# Developing a vine copula model to simulate and predict long serial lake water levels

Shixiang Gu [1, 2, 3], Yanling Wei [2*], Jing Chen [1], Zhong Zhao [4], Rong Gao [1], Jinming Chen [1, 3], Zekun Gao [2], Mengqi He [2], Gang Chen [1], Jing Li [2]

[1]Yunnan Institute for Investigation. Design and Research of Water Resources & Hydropower Engineering, Kunming 650021, China
[2]Key Laboratory of Water Security and Water Saving and Emission Reduction in Yunnan Agricultural University, Kunming 650201, China
[3]State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China
[4]Yunnan Agricultural Machinery Safety Supervision Station, Kunming 650051, China

**Abstract:** Lake water level changes show randomness and the complexity of basin hydrological simulation and lake water level response. We constructed a vine copula model to simulate and predict lake water level that incorporated rolling decisions and real-time correction of prediction results. The model was applied to predict the long- and short-term water levels in Erhai Lake on the Yun-gui Plateau, southwest China. The results showed that (1) the predicted daily water levels (with ME=0.02~0.09, RMSE=0.02~0.024, NSE=0.99, and IA=0.99) were more accurate than the predicted monthly water levels (with the ME=0.039~0.444, RMSE=0.194~0.279, NSE=0.913~0.958, and IA=0.977~0.989), and the accuracy of the predictions improved as the number of variables increased. (2) The vine copula model outperformed the back-propagation neural network and support vector regression models, and, of the three model types, gave the best estimate of the nonlinear relationships between the predicted water level and climatic factors, especially in the wet season (May to October). (3) The prediction accuracy of the vine copula model was lower for small sample sizes and when there was a lack of runoff data. By improving the analysis of the model's errors, the percentages of the relative errors of the prediction accuracy less than 5%, 10%, 15%, and 20% increased to 70%, 83%, 95%, and 98%, respectively.

## 1. Introduction

Global climate change can drive increases in lake water temperature, decreases in the length of the ice cover period, expansion of lake areas, and increases in evaporation from lake surfaces [1,2]. Water level is the most important factor in lake management, as its fluctuations indicate changes in the volume of water in the basin, and determine the carrying capacity and self-regulation thresholds of the water resource system. Water level fluctuations also influence a lake's eco-environmental characteristics, such as the water quality, sediment, and aquatic life. Accurate simulations and predictions of lake water levels are therefore important for supporting artificial regulation, operation, and remediation of lakes, and for supporting decisions about the timing of the implementation of lake functions. The development of high-performance computers, modern intelligent algorithms, and air-space-ground stereoscopic monitoring technology has allowed researchers to simulate and predict lake water levels [3–5]. To date, three approaches have been used to predict lake water levels, namely physically based models, data-driven and artificial intelligence models, and optical remote-sensing monitoring [6–36].

Physically based models combine a generalized model of the hydrological water resources, hydrodynamic forces, hydrothermal conditions, and other elements in a lake basin; data about the climate patterns, ocean circulation, and geology of the environment, and supporting algorithmic models, such as the copula model and the support vector regression (SVR) model. While physically based models produce highly accurate simulations and predictions, they require detailed data of variables such as lake topography, hydrometeorology, socioeconomics, and the model parameters, which can be difficult to acquire. Furthermore, these models must be able to describe the complexities of the natural and social aspects of the water cycle by finely simulating the water resources supply, usage, consumption, drainage, and the change from precipitation to runoff, meaning that they demand substantial computing time. The complexity of physically based models, their data requirements, and computing demands, therefore, tend to limit their wide application [37]. Data-driven models, also known as artificial intelligence and optimization computational models, are based on the relationships between long-term synchronous observation data of lake water levels and the drivers of change, namely the hydro-meteorological factors. Using the autocorrelation characteristics of a lake's long-term water level data, one or more hydro-

*swling503@qq.com

meteorological factors that are closely related to the water level changes are overlain to build a model that simulates and predicts the water levels. Data-driven models mainly include linear regression models [18, 29], adaptive-network based fuzzy inference systems (ANFIS) [38], artificial neural network (ANN) models [39], SVR models [31], and wavelet analysis [40], with each having its limitations. Hossein et al. [19] used two hydraulically conventional models (WBE and MRM), two linear models (ARX and BJ), and two nonlinear intelligent models (MLP neural network and LLNF) to simulate and predict the water levels in Urmia Lake, and found that the nonlinear intelligent models outperformed the other models. Meral et al. [41] predicted the monthly water levels in Beysehir Lake with five methods, PSO-ANN, SVR, MLP, RBNN, and found that the SVR model had the highest prediction accuracy. Elsewhere, statistical machine learning and ARIMA models were combined to predict water levels in the Red River Delta over different time periods [20]. The third method is based on optical remote sensing and uses high-definition cameras or radars mounted on various flight platforms, such as aerospace satellites, spacecraft, and unmanned aerial vehicles. The information collected from the cameras or radars is verified with in-situ observations, and air-space-ground integrated systems may be used to support large-scale, long-term continuous monitoring of lake water levels [33–36].

Changes in lake water level are the combined result of many natural and social factors, such as precipitation, evaporation, runoff, temperature, wind speed, water resource development and utilization, and lake management regimes in a watershed [14]. The relationships between these factors are complex and nonlinear, with the result that it is difficult to predict how the water level will respond to changes in hydro-meteorological variables using simple statistical models. In recent years, copula functions have been increasingly used in multivariate statistics and stochastic simulations, because of their ability to describe the interdependence of variables. Copula functions have been used extensively in hydrology and water resources research for analyzing flood frequency, identifying drought return periods, geostatistical interpolation, simulating precipitation, simulating runoff from multiple locations, and predicting droughts [42–48]. Most studies so far have used two- or three-dimensional copulas as high-dimensional copulas do not provide accurate simulations of the correlation between high-dimensional variables [49]. Vine copulas, however, are known for their ability to decompose high-dimensional joint distributions into a hierarchical structure of binary connections, and are more flexible than multivariate copulas for constructing complex dependency structures between variables. While vine copulas have been used successfully to predict droughts and simulate flood characteristics, precipitation, and runoff [50–55], they have rarely been used to simulate and predict lake water levels over the long-term time-series.

The objective of this study was to capture the dependence between the lake water level and hydro-meteorological variables, and explore the possibility of using a vine copula model for predicting the long-term water level. To achieve this objective, a multidimensional variable vine copula model was constructed and the monthly and daily water levels were predicted from different combinations of the hydro-meteorological factors. The accuracy of the predictions from the constructed vine copula model was tested by comparing the simulation results with those from a BP neural network model and a support vector regression (SVR) model. A vine copula model for lake management was established to predict the lake water level for small data sets. The accuracy of the model was tested and improved by error analysis, such that the obtained lake water level values were further refined. The vine copula model produced in this study will provide information to support decisions about lake water level regulation, protect the aquatic ecological environment, manage the water resources, and plan the allocation of water resources in a basin.

## 2. Materials and Methods

### 2.1. Study area and data

Erhai Lake, in the Dali Bai Autonomous Prefecture of Yunnan Province, is the seventh largest freshwater lake in China and the second largest freshwater lake on the Yungui Plateau . It is at the heart of the Cangshan Mountain-Erhai Lake National Nature Reserve [56]. This lake performs seven major functions, namely municipal water supply, irrigation, hydropower generation, climate regulation, fishery, shipping, and tourism. The Erhai Lake area encountered continuous drought disasters from 2010 to 2015 because of global climate change and high-intensity human activities. For example, the agricultural water use and sources of agricultural non-point pollutants in the basin increased, causing a decline in the lake water level, and triggering local cyanobacterial bloom outbreaks. Because of the lake's value as a multi-purpose resource, the protection and management of the Erhai Lake Basin aquatic ecosystem has attracted national attention.

The study was supported by data of the (1) mean monthly (or daily) water level (1954–2021), obtained from the Daguanyi Station in Erhai Lake Basin; (2) monthly (or daily) precipitation, mean temperature, and evaporation data (1954–2021) retrieved from Dali Meteorological Station; (3) monthly runoff for the Erhai Lake Basin (1954–2016) that was calculated from measured water level data from the Daguanyi Station at Erhai Lake, outflow rate data from the Tianshengqiao Hydrologic Station on the Xi'er River downstream and the tunnel of Trans- basin water supply engineering from Erhai Lake to Binchuan, and observed evaporation data from Dali Meteorological Sta-tion, and data for industrial and agricultural water use in the Erhai Lake Basin. We al-so used data from (4) the Water Resources Bulletins and Water Conservancy Censuses of the Yunnan Province.

### 2.2. Methods

The aim of the study was to predict the lake water level under a range of scenarios with different combinations of hydro-meteorological variables, with evaporation (E), temperature (T), precipitation (P), and the runoff flowrate (F) as the hydro-meteorological variables. Because there is significant autocorrelation in the lake water level series over time, the water level state for an earlier time period was used as an input variable to the model. When the prediction process involved the previous water level, the water level measured at the end of the previous time period was used to correct the water level predicted at the beginning of the time period. Then, the measured water level of the previous period was used as an input when predicting the level of the next time period. This approach, of combining rolling decisions with real-time corrections, helped prevent the accumulation of model errors in the prediction process [57]. The idea behind the model was to select different combinations of conditional variables, use the most suitable vine copula to connect the function between the water level and the hydro-meteorological variables, simulate the predicted water level through the quantile conditional form of the function, and then select the optimal prediction model.

Lake water level changes are the result of the interactions of multiple factors that change continuously. The relationships between the factors that affect the lake water level are variable. Various vine copula structures have been used to deal with the complex relationships between variables. In this study, two vine copula structures were selected and compared, namely the C-vine and the D-vine copula functions. The C-vine is a star-like structure that describes a situation where one variable dominates the remaining variables, and the D-vine is a parallel line structure that describes a situation when the dependencies of the variables are almost the same [58]. The C-vine copula and D-vine copula structures are displayed using five variables, namely, the water level of the current time period (Z), water level of the previous time period ($Z_{t-1}$), evaporation (E), temperature (T), and the runoff flowrate (F) (Fig. 1).

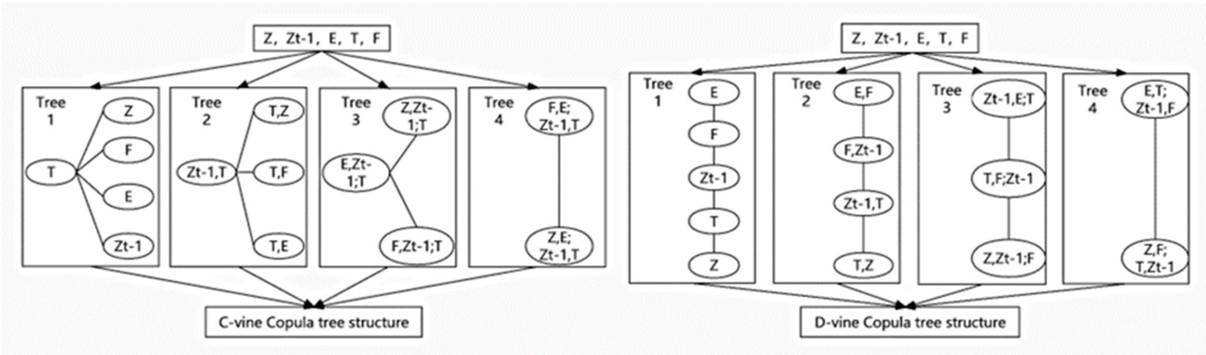



**Figure 1**. Five-dimensional variable structure of the vine copula

For example, when the five variables shown in Fig. 1 are arranged in different orders, the combined probability density function of the C- and D-vine copulas can be decomposed into:

$$f(T,Z,F,E,Z_{t-1}) = f(T)\cdot f(Z)\cdot f(F)\cdot f(E)\cdot f(Z_{t-1})\cdot c_{TZ}\cdot c_{TF}\cdot c_{TE}\cdot c_{TZt-1}\cdot c_{Zt-1,Z|T}\cdot c_{Zt-1,F|T}\cdot c_{Zt-1,E|T}\cdot c_{E,Z|Zt-1,T}\cdot c_{E,F|Zt-1,T}\cdot c_{Z,F|E,Zt-1,T} \quad (1)$$

$$f(E,F,Z_{t-1},T,Z) = f(E)\cdot f(F)\cdot f(Z_{t-1})\cdot f(T)\cdot f(Z)\cdot c_{EF}\cdot c_{FZt-1}\cdot c_{Zt-1T}\cdot c_{TZ}\cdot c_{E,Zt-1|F}\cdot c_{F,T|Zt-1}\cdot c_{Zt-1,Z|T}\cdot c_{E,F|T,Zt-1}\cdot c_{T,Z|Zt-1,F}\cdot c_{E,Z|T,Zt-1,F} \quad (2)$$

where $f$(X) is the edge density function; $c$ is the density function of the two-dimensional copula in the vine structure, and $c_{\cdot|\cdot}$ is a conditional distribution function. With a three-dimensional variable as an example, $f(Z|Z_{t-1},E)$ can be expressed by equation (3) when the $h$ function is introduced into the conditional distribution function:

$$f(Z|Z_{t-1}E) = \frac{\partial c_{Z,Z_{t-1}|E}[f(Z|Z_{t-1}),G(E|Z_{t-1})]}{\partial f(E|Z_{t-1})} = h\{h(u_Z|u_{Zt-1};\theta_{Zt-1,E})|h(u_E|u_{Zt-1};\theta_{Zt-1,Zt-1});\theta_{E,Zt-1}\} \quad (3)$$

where θ is the parameter of the copula function when there is a joint distribution between two variables, and u represents the edge cumulative distribution function of the variable, which represents the conditional distribution function of water level (Z) given the conditional variable water level of the previous time period ($Z_{t-1}$) and evaporation (E). The quantile transformation is then carried out to derive the inverse function to predict the water level(Z), which can be expressed as:

$$Z = F^{-1}(u_Z) = F^{-1}[h^{-1}\{h^{-1}(\Gamma|h(u_E|u_{Zt-1};\theta_{Zt-1,Zt-1});\theta_{E,Zt-1})|u_{Zt-1};\theta_{Zt-1,E}\}] \quad (4)$$

where $F^{-1}$ is the inverse of $u_{z;}$ and $h^{-1}$ is the inverse of the $h$ function, and Γ is the probability.

Building a vine copula model for simulating and predicting water levels involved the following steps:

1) The marginal distribution for the variable was fitted.

2) Different conditional variable combination scenarios were selected to construct three-, four-, five-, and six-dimensional C- and D-vine copula joint distributions. After the vine copula model was completed, the appropriate vine copula model was selected for the different combinations of conditional variables.

3) The distribution of the conditions was constructed, to account for the long-term impact of meteorological factors and water level in the early stage.

4) Vine copula sampling has inherent uncertainty. The sampling algorithms were improved and used to sample from the conditional C- or D-vines (see [59]). To obtain better predictions, 1000 samplings were carried out in the conditional C- and D-vines. Using Equation (4), the average of the 1000 predicted water level values was calculated as the final predicted water level value.

5) The above procedures were repeated until all the conditional variable combina-tions were processed. The vine copula model was then evaluated for each combination of the conditional variables to select the optimal conditional variable combination. The technical framework is presented in Fig. 2.

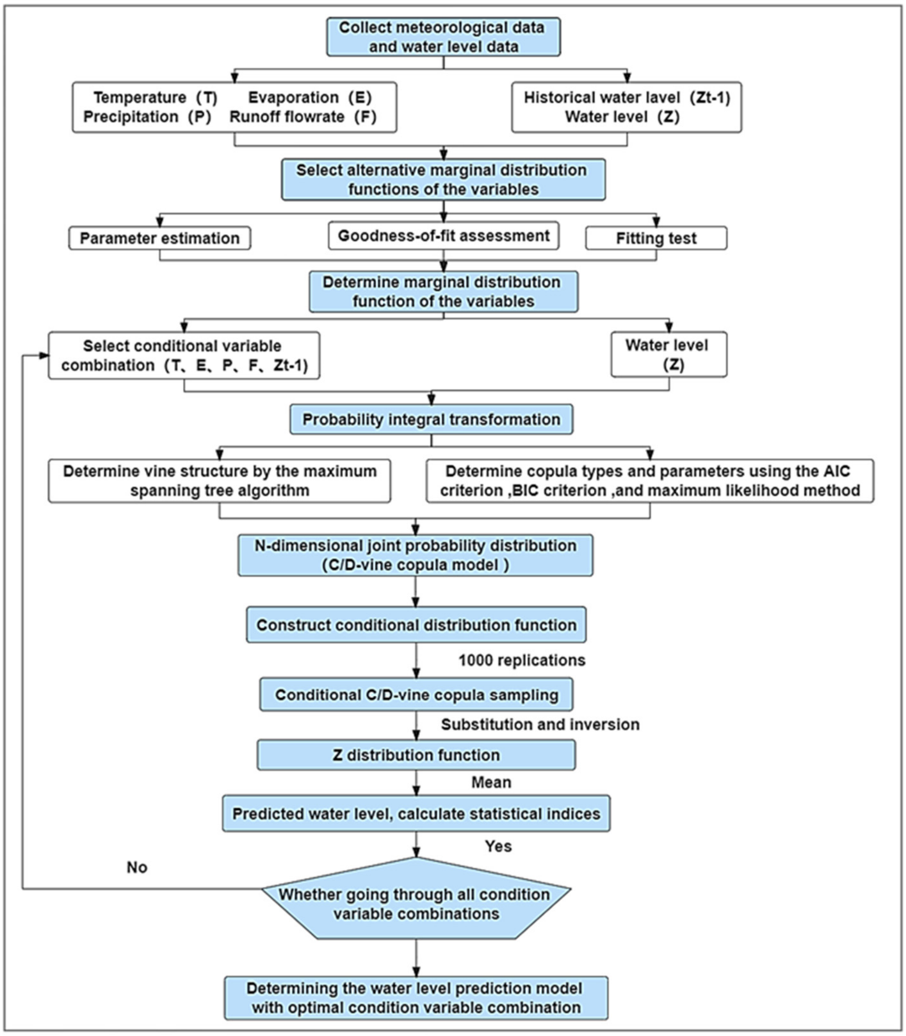



**Figure 2.** Framework for predicting the lake water level based on the vine copula

## 2.3. Model error evaluation

The vine copula simulation method was compared with two other models, a BPNN and an SVR, to evaluate the accuracy of the forecasted lake water levels. The model's ability to predict the water level of Erhai Lake was evaluated using four common statistical indices of error, namely, the mean error (ME), root mean square error (RMSE), index of agreement (IA), and Nash-Sutcliffe efficiency coefficient (NSE), that were calculated with formulae (5)–(8). The IA values are between 0–1, and the model performance improves as the IA value increases. The NSE values range from –∞ to 1; an NSE value close to 1 indicates that the simulated value is close to the observed value, whereas an NSE value close to 0 indicates that the simulated result is close to the mean level of the observed value, that is, the overall result is reliable, but there are large errors in the process simulation range [11].

$$ME = \sum_{i=1}^{n} \frac{S_i - O_i}{n} \tag{5}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(S_i - O_i)^2}{n}} \tag{6}$$

$$IA = 1 - \frac{\sum_{i=1}^{n}(S_i - O_i)^2}{\sum_{i=1}^{n}(|S_i - \bar{O}| + |O_i - \bar{O}|)^2} \tag{7}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(S_i - O_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \tag{8}$$

where $n$ is the number of months/days, $S_i$ is the predicted water level, $O_i$ is the actual observed water level, and $\bar{O}$ is the mean of the actual observed water level.

# 3. Results and Analysis

## 3.1. Building vine copula models to simulated long-term changes in the Erhai Lake water leve

The P-III distribution was selected for the monthly runoff flowrate (F), and distributions typically used for hydrological variables (Normal, Gamma, Lognormal, and Weibull) were selected to construct the marginal distribution function of the other variables. The Weibull distribution was selected for the water level (Z), mean water level of the previous month (Zt-1), monthly precipitation (P), and the mean monthly temperature (T), and the lognormal distribution was selected for monthly evaporation (E) (Table 1).

**Table 1.** Goodness-of-fit test of the univariate marginal distribution functions

| Distribution function | Test parameters | Z | $Z_{t-1}$ | P | T | E |
|---|---|---|---|---|---|---|
| Normal distribution | P | 0.1072 | 0.1065 | 0.158 | 0.127 | 0.107 |
| | AIC | 2071.852 | 2073.708 | 8687.608 | 4359.168 | 7437.763 |
| | BIC | 2081.108 | 2082.964 | 8696.808 | 4368.424 | 7447.019 |
| Gamma distribution | P | 0.1073 | 0.1066 | 0.101 | 0.146 | 0.065 |
| | AIC | 2071.995 | 2073.851 | 8094.378 | 4414.317 | 7338.605 |
| | BIC | 2081.251 | 2083.107 | 8103.578 | 4423.573 | 7347.861 |
| Lognormal distribution | P | 0.1073 | 0.1067 | 0.212 | 0.158 | 0.051 |
| | AIC | 2072.067 | 2073.922 | 9453.176 | 4480.846 | 7314.036 |
| | BIC | 2081.323 | 2083.179 | 9462.376 | 4490.102 | 7323.292 |
| Weibull distribution | P | 0.046 | 0.046 | 0.072 | 0.123 | 0.104 |
| | AIC | 2013.296 | 2015.317 | 8055.417 | 4319.576 | 7447.411 |
| | BIC | 2022.552 | 2024.574 | 8064.616 | 4328.832 | 7456.667 |

Three-, four-, five-, and six-dimensional vine copula models were built with dif-ferent combinations of water level (Z) and the other five variables (abbreviated as VC(x1, ......, xn), n=3, 4, 5, 6). The three-dimensional C-vine copula and the D-vine copula had the same goodness-of-fit (Table 2), but the error analysis of the fitting results showed that the C-vine copula model should be selected. The C-vine copula model was also chosen for the VC(Z,Zt-1,E,P) four-dimensional vine copula models, VC(Z,Zt-1,E,T,F) five-dimensional vine copula models, and the VC(Z,Zt-1,E,T,P,F) six-dimensional vine copula model, and the D-vine copula model was chosen for other models. The C-vine copula was there-fore chosen for most of the different variable combinations used to predict the monthly water levels in Erhai Lake. This analysis indicates that the correlations between the hydro-meteorological variables and the water level were variable. The C-Vine copula model had a better ability to capture the dependent structures between the hydro-meteorological variables and the water level.

**Table 2**. Goodness-of-fit testing of the Multi- dimension vine copula models

| Simulated Model | C-vine copula function | | | D-vine copula function | | |
|---|---|---|---|---|---|---|
| | logLik | AIC | BIC | logLik | AIC | BIC |
| $VC(Z,Z_{t-1},P)$ | 1112.87 | -2217.74 | -2199.23 | 1112.87 | -2217.74 | -2199.23 |
| $VC(Z,Z_{t-1},T)$ | 992.62 | -1977.23 | -1958.72 | 992.62 | -1977.23 | -1958.72 |
| $VC(Z,Z_{t-1},F)$ | 1026.63 | -2043.27 | -2020.13 | 1026.63 | -2043.27 | -2020.13 |
| $VC(Z,Z_{t-1},E)$ | 1032.72 | -2055.45 | -2032.31 | 1032.72 | -2055.45 | -2032.31 |
| $VC(Z,Z_{t-1},E,P)$ | 1296.17 | -2576.34 | -2539.32 | 1286.2 | -2554.41 | -2512.75 |
| $VC(Z,Z_{t-1},E,T)$ | 1221.13 | -2424.26 | -2382.61 | 1227.49 | -2436.99 | -2395.34 |
| $VC(Z,Z_{t-1},E,F)$ | 1183.58 | -2351.17 | -2314.14 | 1185.29 | -2352.57 | -2310.92 |
| $VC(Z,Z_{t-1},E,T,P)$ | 1484.18 | -2942.35 | -2882.19 | 1489.24 | -2950.49 | -2885.7 |
| $VC(Z,Z_{t-1},E,T,F)$ | 2495.35 | -4966.71 | -4911.17 | 2470.53 | -4909.07 | -4835.02 |
| $VC(Z,Z_{t-1},E,T,P,F)$ | 1749.23 | -3456.45 | -3359.26 | 1731.47 | -3420.93 | -3323.74 |

## 3.2. Prediction results and analysis

### *3.2.1. Selection of the optimal variable combination*

To analyze how the different combinations and numbers of variables influenced the predictions of the lake water level by the vine copula model, the measured values were compared with the predicted values, and the error indices were calculated. The ME, RMSE, NSE, and IA index values ranged from 0.039–0.444, 0.194–0.279, 0.913–0.958, and 0.977–0.989, respectively. The most accurate prediction was achieved with a three-dimensional vine copula model with the $VC(Z,Z_{t-1},E)$ combination (Fig. 3). In other cases, the predictions of the three-dimensional vine copula models built with the lake water level (Z), evaporation (E), precipitation (P), temperature (T), and runoff flowrate (F) variables all produced large errors. The NSE values were very small or close to 0, which indicates that a vine copula model for predicting water levels would not be reliable if built with only the meteorological factors that influenced the water level changes. The results for artificially regulating the lake water level were mainly for the previous period. Stochastic analysis showed that the lake water level was highly autocorrelated with the water level of the previous period, so it was added to the model as an input factor to investigate whether it might improve the simulation accuracy.

To explore the dependency between the lake water level and the other variables, four-, five-, and six-dimensional vine copula models were built continuously using the combination of variables that showed the best dependency in each dimension together with other variable combinations. The prediction accuracies were best with the $VC(Z,Z_{t-1},E,T)$, $VC(Z,Z_{t-1},E,T,P)$, and $VC(Z,Z_{t-1},E,T,P,F)$

combinations, and the prediction accuracy tended to increase as the number of variables increased. The prediction accuracy was highest for the VC($Z,Z_{t-1},E,T,P,F$) model, and the percentages of samples with relative errors less than 5%, 10%, and 15% were 81%, 98%, and 99%, respectively (for relative error statistics, the difference between the maximum and minimum water levels were used as the true values because of the high elevations at Erhai Lake). Because of the advantageous structure of the vine copula, inputs of different variable factors can give different absorption results for the variable factors. It is better to use strongly correlated variables, to maximize the accuracy of the predictions and to characterize the structure of the variable dependency accurately. The analysis results showed that the long-term interannual water level changes in Erhai Lake were correlated with the temperature, evaporation, precipitation, and the runoff entering the lake, and that the water levels were more strongly correlated with temperature and evaporation than with the runoff and precipitation.

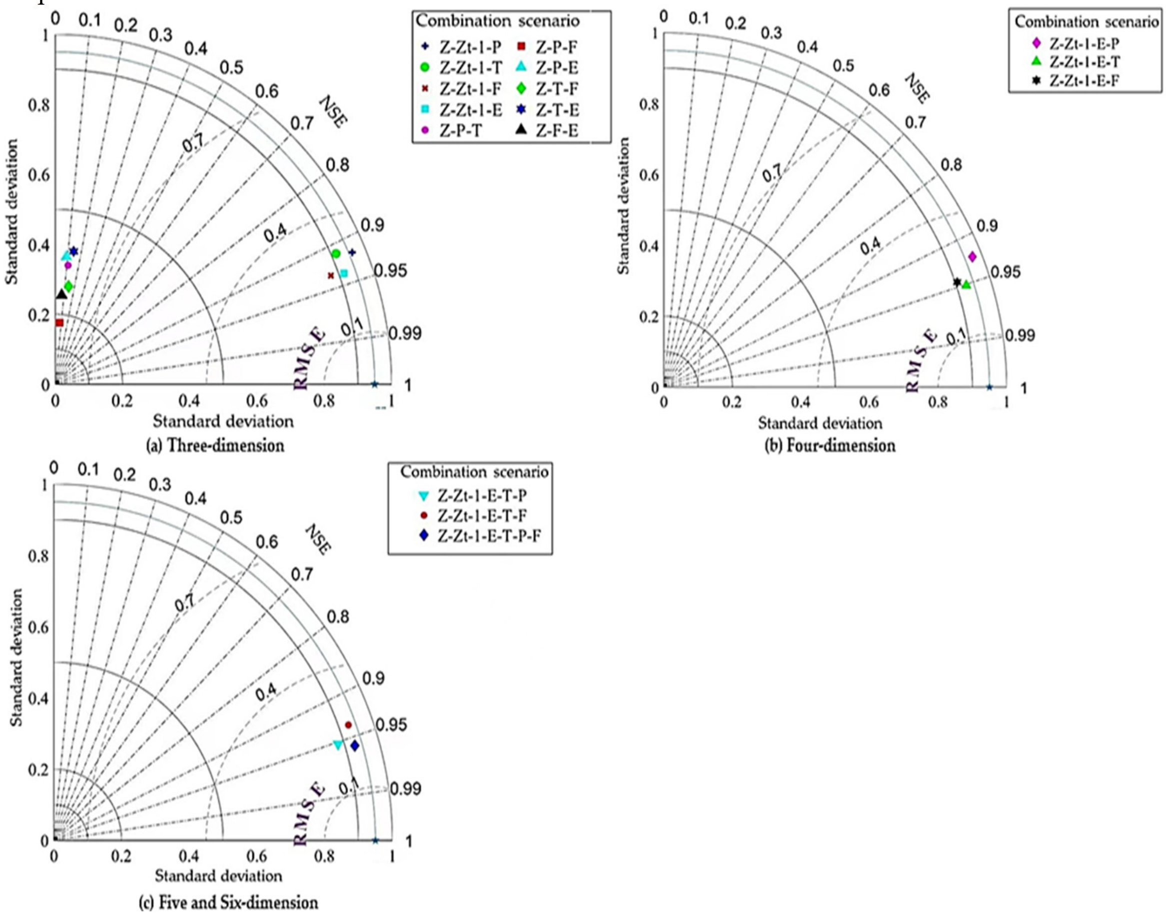



**Figure 3.** Taylor plot of the measured values by different vine copula models.

### *3.2.2. Prediction results*

The vine copula models of the three-, four-, five-, and six-dimensional variables with the best dependencies were used to predict the water level of Erhai Lake from 1954–2016. The results showed that the predicted time series and the observed time series almost overlapped. Over the years, the predicted water level of Erhai Lake tended to fall, rise, and then fall. The mean annual minimum (1964.48 m altitude) and maximum (1965.69 m altitude) water levels occurred in June and November respectively, and were consistent with the mean annual minimum (1964.39 m altitude) and maximum (1965.64 m altitude) observed values. These values indicate that the models gave accurate predictions of the inter-annual characteristics of the water level, and closely reflected the water level trends in Erhai Lake. However, the prediction results for some of the peaks were not ideal and deviated somewhat from the observed time series. It may be that, when predicting high and low water levels with a vine copula model, the marginal distribution of the variables cannot represent the variable extreme values accurately, which then affects the ability of the vine copula model to simulate the joint distribution structure of different variables.

## 3.3. Model pridected accuracy comparison

The results of the vine copula models developed in this study were compared with the results of the SVR and BPNN models using the same variables. For this testing, the inputs to the SVR and BPNN models were the conditional variables of the combination that achieved the highest prediction accuracy in vine copula models, and the output was the water level. Comparison of the results showed that the lake water levels predicted by the BPNN model deviated considerably from the measured values (correlation coefficient R = 0.96), whereas the water levels predicted by the vine copula and SVR models were relatively close to the measured values (correlation coefficient R =

0.98; Fig. 4). The vine copula and SVR models performed well in the dry season (from November to April of the next year), but the vine copula model gave better predictions in the wet season (May–October) than the other two model types. As well as producing graphs of the predicted and observed values, the error indices of the different models were calculated and compared. The vine copula model produced relatively low ME and RMSE values and high IA and NSE values (Table 3). It also achieved the highest percentage of relative errors less than 5%, showing that it was more accurate than the other two models. This is because the vine copula model can deal with variable factors and break down high-dimensional distributions into numerous two-dimensional distributions and fully absorb relevant information of the variables. Further, it has a higher capacity to capture the nonlinear relationships between the predicted water level and climatic factors than the other model types. The BPNN model is highly dependent on the data and the accuracy of the predictions may be influenced by irrelevant observed points, resulting in an overall low model accuracy. Many of the values predicted by the BPNN model deviated considerably from the measured values, as shown in Fig. 4.

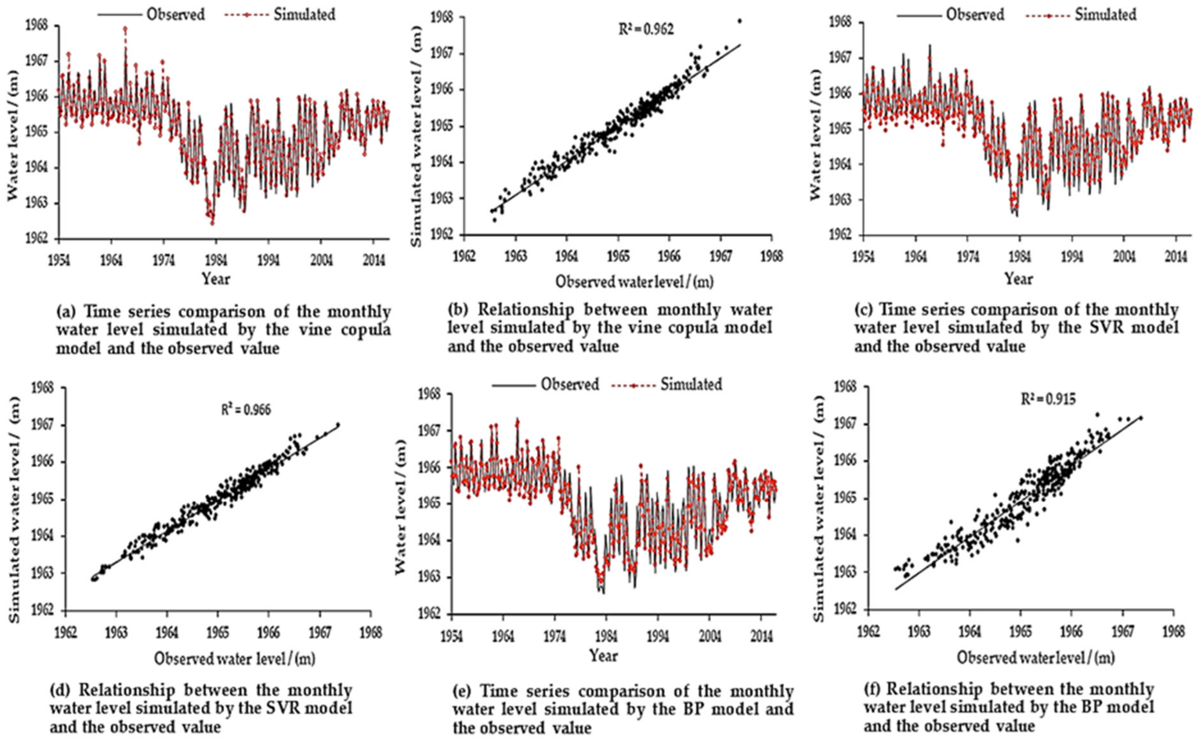



**Figure 4**. Comparison of the simulated monthly water levels from the different models and the observed values

**Table 3**. Comparison of the prediction performance of the different models

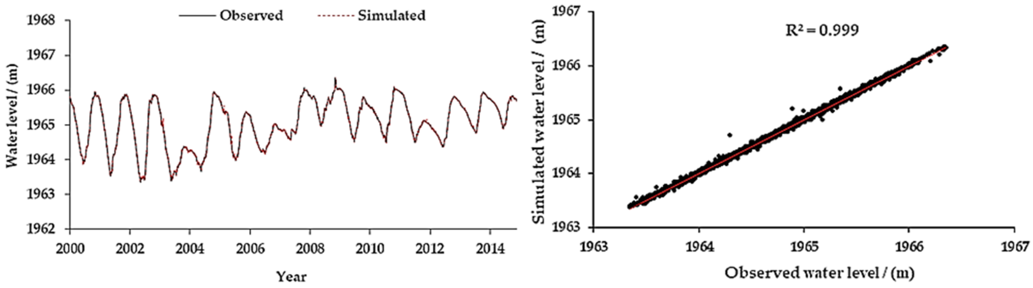
| Model type | Statistical index | | | | Relative error statistics | | | |
|---|---|---|---|---|---|---|---|---|
| | ME | RMSE | IA | NSE | 5% | 10% | 15% | 20% |
| C-Vine copula | 0.048 | 0.194 | 0.989 | 0.958 | 0.81 | 0.98 | 0.99 | 1 |
| SVR | 0.020 | 0.212 | 0.985 | 0.950 | 0.72 | 0.99 | 1 | 1 |
| BP | 0.057 | 0.294 | 0.976 | 0.904 | 0.57 | 0.91 | 0.99 | 1 |

## 4. Method application

### 4.1. Building vine copula models for predicting daily water level in Erhai Lake

The vine copula models with different combinations of conditional variables were applied to simulate and predict the daily water level for 2000–2014 using the meteorological dataset (including E, T, P) and the verified daily water level data of Erhai Lake (2000–2014) for the same period. The error indices of the models were calculated and compared. The ME values ranged from 0.02–0.09, the RMSE values ranged from 0.02–0.024, and the NSE and IA values both reached 0.99. These values show that the accuracy of the predicted daily water level was considerably higher than that of the monthly water level, and the higher accuracy reflects the larger data set that was used in the prediction (5479 samples). The daily water level was highly correlated with the water level of the previous day, and the Kendall correlation coefficients reached 0.99. Therefore, when the daily water level was predicted with different combinations of meteorological variables, the effect on the prediction results was minimal. The predicted time series of the daily water level obtained with the $VC(Z,Z_{t-1},E,T,P)$ achieved the highest prediction accuracy, and essentially overlapped with the measured time series

(Fig. 5). The predicted values and observed values of the daily water level were strongly correlated, with a correlation coefficient (R) of 0.99, which shows that the prediction was highly accurate.




**(a) Time series comparison of the daily water level simulated by the vine copula model and the observed value**
**(b) Relationship between daily water level simulated by the vine copula model and the observed value**

**Figure 5**. Simulated and observed value of the daily water level from vine copula mode

## 4.2. Rapid prediction output of lake water level at monthly scale

There was no systematic analysis of runoff in Erhai Lake and its surrounding areas after 2016. Therefore, the water levels for 2017–2021 were predicted from the monthly meteorological data (including E, T, P), and different combinations of conditional variables were substituted into the vine copula model. The ME, RMSE, NSE, and IA values ranged from 0.009–0.048, 0.122–0.198, 0.706–0.888, and 0.905–0.968, respectively. The prediction accuracy was lower than that for 1954–2016. The information about the variable factors could not be fully extracted because of a lack of runoff data and the small amount of sample data. However, the prediction accuracy was always highest for the combination with the highest number of variables, namely, the vine copula model with $VC(Z, Z_{t-1}, E, T, P)$. For this combination, the percentages of relative errors less than 5%, 10%, 15%, and 20% were 42%, 73%, 87%, and 95%, respectively. All the vine copula models with different variable combinations produced smaller errors for November–April than for May–October. The trends of the errors were basically the same, and displayed an M-shaped pattern. There were two peaks in June, July, and September, which was similar to the pattern of systematic errors calculated when predicting the daily reference crop evapotranspiration [47]. The mean error of the predicted values and the measured values corresponding to each month from 1954 to 2016 was used as a correction term and was substituted into the vine copula model with the highest prediction accuracy. Using this process, the percentages of relative errors less than 5%, 10%, 15%, and 20% increased to 70%, 83%, 95%, and 98%, respectively. The monthly water level changes in Erhai Lake are shown in Fig. 6. The results for the monthly observed water level and predicted values were still unsatisfactory for a few specific time periods, with deviations mainly in March–July 2019 and July–October 2021. It is possible that the precipitation and evaporation were distinctly different and showed sudden increases or decreases between these months and the adjacent months, which influenced the description of the data characteristics and reduced the accuracy of the predictions from the vine copula models. The monthly scale vine copula model for the short-term water level was tested for stationarity, and the results showed that this process was stationary and random. The construction of the vine copula model was more straightforward when the time indicator was omitted. The long term water level series predicted by the monthly scale vine copula model was non-stationary and the prediction accuracy was good, even with no processing of the data.

The minimum and suitable ecological water levels of Erhai Lake were determined from the water level guarantee rate [60]. These were overlain on the graph of the predicted monthly water level to show and compare the relationship between the rises and falls of the water level and the regulation and early warning for the lake ecological water quantity. The maximum and minimum monthly operating water levels of Erhai Lake are determined from lake regulation and water supply, flood protection, and water resources allocation planning [61]. It was selected as the monthly minimum water level control threshold of a given time period that the higher value between the minimum operating water level and the minimum ecological water level. For most of 2017–2021, the measured and predicted water levels of Erhai Lake were between the suitable ecological water level and the minimum water level (Fig. 6). They were only below the minimum ecological water level from November 2019 to January 2020, possibly because of the low precipitation during this period. Thus, to ensure the water level remains above the minimum water level for operation, the water storage should be increased and the lake outflow rate should be reduced. It may have been appropriate to increase the water supply to the Binchuan Basin from September 2020 to July 2021 to ensure that the lake water level was maintained between the minimum and suitable ecological water levels for as long as possible and to maximize the benefit from the water supply.

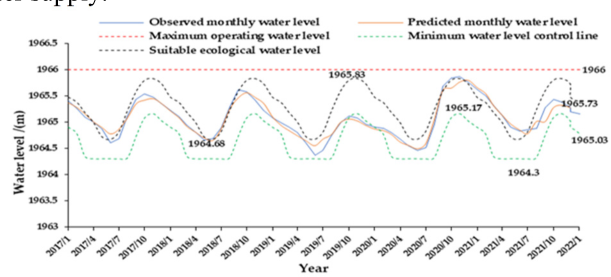



**Figure 6.** Water level changes in Erhai Lake from 2017 to 2021

## 5. Conclusions

In this study, C- and D-vine copula models were introduced, based on the correlation differences between variables. Three-, four-, five-, and six-dimensional joint distribution models were built with different combinations of hydro-meteorological variables (evaporation, temperature, precipitation, and runoff flowrate) and water level (water level of the current time period and the previous period) to predict the long-term time series of the water level in Erhai Lake. Errors in the prediction models were reduced by making rolling decisions and correcting the water level in real-time. The selection of the variable combination and the vine copula structure affected the accuracy of the predictions of the lake water level. The C-vine copula was used more frequently than the D-vine copula, as it had a stronger ability to capture the dependent structure between the hydro-meteorological variables and the water level, and the prediction accuracy was improved by introducing strongly correlated variables. The models were then applied to predict the daily and monthly water levels for the period with no runoff data, and the results were compared. The prediction accuracy was always highest for the model combination with the most variables, and the prediction of the daily water level was more accurate than that of the monthly water level. In addition, the autocorrelation was stronger for the daily water level time series than for the monthly water level time series. For the monthly water levels predicted by the vine copula model, the ME, RMSE, NSE, and IA values ranged from 0.039–0.444, 0.194–0.279, 0.913–0.958, and 0.977–0.989, respectively. The ME and RMSE values for the daily water level in the lake were between 0.02–0.09 and 0.02–0.024, respectively, while the NSE and IA values reached 0.99. Accordingly, the vine copula model was more accurate for predicting the water levels over the short-term. When water levels are predicted without runoff data, the results should be corrected by inversely substituting systematic errors into the vine copula model with the highest prediction accuracy. With this approach, the percentages of the relative errors in the prediction accuracy less than 5%, 10%, 15%, and 20% reached 70%, 83%, 95%, and 98%, respectively. It is worth noting that the vine copula model was limited by the sample size and the prediction accuracy decreased as the number of samples decreased. As far as possible, long-term time-series data should be used when applying the vine copula model.

Comparison of the monthly water levels predicted by the vine copula, BPNN, and SVR models showed that the vine copula model had the best performance. The vine copula and the SVR models performed well in the dry season (November–April of the next year), whereas the vine copula model gave the best performance in the wet season (May–October). The vine copula model is better at dealing with the nonlinear relationships between the predicted water level and climatic factors than other model types, and can also indicate the trends in water level and the interannual variations in the water level. However, there were deviations in some peaks, possibly because the marginal distribution of the variables poorly represented the extreme values of the variables when the vine copula was predicting low- or high-water levels. These promising results show that it is worth evaluating the advantages and disadvantages of the vine copula simulation method in future studies.

## Acknowledgments

This research was funded by National Key R&D Program for the 14th Five-Year Plan (No. 2021YFC3000205-06), demonstration project of comprehensive government management and large -scale industrial application of the major special project of CHEOS (grant number 89-Y50G31-9001-22/23-05), and the Scientific Research and Technical Innovation Team Construction of Yunnan Province (No. 2018HC024).

## References

1. Woolway, R. I.; Kraemer, B. M.; Lenters, J. D.; et al. (2020) Global lake responses to climate change. J. Nature Reviews: Earth & Environment., 13 (3686): 1–16.
2. Zhao, G.; Li, Y.; Zhou, L. M.; et al. (2022) Evaporative water loss of 1.42 million global lakes. J. Nature Communications., 1: 388–403.
3. Jiang, T.; Zhang, Y. J.; Bao, Z. F.; et al. (2018) Water level control for quality of lakes: A case study of Cihu Lake. J. Engineering Journal of Wuhan University., 51 (7): 570-576.
4. He, K. D.; Gao, W.; Duan, C. Q.; et al. (2019) Water level variation and its driving factors in lake Dianchi, Fuxian and Yangzong during 1988-2015. J. Journal of Lake Sciences., 31 (05): 1379-1390.
5. Liu, X. D.; Hou, Z. Y.; Xie, Y. H.; et al. (2021) Influence of water level on four typical submerged plants in wetlands of Lake Dongting. J. Journal of Lake Sciences., 33 (01): 181-191.
6. VanDeWeghe, A.; Lin, V.; Jayaram, J.; et al. (2022) Changes in Large Lake water level dynamics in response to climate change. J. Frontiers in water., 4(4): 1-12.
7. Tyler, J. J.; Marshall, J. C.; Schulz, C.; et al. (2022) Hydrological and Isotopic Variability of Perched Wetlands on North Stradbroke Island (Minjerribah), Australia: Implications for Understanding the Effects of Past and Future Climate Change. J. Frontiers in Environmental Science., 10 (7): 1-21.
8. Tetsuya, T.; Ralph, D. L. (2019) Modeling of Seasonal Lake Level Fluctuations of Titan's Seas/Lakes. J. Journal of Geophysical Research: Planets., 124(2): 617-635.
9. Jayanthi, S. L. S. V.; Keesara, V. R.; Sridhar, V. (2022) Prediction of Future Lake Water Availability Using SWAT and Support Vector Regression. J. Sustainability., 14 (12): 6974: 1-17.
10. Yang, Z. H.; Zhu, Z. T.; Huai, W. X.; et al. (2018) Study on the influence of Poyang Lake Hydraulic project on hydrodynamics and water-quality in wet and

dry year. J. Journal of Hydraulic Engineering., 49(2): 156-167.

11. Li, J. L.; Li, H. Y.; Luo, L. C.; et al. (2022) Water level retrieval for the past and prediction for the next 30 years at Lake Fuxian. J. Journal of Lake Sciences., 34 (3): 958–971.
12. Du, Y.; Xie, M. W.; Hu, M.; et al. (2014) Construction of Nam Co Lake water level variation model and its driving mechanism identification. J. Yangtze River., 45 (5): 19–23.
13. Dong, C. Y.; Wang, N. A.; Li, Z. L.; et al. (2009) Forecast of lake level in Lake Qinghai based on energy-water balance model. J. Journal of Lake Sciences., 21 (4): 587–593.
14. Li, Y.; Zhang, Y. X.; Zhang, X. Z.; et al. (2020) A continuous simulation of Holocene effective moisture change represented by variability of virtual lake level in East and Central Asia. J. Scientia Sinica(Terrae)., 50 (8): 1106-1121.
15. Yasuda, H.; Fenta, A. A.; Berihun, M. L.; et al.(2022) Water level change of Lake Tana, source of the Blue Nile: Prediction using teleconnections with sea surface temperatures. J. Journal of Great Lakes Research., 48 (2): 468-477.
16. Fathian, F.; Modarres, R.; Dehghan, Z. (2016) Urmia Lake water-level change detection and modeling. J. Model. Earth Syst. Environ., 2 (203): 1-16.
17. Kakahaji, H.; Banadaki, H. D.; Kakahaji, A.; et al. (2013) Prediction of Urmia Lake Water-Level Fluctuations by Using Analytical, Linear Statistic and Intelligent Methods. J. Water Resources Management., 27: 4469-4492.
18. Ehteram, M.; Ferdowsi, A.; Faramarzpour, M.; et al. (2021) Hybridization of artificial intelligence models with nature inspired optimization algorithms for lake water level prediction and uncertainty analysis. J. Alexandria Engineering Journal., 60(2): 2193-2208
19. Hossein, K.; Hamed, D.; Abbas, K.; et al. (2013) Prediction of Urmia Lake Water-Level Fluctuations by Using Analytical, Linear Statistic and Intelligent Methods. J. Water resources management., 27(13): 4469-4492.
20. Phan, T. T. H.; Nguyen, X. H. (2020) Combining statistical machine learning models with ARIMA for water level forecasting: The case of the Red River. J. Advances in Water Resources., 142 (103656): 1-15.
21. John, R.; John, M. (2019) Adaptation of the visibility graph algorithm for detecting time lag between rainfall and water level fluctuations in Lake Okeechobee. J. Advances in Water Resources., 134(9), 103429: 1-16.
22. Cheng, V. Y. S.; Saber, A.; Arnillas, C. A.; et al. (2021) Effects of hydrological forcing on short- and long-term water level fluctuations in Lake Huron-Michigan: A continuous wavelet analysis. J. Journal of Hydrology., 603: 127164.
23. Zhu, S. L.; Hrnjica, B.; Ptak, M.; et al. (2020) Forecasting of water level in multiple temperate lakes using machine learning models. J. Journal of Hydrology., 585, 124819.
24. Morovati, K.; Nakhaei, P.; Tian, F. Q.; et al. (2021) A Machine learning framework to predict reverse flow and water level: A case study of Tonle Sap Lake. J. Journal of Hydrology., 603, 127168.
25. Barzegar, R.; Aalami, M. T.; Adamowski, J. (2021) Coupling a hybrid CNN-LSTM deep learning model with a Boundary Corrected Maximal Overlap Discrete Wavelet Transform for multiscale Lake water level forecasting. J. Journal of Hydrology., 598, 126196.
26. Mohammadi, B.; Guan, Y. Q.; Aghelpour, P.; et al. (2021) Simulation of Titicaca Lake Water Level Fluctuations Using Hybrid Machine Learning Technique Integrated with Grey Wolf Optimizer Algorithm. J. Water., 12(11), 2015: 1-18.
27. Fan, H. X.; He, H. D.; Xu, L. G.; et al. (2021) Simulation and attribution analysis based on the long-short-term-memory network for detecting the dominant cause of runoff variation in the Lake Poyang Basin. J. Journal of Lake Sciences., 33(3): 866-878.
28. Zhang, F.; Wang, J. Q.; Luo, H. (2016) Analysis and Forecast of Poyang Lake's water level base on Time Series. J. Geomatics & Spatial Information Technology., 39 (08): 35-37+41.
29. Li, Y. L.; Zhang, Q.; Li, M.; et al. (2015) Using BP neural networks for Water level simulation in Poyang Lake. J. Resources and Environment in the Yangtze Basin., 24 (2): 233–240.
30. Guo, Y.; Lai, X. J. (2021) Research on water level prediction of Dongting lake based on Recurrent Neural Network. J. Resources and Environment in the Yangtze Basin., 30 (03): 689-698.
31. Wang, M. M.; Dai, L. Q.; Dai, H. C.; et al. (2017) Support vector regression based model for predicting water level of Dongting Lake. J. Journal of Drainage and Irrigation Machinery., 35 (11): 954-961.
32. Li, Y.; Shi, H.; Liu, H. (2020) A Hybrid Model for River Water Level Forecasting: Cases of Xiangjiang River and Yuanjiang River, China. J. Journal of Hydrology., 587, 124934.
33. Nair, A. S.; Kumar, N.; Indu, J. Vivek B. (2021) Monitoring Lake Levels From Space: Preliminary Analysis With SWOT. J. Frontiers in water., 3(12), 717852: 1-9.
34. Bandini, F.; Jakobsen, J.; Olesen, D.; et al. (2017) Measuring water level in rivers and lakes from lightweight Unmanned Aerial Vehicles. J. Journal of Hydrology., 548: 237-250.
35. Wu, G. P.; Liu, Y. B.; Liu, R. (2022) Assessing the performance of the Tiangong-2 wide-swath imaging altimeter observations for water level monitoring over complex and shallow lakes. J. Journal of Hydrology., 612, 124819.
36. Zhu, C. M.; Zhang, X.; Huang, Q. H. (2018) Lake hydrological information estimation based on remote

sensing. J. Journal of China Hydrology., 38(5): 29-33.

37. Devia, G. K.; Ganasri, B. P.; Dwarakish, G. S. (2015) A Review on Hydrological Models. J. Aquatic Procedia., 4: 1001-1007.
38. Karimi, S.; Shiri, J.; Kisi, O.; et al. (2012) Forecasting Water Level Fluctuations of Urmieh Lake Using Gene Expression Programming and Adaptive Neuro-Fuzzy Inference System. J. The International Journal of Ocean and Climate Systems., 3(2): 109-125.
39. Karimi, S.; Kisi, O.; Shiri, J.; et al. (2013) Neuro-fuzzy and neural network techniques for forecasting sea level in Darwin Harbor, Australia. J. Computers and Geosciences., 52: 50-59.
40. Wang, C. H. (2016) Application of wavelet neural network model in Liaoning plain groundwater prediction. J. Water Sciences and Engineering Technology., (03): 44-46.
41. Buyukyildiz, M.; Tezel, G.; Yilmaz, V. (2014) Estimation of the Change in Lake Water Level by Artificial Intelligence Methods. J. Water Resources Management., 28(13): 4747-4763.
42. Wang, X. J.; Mao, H. T.; Huang, Q. H.; et al. (2021) Analysis of rainfall-runoff relationship before and after impoundment of Three Gorges Reservoir Wan County section based on Copula function. J. Journal of Water Resources and Water Engineering., 32 (02): 23-30+37.
43. Haff, L.H.; Frigessi, A.; Maraun, D. (2015) How well do regional climate models simulate the spatial dependence of precipitation? An application of pair-copula constructions. J. Geophys. Res. Atmos., 120 (7): 2624-2646.
44. Wu, H. O.; Tu, X. J.; Du, Y. L.; et al. (2019) Multidimensional analysis of wetness-dryness encountering of streamflow based on the Copula function in Lake Poyang Basin. J. Journal of Lake Sciences., 31 (03): 801-813.
45. Chen, L.; Guo, S. L. (2019) Copulas and Its Application in Hydrology and Water Resources. M. Singapore., 97-116.
46. Scott, C. W.; Scott, S.; William, F.; et al. (2019) Copula theory as a generalized framework for flow-duration curve based streamflow estimates in ungaged and parially gaged catchments. J. Water Resources Research., 55 (11): 9378-9397.
47. Gu, S. X.; Zhao, Z.; Chen, J.; et al. (2020) Daily reference evapotranspiration and meteorological drought forecast using high-dimensional Copula joint distribution model. J. Transactions of the Chinese Society of Agricultural Engineering., 36 (09): 143-151.
48. Yan, B. W.; Pan, Z.; Xue, Y.; et al. (2017) Modeling dependence and correlation in hydrological calculation. J. Journal of Hydraulic Engineering., 48 (09): 1039-1046.
49. Fan, G. B.; Zeng, Y.; Huang, W. G. (2013) A new approach for measuring the risk of portfolios with Multiple Assets. J. Journal of Quantitative & Technological Economics., 30 (01): 88-102.
50. Daneshkhah, A.; Remesan, R.; Chatrabgoun, O.; et al. (2016) Probabilistic Modeling of Flood Characterizations with Parametric and Minimum Information Pair-Copula Model. J. Journal of Hydrology., 540: 469-487.
51. Haberlandt, U.; Ebner von Eschenbach, A.-D.; Buchwald, I. A. (2008) Space-Time Hybrid Hourly Rainfall Model for Derived Flood Frequency Analysis. J. Hydrology and Earth System Sciences., 12(6): 1353-1367.
52. Ji, C. M.; Ma, H. Y.; Peng, Y.; et al. (2021) Prediction of Meteorological drought in Huai River Basin considering multiple climatic indices. J. China Rural Water and Hydropower., (04): 16-21.
53. Pereira, G.; Veiga, A. (2018) PAR(p) -vine copula based model for stochastic streamflow scenario generation. J. Stoch. Environ. Res. Risk Assess., 32 (3): 833-842.
54. Pereira, G. A. D.; Veiga, A. (2019) Periodic copula autoregressive model designed to multivariate streamflow time series modelling. J. Water Resour. Manag., 33 (10): 3417-3431.
55. Pham, M. T.; Vernieuwe, H.; De Baets, B.; et al. (2016) Stochastic simulation of precipitation-consistent daily reference evapotranspiration using vine copulas. J. Stochastic Environmental Research and Risk Assessment., 30(8): 2197-2214.
56. Edited by Nanjing Institute of geography and lakes, Chinese Academy of Sciences China Lake survey report. M. Beijing: Science Press, 2019.
57. Gowing, J. w.; Ejieji, C. J. (2001) Real-time scheduling of supplemental irrigation for potatoes using a decision model and short-term weather forecasts . J. Agricultural Water Management., 47: 137-153.
58. Kjersti, A.; Claudia, C.; Arnoldo, F.; et al. (2007) Pair-copula constructions of multiple dependence. J. Insurance Mathematics and Economics., 44(2): 182-198.
59. Shan, B. Y.; Guo, S. S.; Wang, Y. Z.; et al. (2021) Vine copula and cloud model-based programming approach for agricultural water allocation under uncertainty. J. Stochastic Environmental Research and Risk Assessment., : 1-21.
60. Su, J. G.; Wang, F.; Gao, R.; et al. (2020) Water resources allocation in Erhai irrigated district considering catchment water-cycle health. A. IOP Conference Series: Earth and Environmental Science., 508(1), 01215.
61. Deng, W.; Chen, J.; Gu, S. X.; et al. (2020) Hydrological Variability of Water Level of Dianchi Lake and Its Application. A. IOP Conference Series: Materials Science and Engineering., 780(1), 062048.