\max}$, DPT)	Iran (Urmia)	0.421
			Iran (Kerman)	0.396
			Iran (Gorgan)	0.496
			Iran (Bobolsar)	0.495
			Iran (Rashat)	0.422
			Iran (Ahvaz)	0.617

RMSE: root mean square error; ANFIS: adaptive neuro fuzzy inference system; MLR: multiple linear regression; ANN: artificial neural networks; SVR: support vector regression; MLP: multilayer perceptron; GRNN: generalized regression neural network.

References

- Aas K, Berg D. 2009. Models for construction of multivariate dependence – A comparison study. *The European Journal of Finance* 15: 639-659. <https://doi.org/10.1080/13518470802588767>
- Aas K, Czado C, Frigessi A, Bakken H. 2009. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics* 44: 182-198. <https://doi.org/10.1016/j.insmatheco.2007.02.001>
- Baghban A, Bahadori M, Rozyn J, Lee M, Abbas A, Bahadori A, Rahimali A. 2016. Estimation of air dew point temperature using computational intelligence schemes. *Applied Thermal Engineering* 93: 1043-1052. <https://doi.org/10.1016/j.applthermaleng.2015.10.056>
- Bedford T, Cooke R. 2001. Probabilistic risk analysis: Foundations and methods. Cambridge University Press. <https://doi.org/10.1017/CBO9780511813597>
- Bevacqua E, Maraun D, Hobæk Haff I, Widmann M, Vrac M. 2017. Multivariate statistical modelling of compound events via pair-copula constructions: analysis of floods in Ravenna (Italy). *Hydrology and Earth System Sciences* 21: 2701-2723. <https://doi.org/10.5194/hess-21-2701-2017>
- Brechmann EC, Schepsmeier, U. 2013. Modeling dependence with C- and D-vine copulas: The R package CDVine. *Journal of Statistical Software* 52: 1-27. <https://doi.org/10.18637/jss.v052.i03>
- Czado C. 2019. Analyzing dependent data with vine copulas: A practical guide with R. *Lecture Notes in Statistics*. Springer Nature Switzerland, Cham. <https://doi.org/10.1007/978-3-030-13785-4>
- De Michele C, Salvadori G. 2003. A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. *Journal of Geophysical Research: Atmospheres* 108: D2. <https://doi.org/10.1029/2002JD002534>
- De Michele C, Salvadori G, Canossi M, Petaccia A, Rosso R. 2005. Bivariate statistical approach to check adequacy of dam spillway. *Journal of Hydrologic Engineering* 10: 50-57. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2005\)10:1\(50\)](https://doi.org/10.1061/(ASCE)1084-0699(2005)10:1(50))
- Dißmann J, Brechmann EC, Czado C, Kurowicka D. 2013. Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis* 59: 52-69. <https://doi.org/10.1016/j.csda.2012.08.010>
- Dong J, Zeng W, Lei G, Wu L, Chen H, Wu J, Huang J, Gaiser T, Srivastava AK. 2022. Simulation of dew point temperature in different time scales based on grasshopper algorithm optimized extreme gradient boosting. *Journal of Hydrology* 606: 127452. <https://doi.org/10.1016/j.jhydrol.2022.127452>
- Harville DA. 1974. Bayesian inference for variance components using only error contrasts. *Biometrika* 61: 383-385. <https://doi.org/10.1093/biomet/61.2.383>
- Hubbard KG, Mahmood R, Carlson C. 2003. Estimating daily dew point temperature for the Northern Great Plains using maximum and minimum temperature. *Agronomy Journal* 95: 323-328. <https://doi.org/10.2134/agronj2003.3230>
- Khalili K, Tahoudi MN, Mirabbasi R, Ahmadi F. 2016. Investigation of spatial and temporal variability of precipitation in Iran over the last half century. *Stochastic Environmental Research and Risk Assessment* 30: 1205-1221. <https://doi.org/10.1007/s00477-015-1095-4>
- Khashei-Siuki A, Shahidi A, Ramezani Y, Nazeri Tahroudi M. 2021. Simulation of potential evapotranspiration values based on vine copula. *Meteorological Applications* 28: e2027. <https://doi.org/10.1002/met.2027>
- Khozayemehnezhad H, Nazeri-Tahroudi M. 2020. Analyzing the frequency of non-stationary hydrological series based on a modified reservoir index. *Arabian Journal of Geosciences* 13: 232. <https://doi.org/10.1007/s12517-020-5226-y>
- Kim S, Singh VP, Lee CJ, Seo Y. 2014. Modeling the physical dynamics of daily dew point temperature using soft computing techniques. *KSCE Journal of Civil Engineering* 19: 1930-1940. <https://doi.org/10.1007/s12205-014-1197-4>
- Kurowicka D, Cooke RM. 2005. Distribution-free continuous Bayesian belief nets. *Modern Statistical and Mathematical Methods in Reliability* 10: 309-322. https://doi.org/10.1142/9789812703378_0022
- Li H, Huang G, Li Y, Sun J, Gao P. 2021. A C-vine copula-based quantile regression method for streamflow forecasting in Xiangxi River basin, China. *Sustainability* 13: 4627. <https://doi.org/10.3390/su13094627>
- Ma J, Sun Z. 2011. Mutual information is copula entropy. *Tsinghua Science & Technology* 16: 51-54. [https://doi.org/10.1016/S1007-0214\(11\)70008-6](https://doi.org/10.1016/S1007-0214(11)70008-6)
- Mehdizadeh S, Behmanesh J, Khalili K. 2017. Application of gene expression programming to predict daily dew point temperature. *Applied Thermal Engineering* 112: 1097-1107. <https://doi.org/10.1016/j.applthermaleng.2016.10.181>

- Mehdizadeh S, Mohammadi B, Ahmadi F. 2022. Establishing coupled models for estimating daily dew point temperature using nature-inspired optimization algorithms. *Hydrology* 9: 9. <https://doi.org/10.3390/hydrology9010009>
- Mohammadi K, Shamshirband S, Petković D, Yee PL, Mansor Z. 2016. Using ANFIS for selection of more relevant parameters to predict dew point temperature. *Applied Thermal Engineering* 96: 311-319. <https://doi.org/10.1016/j.applthermaleng.2015.11.081>
- Morales-Nápoles O. 2010. Counting vines. In: Dependence modeling: Vine copula handbook (Kurowicka D, Joe H, Eds.). World Scientific, 189-218. https://doi.org/10.1142/9789814299886_0009
- Nadig K, Potter W, Hoogenboom G, McClendon R. 2013. Comparison of individual and combined ANN models for prediction of air and dew point temperature. *Applied intelligence* 39: 354-366. <https://doi.org/10.1007/s10489-012-0417-1>
- Nash JE, Sutcliffe JV. 1970. River flow forecasting through conceptual models part I – A discussion of principles. *Journal of hydrology* 10: 282-290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Nazeri-Tahroudi M, Ramezani Y. 2020. Estimation of dew point temperature in different climates of Iran using support vector regression. *Időjárás* 124: 521-539. <http://doi.org/10.28974/idojaras.2020.4.6>
- Nazeri-Tahroudi M, Ramezani Y, De Michele C, Mirabbasi R. 2021. Multivariate analysis of rainfall and its deficiency signatures using vine copulas. *International Journal of Climatology* 42: 2005-2018. <https://doi.org/10.1002/joc.7349>
- Nazeri-Tahroudi M, Ramezani Y, De Michele C, Mirabbasi R. 2022. Bivariate Simulation of Potential Evapotranspiration Using Copula-GARCH Model. *Water Resources Management* 36: 1007-1024. <https://doi.org/10.1007/s11269-022-03065-9>
- Raji M, Tahroudi MN, Ye F, Dutta J. 2022. Prediction of heterogeneous Fenton process in treatment of melanoidin-containing wastewater using data-based models. *Journal of Environmental Management* 307: 114518. <https://doi.org/10.1016/j.jenvman.2022.114518>
- Salvadori G, de Michele C. 2004. Frequency analysis via copulas: Theoretical aspects and applications to hydrological events. *Water Resources Research* 40: W12511. <https://doi.org/10.1029/2004WR003133>
- Salvadori G, De Michele C, Kottegoda NT, Rosso R. 2007. Extremes in nature: An approach using copulas. *Water Science and Technology Library*, 56. Springer, Dordrecht.
- Salvadori G, de Michele C. 2007. On the use of copulas in hydrology: Theory and practice. *Journal of Hydrologic Engineering* 12: 369-380. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2007\)12:4\(369\)](https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(369))
- Salvadori G, de Michele C. 2010. Multivariate multi-parameter extreme value models and return periods: A copula approach. *Water Resources Research* 46: W10501. <https://doi.org/10.1029/2009WR009040>
- Salvadori G, de Michele C, Durante F. 2011. Multivariate design via copulas. *Hydrology Earth System Sciences Discussions* 8: 5523-5558. <https://doi.org/10.5194/hessd-8-5523-2011>
- Shank DB, McClendon RW, Paz J, Hoogenboom G. 2008. Ensemble artificial neural networks for prediction of dew point temperature. *Applied Artificial Intelligence* 22: 523-542. <https://doi.org/10.1080/08839510802226785>
- Shiri J, Kim S, Kisi O. 2014. Estimation of daily dew point temperature using genetic programming and neural networks approaches. *Hydrology Research* 45: 165-181. <https://doi.org/10.2166/nh.2013.229>
- Tahroudi MN, Ramezani Y, de Michele C, Mirabbasi R. 2020a. A new method for joint frequency analysis of modified precipitation anomaly percentage and streamflow drought index based on the conditional density of copula functions. *Water Resources Management* 34: 4217-4231. <https://doi.org/10.1007/s11269-020-02666-6>
- Tahroudi MN, Ramezani Y, de Michele C, Mirabbasi R. 2020b. Analyzing the conditional behavior of rainfall deficiency and groundwater level deficiency signatures by using copula functions. *Hydrology Research* 51: 1332-1348. <https://doi.org/10.2166/nh.2020.036>
- Zhang G, Band SS, Ardabili S, Chau KW, Mosavi A. 2022. Integration of neural network and fuzzy logic decision making compared with bilayered neural network in the simulation of daily dew point temperature. *Engineering Applications of Computational Fluid Mechanics* 16: 713-723. <https://doi.org/10.1080/19942060.2022.2043187>
- Zhang L, Singh VP. 2006. Bivariate Flood frequency Analysis Using the Copula Method. *Journal of Hydrologic Engineering* 11: 150-164. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2006\)11:2\(150\)](https://doi.org/10.1061/(ASCE)1084-0699(2006)11:2(150))
- Zounemat-Kermani M. 2012. Hourly predictive Levenberg-Marquardt ANN and multi linear regression models for predicting of dew point temperature. *Meteorology and Atmospheric Physics* 117: 181-192. <https://doi.org/10.1007/s00703-012-0192-x>