

Article

A Vine Copula-Based Modeling for Identification of Multivariate Water Pollution Risk in an Interconnected River System Network

Ruolan Yu ¹, Rui Yang ¹, Chen Zhang ^{1,*} , Maria Špoljar ², Natalia Kuczyńska-Kippen ³ and Guoqing Sang ^{4,*}

¹ State Key Laboratory of Hydraulic Engineering Simulation and Safety, Tianjin University, Tianjin 300072, China; yuruolan97@163.com (R.Y.); rayyang@whu.edu.cn (R.Y.)

² Department of Biology, Zoology, Faculty of Science, University of Zagreb, Rooseveltov trg 6, HR-10000 Zagreb, Croatia; maria.spoljar@biol.pmf.hr

³ Department of Water Protection, Faculty of Biology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznan, Poland; kippen@hotmail.pl

⁴ School of Water Conservancy and Environment, University of Jinan, Jinan 250012, China

* Correspondence: emil@tju.edu.cn (C.Z.); sangguoqing@163.com (G.S.)

Received: 13 August 2020; Accepted: 29 September 2020; Published: 30 September 2020



Abstract: The Interconnected River System Network (IRSN) has become a popular and useful measure to realize the long-term health and stability of water bodies. However, there are lots of uncertain consequences derived from natural and anthropogenic pressures on the IRSN, especially the water pollution risk. In our study, a Vine Copula-based model was developed to assess the water pollution risk in the IRSN. Taking the ponds around Nanyang station as research objects, we selected five proxy indicators from water quality indexes and eutrophication indexes, which included dissolved oxygen (DO), total nitrogen (TN), total phosphorus (TP), chlorophyll-a (*Chla*), and ammonia nitrogen (*NH₃-N*). Models based on three classes of vine copulas (C-, D-, and R-vine) were utilized respectively to identify the water quality indicators before and after the operation of the connection project. Our results showed that *TN*, *Chla*, and *NH₃-N* should be considered as key risk factors. Moreover, we compared the advantages and prediction accuracy of C-, D-, and R-vine to discuss their applications. The results reveal that the Vine Copula-based modeling could provide eutrophication management reference and technical assistance in IRSN projects.

Keywords: risk identification; water quality; Interconnected River System Network; multiple uncertainties; Copula function

1. Introduction

River and lake function are mutually affected by river-lake connectivity, which relates to the water supply, flood control, and even economic development. However, water issues are constantly changing along with the development of society and the economy. Human activities result in changes of river-lake connectivity and acceleration of eutrophication [1]. Water transition from rivers to lakes has become a popular and useful measure to improve water quality and prevent or mitigate the ecological deterioration of water systems [2], such as Lake Monsees in USA [3], Banas catchment in India [4], and Lake Dianchi in China [5].

The Interconnected River System Network (IRSN) is a new water transition method with a certain hydraulic connection, which was established among different water systems [6]. By increasing the hydraulic liquidity and continuity in water bodies, this new method can improve self-recovery ability and realize the long-term health and stability of water bodies to solve the environmental issues

associated with water [7]. For water resource conservation and sustainability, IRSN improves the heterogeneous distribution of water resources and freshwater ecological restoration and has become an effective measure to increase efficiency and allocation of water resources [8].

However, there are subsequently a number of uncertain consequences derived from natural and anthropogenic pressures on the IRSN, especially ecological risk to the water environment, which has caused widespread interest and attention. Smaller area and lower depth promote higher exposure to resuspension and higher nutrient concentrations; therefore, small water bodies need more concern [9]. The Water Quality Index (WQI) is one of the most frequently used tools for assessing water quality [10]. The Nemerow pollution index [11], Organic Pollution Index, and Eutrophication Index [12] have also been used to quantify the level of water pollution and water quality condition. In addition, cross-coupling among different methods and models has emerged in water quality assessment to reduce the shortages of each model. The SWAT-ANN model combined the small watershed model SWAT and the Artificial Neural Network (ANN) for improving the accuracy of water quality prediction [13,14]. Wu et al. [15] coupled Grey Target Theory (GTT) with the traditional Analytic Hierarchy Process (AHP) to eliminate the subjective errors of the AHP. The above-mentioned models are commonly applied to single water quality index analysis, and then the identified critical indicators are formed into a data set to conduct an overall assessment of the water environment without taking the correlation between different indicators into consideration [16].

As a powerful and flexible toolbox to characterize dependence profiles for multi-variables [17], Copula has been the subject of widespread interest in hydrology. Compared with other models, Copula could directly establish the joint distribution function of critical indicators and analyze the multifactorial combined risk caused by key indicators which exceed standard limits [18]. Recently, Copula has been widely used in studies of multivariate hydrological indicators analysis, particularly in flood frequency calculation [19,20], rainfall frequency analysis [21,22], and drought characteristics analysis [23–25]. In the field of water pollution risk assessment, Copula has mainly been used for two-dimensional and three-dimensional research, eutrophication of water quality analysis [26,27], and water quantity and quality risk assessment [28,29]. When facing high-dimensional problems, the above method is not powerful enough to model possible mutual dependencies among all variables [30].

The work of Joe [31], Bedford, and Cooke [32,33] proposed a radically reliable way to construct complex multivariate highly dependent models called Vine Copula [30,34]. Vine Copula decomposes the high-dimensional copula into several bivariate copulas, which greatly improves the flexibility of the model. Different decomposition structures result in different Vine Copula models. The regular vine (R-vine) first appeared in public as a graphical tool by Bedford and Cooke [32,33]; they also introduced the so-called canonical (C-vine) and drawable vine (D-vine). On these grounds, Aas et al. [34] extended these models and indicated that C-vine exhibited star shape structures having a tree sequence, whereas, D-vine possessed a path structure [35].

In this paper, a Vine Copula-based model was developed to identify the water pollution risk factors in the IRSN. As an efficient multiple-variables analysis tool in water ecological environment, the model based on C-vine and D-vine [36,37] has found an increasingly wide utilization, and there have been relatively few cases of application of R-vine [38,39]. However, the difference of identification results for three classes of vine copulas has not been fully addressed, that is, the uncertainties of the model structure need to be further investigated. Accordingly, we propose to use Vine Copula to identify the key factors from multiple water quality indicators, which could provide early warning and forecasting of any impending water pollution risks. Therefore, the objective of this study is to develop a Vine Copula-based analysis system, considering multiple uncertainties in a water environment, and apply it concretely to identifying the key water pollution indicators. Furthermore, we analyze the difference of identification results of C-, D-, and R-vine to discuss the distinctions of different vine structures for the purpose of choosing the most suitable structure for a case study in the future work.

2. Materials and Methods

2.1. Study Area

Nanyang Lake ($35^{\circ}4'–35^{\circ}20' N$; $116^{\circ}34'–116^{\circ}42' E$) is located in Jining city, Shandong province, central western China, which is generally known as a shallow and eutrophic lake. Nanyang Lake has an average water depth of 1.5 m and a total surface area of 215 km^2 , with an annual average precipitation of 717 mm and an annual average evaporation of 1074 mm. The safety flow rates of the Si River, Zhuzhao New River, Wanfu River, and Baima River, which are connected to Nanyang Lake are 5140, 1700, 830 and $558 \text{ m}^3/\text{s}$, respectively. Over the past two decades, with the development of industry and urbanization, great quantities of industrial wastewater and sanitary sewage without effective treatment have been discharged into Nanyang Lake. Point pollution sources were the major pollution, among which the effluent of the coal, medical, and bio-chemical industries are of importance. The level of pollution was greatly beyond the self-purification capacity of the water system and caused the water quality to deteriorate gradually with increased concentration of nitrogen, phosphorus, suspended matters, and other organisms.

As the water quality monitoring site of the Lake, Nanyang station is situated on an island in the center of Nanyang Lake (Figure 1). Ponds around Nanyang station form a reedy, marsh-covered wetland area. In this paper, we selected five proxy indicators from water quality indexes and eutrophication indexes and collected the indicator data monthly from 2008 to 2014, which included dissolved oxygen (DO), total nitrogen (TN), total phosphorus (TP), chlorophyll-a (Chla), and ammonia nitrogen ($\text{NH}_3\text{-N}$). To relieve water shortage in North China, the central government launched the strategically significant South-to-North Water Transfer Project (SNWTP). As the water transfer channel and impounded lake of the East Route of SNWTP, Nanyang Lake changed its hydraulic regulation after the operation of the project began, which gave rise to the changes in the water environment factors. The IRSN study of Nanyang Lake formally transferred water in November 2013, which was taken as a demarcation line; the model was divided into two periods—before the operation of the IRSN (B-IRSN), the period from January 2008 to October 2013, and after the operation of the IRSN (A-IRSN), for the period from November 2013 to December 2014.

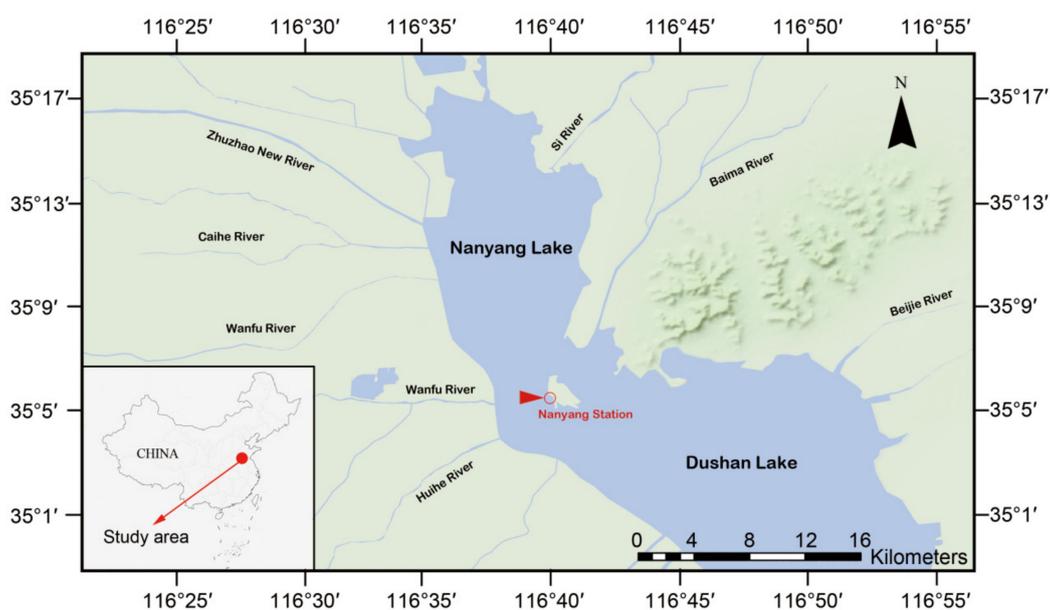


Figure 1. Map showing the location of the ponds around Nanyang station.

2.2. Water Pollution Risk Definition

Risk refers to the uncertainty of a certain outcome in a given situation, commonly using probability to make a prediction of potential accidents as well as trends. Tung and Mays [40] defined the risk of the system as the probability when the load l on the system exceeds the resistance r , which can be expressed as

$$R = P_r(l > r) = P_r(l - r > 0) \tag{1}$$

where R represents the risk of the system, $P_r()$ represents probability.

To the water environment system of the IRSN discussed in this paper, "load" could be defined as the pollutant concentrations, while "resistance" is the limited values of pollutants within the specified range. Thus, water pollution risk (R_w) can be computed as:

$$R_w = P_r(p_c > p_v) = \frac{m}{N} \tag{2}$$

where p_c refers to the pollutant concentrations, p_v refers to the maximum allowed value of pollutant concentration, m is the number of times that the measured contraindication was above the limit, N is all the water pollution events. Accordingly, the water pollution risk of the IRSN can be described as the probability of water quality exceeding the standard values under the influence of various uncertain pollutants before and after the operation of the connection project.

2.3. Multivariate Dependence Modeling Based on Vine Copula

In a more formal definition, Copula is a multivariate joint probability distribution function on the unit square $[0, 1]$ with uniform marginal distributions [30]. According to Sklar's theorem, the corresponding joint distribution function $F(X_1, X_2, \dots, X_n)$ can be expressed as follows:

$$F(X_1, X_2, \dots, X_n) = C_\theta(F_1(x), F_2(x), \dots, F_n(x)) \tag{3}$$

where $C_\theta()$ is a copula distribution function, θ represents an explicit parameter to the function, $F_1(x), F_2(x), \dots, F_3(x)$ to denote the marginal distribution function of random variables.

As the system of IRSN is so large and complicated and involves lots of risk factors, which mostly have dependency relationship, it is necessary to build the joint probability distribution for a multivariable. Bedford and Cooke introduced a more advanced and flexible alternative method of constructing the dependence structure called Vine Copula [41]. At least $n(n-1)/2$ bivariate copulas with a free specification can be established between n given variables under this flexible structure [30]. That means Vine Copula addresses the flexibility limitation of classical copulas to provide different tail dependencies for different variable pairs. C-vine and D-vine are the most common copula structures in practice. C-vine has a star structure with a root node that connects all other nodes for each tree, and therefore, it is applicable to fit a multi-variable with a key variable that controls interactions in the data. The general expression of the n-dimensional joint probability density of C-vine is given as

$$F(x_1, x_2, \dots, x_n) = \prod_{k=1}^n f(x_k) \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} C_{j,j+1|1, \dots, j-1} \{F(x_j|x_1, \dots, x_{j-1}), F(x_{j+1}|x_1, \dots, x_{j-1})\} \tag{4}$$

D-vine has a path structure, and each node has at most two edges. D-vines might actually be more beneficial than C-vines when we do not want to assume the existence of a key note that governs the dependencies [42]. The n-dimensional joint probability density of D-vine is expressed as

$$F(x_1, x_2, \dots, x_n) = \prod_{k=1}^n f(x_k) \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} C_{i,i+j|i+1, \dots, i+j-1} \{F(x_i|x_{i+1}, \dots, x_{i+j-1}), F(x_{i+j}|x_{i+1}, \dots, x_{i+j-1})\} \tag{5}$$

where $C(\cdot)$ refers to the bivariate copula with index i running over the edges for each tree and index j identifying the trees, $f(\cdot)$ denotes the marginal density, and $F(x|v)$ is the conditional distribution function. Figure 2a,b shows the four-dimensional structure of the C-vine and D-vine, respectively.

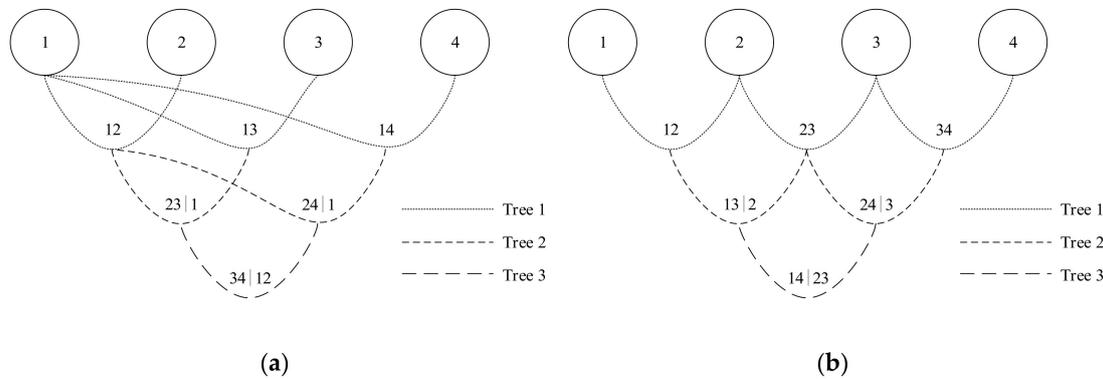


Figure 2. Four-dimensional vine copula structure, where each edge is associated with a pair-copula: (a) A C-vine structure with 4 variables, 3 trees, and 6 edges, where node 1 is the root node; (b) A D-vine structure with 4 variables, 3 trees, and 6 edges. Specifically, the numbers at each edge represent the bivariate copula between any adjacent variables.

However, the structures of C-vine and D-vine are relatively fixed and cannot be used in fitting more sophisticated dependence structures. The class of R-vine distributions is much larger than the class of C- and D-vine distributions. Unfortunately, due to the huge number of possible R-vine tree sequences, it has not been widely applied in practice. Dissmann et al. [39] developed an automated strategy that combined Maximum Spanning Tree (MST) algorithm with an R-vine copula, which simplified the original selection process and provided concise and efficient multi-dimensional data modeling. The n-dimensional density function of an R-vine is defined as

$$F(x_1, x_2, \dots, x_n) = \prod_{k=1}^n f(x_k) \prod_{j=n-1}^1 \prod_{i=n}^{j+1} C_{j,i|i+1, \dots, n} \{F(x_j|x_{i+1}, \dots, x_n), F(x_i|x_{i+1}, \dots, x_n)\} \quad (6)$$

Since the decomposition structure of an R-vine is too complex and does not allow for an easy way to express inference algorithms [39], this article uses the R-vine Matrix (RVM) to represent an R-vine. The following five-dimensional matrix, taken as an example, provides a more simple and efficient method to express a five-dimensional R-vine.

$$M^* = \begin{pmatrix} & & & & 5 \\ & & & & 4 & & & & \\ & & & & 3 & & & & \\ & & & & 2 & & 1 & & & \\ & & & & 1 & & 2 & & & \\ & & & & 1 & & 2 & & 1 & & \\ & & & & & & & & & & 1 \end{pmatrix} \quad (7)$$

All the nodes of Tree 1 are arranged in an orderly way on the master diagonal and combined with the nodes in the last row of M^* , respectively as edges of Tree 1. In this example, (5,1)(4,2)(3,2)(2,1) are the edges of Tree 1. The above combinations with conditioned nodes in the penultimate row of M^* are the edges of Tree2, which is (5,2|1)(4,1|2)(3,1|2), and so on, for each of the nodes in M^* . This completes the construction of R-vine structure.

The construction of fitting a multivariate dependence model using a C-, D- or R-vine includes the following three steps:

1. Selecting the connection order of the variables. For a C-vine, Czado et al. [43] introduced a method that calculates the absolute values of Kendall’s tau coefficients for pairwise variables and selects the variable with the maximum sum of absolute values as the root node. For a D-vine, we order the variables that define a tree that maximizes a given dependence measure used as edge weights according to Kendall’s tau; review Brechmann [44] for details. For the R-vine, this paper uses a maximum spanning tree algorithm such the algorithms of Prim (MST-PRIM), which maximizes the sum of absolute values of Kendall’s tau in every tree to select the suitable RVM. Calculate the sum of the absolute values of Kendall’s tau coefficients by Equation (8).

$$S_i = \sum_{j=1}^n |\tau_{i,j}| \tag{8}$$

where S_i refers to the sum of the absolute values of Kendall’s tau coefficients, τ refers to the Kendall’s tau coefficients. Thus, the Kendall tau coefficient matrix among above five water quality variables is given in Equation (9). The data before “/” represent results in the period of B-IRSN, and the data after “/” represent results in the period of A-IRSN.

$$\tau = \begin{pmatrix} & \text{DO} & & & & & \\ & 0.003 / -0.045 & & & & & \\ & -0.018 / 0.067 & \text{TN} & & & & \\ & 0.057 / 0.006 & 0.003 / -0.007 & & & & \\ & 0.043 / -0.053 & 0.014 / -0.009 & \text{TP} & & & \\ & & -0.013 / 0.047 & 0.014 / 0.022 & & & \\ & & & -0.013 / -0.015 & \text{Chla} & & \\ & & & & -0.007 / -0.033 & & \\ & & & & & \text{NH3-N} & \end{pmatrix} \tag{9}$$

2. Choosing the suitable binary copula function for each pair-copula. Taking the Akaike Information Criterion (AIC) as selection criteria [40], this paper selects the most suitable binary copula function from ten alternative copulas (Gaussian copula, Student’s t copula, Clayton copula, Gumbel copula, Frank copula, Joe copula, and BB copula).
3. Estimating parameters of all vine copulas. Here, we estimate the parameters of vine copulas by the maximum likelihood estimation (MLE). The specific steps are as follows. Firstly, the binary copula parameters of Tree 1 can be estimated using the MLE method. Secondly, the conditional function $F(x|v)$ can be calculated as the observations of Tree 2 using h-functions as follows:

$$h(x|v, \theta) := F(x|v) = \frac{\partial C_{xv_j|v_{-j}}(F(x|v_{-j}), F(v_j|v_{-j}))}{\partial F(v_j|v_{-j})} \tag{10}$$

where v_j refers to an arbitrary element of v , v_{-j} is the vector v excluding v_j , and $C(\)$ represents the bivariate copula distribution function to $F(x|v)$. The observations obtained in the previous step are then taken to estimate the parameters of Tree 2. The parameter estimation of all trees is completed by repeating the above steps [45]. This completes multivariate dependence modeling based on the C-, D-, and R-vine copula.

2.4. Water Pollution Risk Identification Model

In this paper, we developed a Vine Copula-based modeling to identify a multivariate water pollution risk, which includes primarily three parts:

1. Processing of basic data. For the initial observational series, we calculated their eigenvalues and fitted them with six common hydrological distributions (Normal distribution, Log-normal distribution, exponential distribution, logistic distribution, Gamma distribution, and Weibull distribution). According to the goodness-of-fit test criteria, we selected the best-fit distribution which has the lowest AIC value. In this study, we collected 70 sets of data before the operation of the IRSN, for the period from January 2008 to October 2013, and 14 sets of data after the operation

- of the IRSN, for the period from November 2013 to December 2014 [46]. Summary statics for the water quality variables were reported in the period of B-IRSN and A-IRSN, respectively (Table 2).
- Constructing vine copulas. Using the methods discussed in Section 2.3, we constructed appropriate five-dimensional vine copulas for five variables, *DO*, *TN*, *TP*, *Chla*, and *NH3-N*, before and after the operation of the IRSN. To compare the fit of the models, the Akaike Information Criterion (AIC), the Bayesian Information Criteria (BIC), and the Log-likelihood method were used to test the prediction accuracy of the C-, D-, and R-vine, respectively. In addition, to verify the models and further compare the C-, D-, and R-vine copula, the following work was accomplished. We brought the copula structure and parameter estimation results into the simulation function to generate 500 groups of simulated data and calculated Kendall's tau coefficients for pairwise variables. After 200 cycles, box plots were created according to the results of Kendall's tau coefficients in the previous step.
 - Identification of water pollution risk. In order to analyze the sensitivity of the water pollution risk and identify the key risk indicators, the influence extent of each water quality factors on the water pollution risk joint probability was analyzed as follows. According to the National Environmental quality standards of the surface water of China and water quality standard proposed by the Organization for Economic Co-operation and Development, the mean values of the initial series of risk factors were taken as the base-value, and the risk probability density calculated when the concentration of one of those variables varied in accordance with the water I-V quality standard described in Table 1. The higher class means the worse water quality. By comparing the evolution of the risk probability density before and after the operation of the connection project, it was possible to ascertain the effect of the IRSN operation on the environment and fully complete the whole water pollution risk identification. All statistical computations were performed using the **Vines**, **Vine copula**, and **CDVine** packages available in R software.

Table 1. Standard limits of water quality variables.

Variable	Standard Limits				
	I	II	III	IV	V
<i>DO</i> (mg/L)	7.5	6	5	3	2
<i>TN</i> (mg/L)	0.2	0.5	1	1.5	2
<i>TP</i> (mg/L)	0.01	0.025	0.05	0.1	0.2
<i>Chla</i> (μ g/L)	1	2.5	8	25	35
<i>NH3-N</i> (mg/L)	0.15	0.5	1	1.5	2

Table 2. Summary statistics of water quality variables. The data before “/” represent statistics in the period of B-IRSN, and the data after “/” represent statistics in the period of A-IRSN.

Variable	Max	Min	Mean	Std	Skewness	Kurtosis
<i>DO</i> (mg/L)	13.53/12.48	2.24/5.80	8.99/7.88	2.12/1.72	0.12/1.26	−0.99/1.12
<i>TN</i> (mg/L)	9.00/4.54	0.39/0.81	2.02/1.96	1.59/0.91	2.05/1.39	4.86/2.00
<i>TP</i> (mg/L)	0.34/0.11	0.01/0.00	0.06/0.04	0.05/0.03	3.39/1.34	13.56/0.50
<i>Chla</i> (μ g/L)	66.75/16.82	5.10/5.61	19.01/12.92	13.74/3.30	1.55/−0.52	1.91/−0.69
<i>NH3-N</i> (mg/L)	1.41/0.98	0.21/0.33	0.48/0.61	0.30/0.26	1.36/0.16	1.03/−1.80

3. Results

3.1. Descriptive Statistics and Marginal Distributions

Studied water quality variables (*DO*, *TN*, *TP*, *Chla*, and *NH3-N*) were the important pollution indicators in the ponds around Nanyang station. For the descriptive statistics of the above water quality variables, the standard deviation is quite small relative to their average, which indicated that the data of water quality were stable (Table 2).

The selected marginal distribution and their estimated parameters for each variable can be found in Table 3 along with the AIC values. From the results, the best fitted distributions of variables were log-norm distribution in most cases.

Table 3. The results of best fitted distribution to each variable along with their estimated parameters. The data before “/” represent results in the period of B-IRSN, and the data after “/” represent results in the period of A-IRSN.

Variable	Distributions	Estimated Parameters		AIC
		Parameter 1	Parameter 2	
DO	gamma/lnorm	17.74/2.04	1.97/0.19	306.10/54.83
TN	lnorm/lnorm	0.47/0.58	0.67/0.41	212.58/35.40
TP	lnorm/lnorm	−2.96/−3.44	0.60/0.62	−282.59/−66.00
Chla	lnorm/weibull	2.73/4.95	0.64/14.12	521.77/75.28
NH3-N	lnorm/gamma	−0.89/5.70	0.53/9.30	−10.51/3.95

Notes: Gamma is the gamma distribution, lnorm is the log-norm distribution, and weibull is the weibull distribution.

3.2. Vine Copula Model Construction

In this section, three classes of vine copula models (C-, D-, and R-vine) are constructed based on the method described in the previous section. The breakdown structures and parameter estimation for B-IRSN and A-IRSN period of three classes of vine copulas are listed and described in Tables 4 and 5, respectively. To facilitate the discussion, the five variables, DO, TN, TP, Chla, and NH3-N are denoted as 1, 2, 3, 4, and 5, respectively.

Table 4. The breakdown structures and best-fit copula families of three classes of vines in the period of B-IRSN.

Tree	C-Vine		D-Vine		R-Vine	
	Edge	Best-Fit Copula	Edge	Best-Fit Copula	Edge	Best-Fit Copula
Tree 1	1,2	C (0.06)	5,3	C270 (−0.05)	1,3	G (−0.04)
	1,3	G (−0.04)	1,5	G (0.07)	4,2	C (0.08)
	1,4	C (0.11)	4,1	C (0.11)	1,4	C (0.11)
	5,1	G (0.07)	2,4	C (0.08)	5,1	G (0.07)
Tree 2	5,2 1	F (−0.07)	1,3 5	G (−0.04)	4,3 1	G (0.03)
	5,3 1	C270 (−0.06)	4,5 1	C90 (−0.06)	1,2 4	G (0.02)
	5,4 1	C270 (−0.06)	2,1 4	G (0.02)	5,4 1	C270 (−0.06)
Tree 3	4,2 5,1	C (0.07)	4,3 1,5	G (0.03)	5,3 4,1	C270 (−0.04)
	4,3 5,1	G (0.03)	2,5 4,1	F (−0.08)	5,2 1,4	F (−0.08)
Tree 4	3,2 4,5,1	SJ (1.05)	2,3 4,1,5	SJ (1.05)	2,3 5,4,1	SJ (1.05)

Notes: It shows the best-fit copula for each edge where C—Clayton copula, G—Gaussian copula, F—Frank copula, C90—rotated Clayton copula (90 degrees), C270—rotated Clayton copula (270 degrees), and SJ—rotated Joe copula (180 degrees, “survival Joe”). 1, 2, 3, 4, and 5 represent DO, TN, TP, Chla, and NH3-N, respectively. The parameter estimation results of copulas are reported in square brackets.

Table 5. The breakdown structures and best-fit copula families of three classes of vines in the period of A-IRSN.

Tree	C-Vine		D-Vine		R-Vine	
	Edge	Best-Fit Copula	Edge	Best-Fit Copula	Edge	Best-Fit Copula
Tree 1	1,2	G90 (−1.04)	2,5	SJ (1.06)	1,3	G (0.11)
	1,3	G (0.11)	1,2	G90 (−1.04)	5,1	F (−0.42)
	1,4	G (−0.02)	3,1	G (0.11)	5,2	SJ (1.06)
	5,1	F (−0.42)	4,3	F (0.21)	5,4	F (−0.30)
Tree 2	5,2 1	SG (1.04)	1,5 2	F (−0.39)	5,3 1	J90 (−1.02)
	5,3 1	J90 (−1.02)	3,2 1	C270 (−0.01)	2,1 5	G270 (−1.04)
	5,4 1	F (−0.28)	4,1 3	G (−0.02)	4,2 5	C270 (−0.04)
Tree 3	4,2 5,1	C270 (−0.04)	3,5 1,2	J270 (−1.02)	2,3 5,1	G (0.01)
	4,3 5,1	F (0.20)	4,2 3,1	C270 (−0.04)	4,1 2,5	G (−0.02)
Tree 4	3,2 4,5,1	G (0.01)	4,5 3,1,2	C270 (−0.05)	4,3 2,5,1	G (0.03)

Notes: It shows the best-fit copula for each edge where G—Gaussian copula, F—Frank copula, G90—rotated Gumbel copula (90 degrees), J90—rotated Joe copula (90 degrees), C270—rotated Clayton copula (270 degrees), J270—rotated Joe copula (270 degrees), SG—rotated Gumbel copula (180 degrees, “survival Gumbel”), and SJ—rotated Joe copula (180 degrees, “survival Joe”). 1, 2, 3, 4, and 5 represent *DO*, *TN*, *TP*, *Chla*, and *NH3-N*, respectively. The parameter estimation results of copulas are reported in square brackets.

The first tree of the Vine Copula often has the greatest influence on the model fit [32]. Therefore, we may more directly observe the dependencies among uncertainty variables by analyzing the structures of the first tree of vine copulas. The results show the C-vine and D-vine structures (Tree 1) with families and their estimated parameters, respectively (Figures 3 and 4).

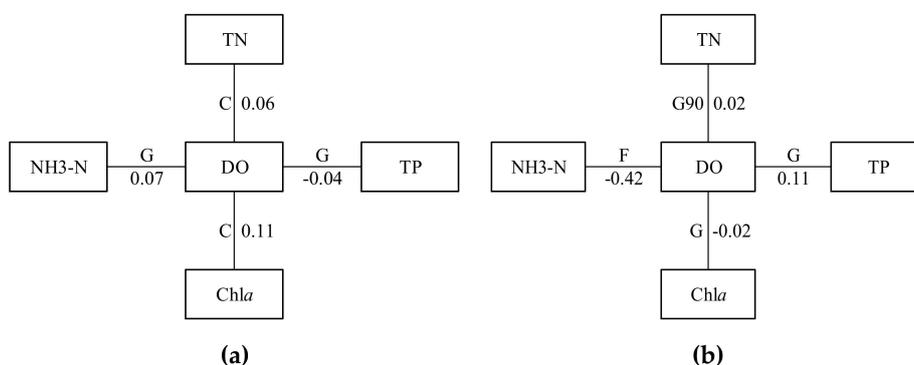


Figure 3. Structures of C-vine Tree 1, where C—Clayton copula, F—Frank copula, G—Gaussian copula, G90—rotated Gumbel copula (90 degrees), J90—rotated Joe copula (90 degrees), C270—rotated Clayton copula (270 degrees), and SG—rotated Gumbel copula (180 degrees, “survival Gumbel”) with estimated parameters shown on the links: (a) C-vine structure in the period of B-IRSN; (b) C-vine structure in the period of A-IRSN.

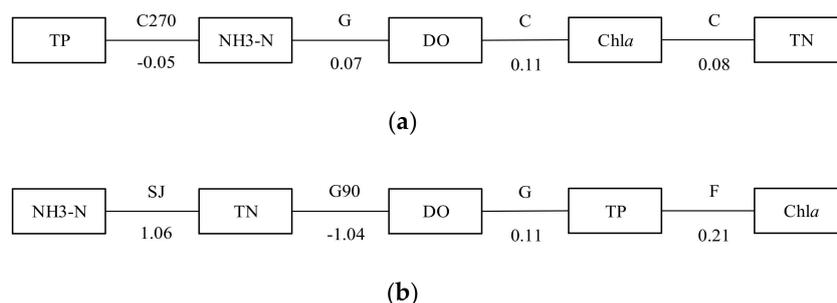


Figure 4. Structures of D-vine Tree 1, where C—Clayton copula, F—Frank copula, G—Gaussian copula, C90—rotated Clayton copula (90 degrees), G90—rotated Gumbel copula (90 degrees), C270—rotated Clayton copula (270 degrees), and SJ—rotated Joe copula (180 degrees, “survival Joe”) with estimated parameters shown on the links: (a) D-vine structure in the period of B-IRSN; (b) D-vine structure in the period of A-IRSN.

The construction results of RVM and best-fit copula family matrix as shown below (Figure 5). The RVM_{before} and RVM_{after} show that the R-vines have four trees in total. Tree 1 in RVM_{before} has four edges, (1,3), (4,2), (1,4), and (5,1). Tree 2 in RVM_{before} has three edges, (4,3|1), (1,2|4), and (5,4|1). Tree 3 in RVM_{before} has two edges, (5,3|4,1) and (5,2|1,4). Tree 4 in RVM_{before} has only one edge (2,3|5,4,1). The edges of trees in RVM_{after} can be defined in this similar manner.

$$RVM_{before} = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 & 0 \\ 5 & 5 & 4 & 0 & 0 \\ 4 & 1 & 5 & 1 & 0 \\ 1 & 4 & 1 & 5 & 5 \end{bmatrix}, C_{before} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ SJ & 0 & 0 & 0 & 0 \\ C270 & F & 0 & 0 & 0 \\ G & G & C270 & 0 & 0 \\ G & C & C & G & 0 \end{bmatrix}$$

(a)

$$RVM_{after} = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ 4 & 1 & 0 & 0 & 0 \\ 2 & 4 & 2 & 0 & 0 \\ 5 & 2 & 4 & 4 & 0 \\ 1 & 5 & 5 & 5 & 5 \end{bmatrix}, C_{after} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ G & 0 & 0 & 0 & 0 \\ G & G & 0 & 0 & 0 \\ J90 & G270 & C270 & 0 & 0 \\ G & F & SJ & F & 0 \end{bmatrix}$$

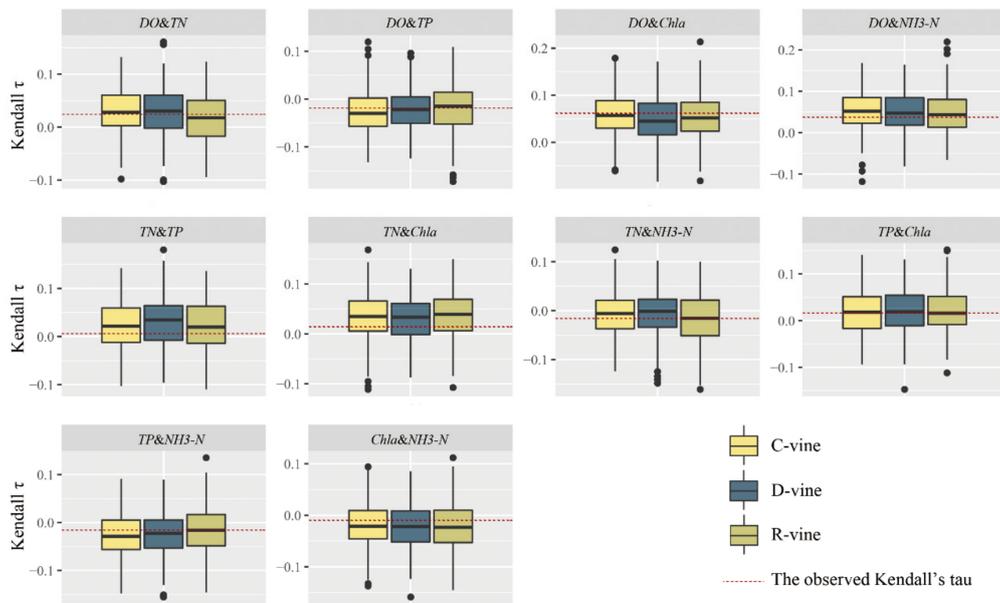
(b)

Figure 5. Structures of RVM, along with the best-fit copula family matrix, where C—Clayton copula, F—Frank copula, G—Gaussian copula, J90—rotated Joe copula (90 degrees), C270—rotated Clayton copula (270 degrees), G270—rotated Gumbel copula (270 degrees), and SJ—rotated Joe copula (180 degrees, “survival Joe”) along with the best-fit copula family matrix. 1, 2, 3, 4, and 5 represent DO, TN, TP, Chla, and NH3-N, respectively, (a) RVM and best-fit copula family matrix in the period of B-IRSN; (b) RVM and best-fit copula family matrix in the period of A-IRSN.

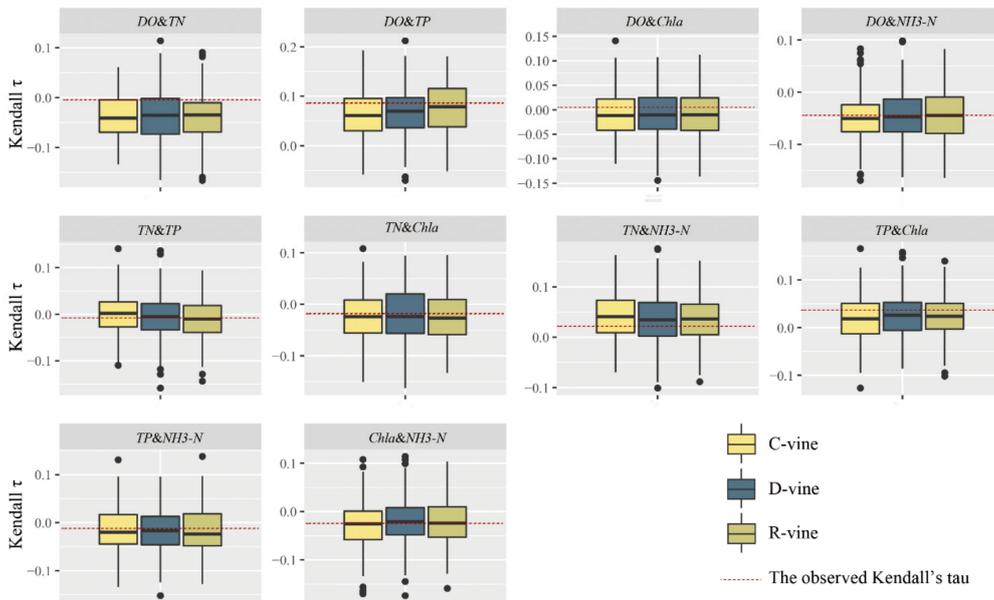
Both, before and after the operation of the IRSN, DO is generally connected with other variables as the key node (Figure 3, Figure 4, and Figure 5). Tail dependence provides measures for the dependence between extreme variables and measures the tendency of one variable to occur with a minimal or maximal outcome simultaneously with another variable [47,48]. The selection of the Gaussian copula suggests the dependence between variables is tail symmetric and exhibits no tail dependence, the selection of the Clayton copula indicates that lower tail dependence without upper tail dependence

between variables, and the Frank copula shows no tail dependence or the tail dependence is very weak. This means the five water quality variables are lowly correlated in general.

The observed and simulated Kendall's tau coefficients among the C-, D-, and R-vine were compared as shown in Figure 6. The observed statistics fall between the upper and lower quartiles of the boxplots, and most of them are near the median. It is indicated that simulated statistics fit the observed statistics well and the C-, D-, and R-vine achieve great performance in characterizing the dependence structure of multiple variables.



(a)



(b)

Figure 6. Comparisons of the observed and simulated Kendall's tau among C-, D-, and R-vine in the period of B-IRSN (a) and A-IRSN (b). The red lines refer to the observed Kendall's tau coefficients.

Furthermore, to evaluate which vine copula could fit the data better, AIC, BIC, and Log-likelihood were used to compare the C-, D-, and R-vine copula in Table 6. Generally, the lower the AIC and BIC values, the better the model; the higher the Log-likelihood value the better the model. C-vine has lower AIC and BIC values and a higher log-likelihood value than the other vines before the operation of the IRSN, while the D-vine was found to achieve greater performance after the operation of the IRSN (Table 6).

Table 6. AIC, BIC, and Log-likelihood for C-, D- and R-vine. The data before “/” represent the AIC, BIC, and Log-likelihood values in the period of B-IRSN, and the data after “/” represent the AIC, BIC, and Log-likelihood values in the period of A-IRSN.

Method	C-Vine	D-Vine	R-Vine
AIC	2/1.92	2.61/1.73	2.82/1.96
BIC	44.15/44.07	44.76/43.84	44.96/44.1
Loglik	9/9.04	8.69/9.14	8.59/9.02

The likelihood-ratio based test proposed by Vuong [49] among the C-, D-, and R-vine copula for pair comparisons listed in Table 7. The Vuong test can be used to compare the predicted probabilities of two copula models by measuring the closeness of a model to the truth. If the Vuong test statistic is positive, we tend to select the structure, families, and parameters of the first model. If the Vuong test statistic is negative, we then prefer the second model. Moreover, the difference between the two models is not statistically significant when the Vuong p -value is greater than 5%. The result shows that the Vuong test selects the C-vine structure, families, and parameters before the operation of the IRSN and the D-vine structure, families, and parameters after the operation of the IRSN (Table 7). In the 5% significance level, the Vuong p -value shows that the general difference for pairwise models is not obvious. This is probably because different observation angles are given according to the special structures of each two models.

Table 7. Vuong test for C-, D-, and R-vine. The data before “/” represent Vuong test values in the period of B-IRSN, and the data after “/” represent Vuong test values in the period of A-IRSN.

Value	C-D	C-R	D-R
Statistic	0.39/-0.10	0.51/0.03	0.65/0.13
p -value	0.70/0.92	0.61/0.98	0.52/0.90

3.3. Sensitivity Analysis of Water Pollution Risk Indicators

The sensitivity analysis of water pollution risk indicators can identify the key factors of joint water pollution risk from multiple uncertainty water quality characteristics. Using the sensitivity analysis method described in Section 2.4, a series of risk scenarios is set and applied to the established vine copula models to calculate the risk probability density and identify the key indicators. The risk scenarios involve 25 different conditions in each period (Figure 7).

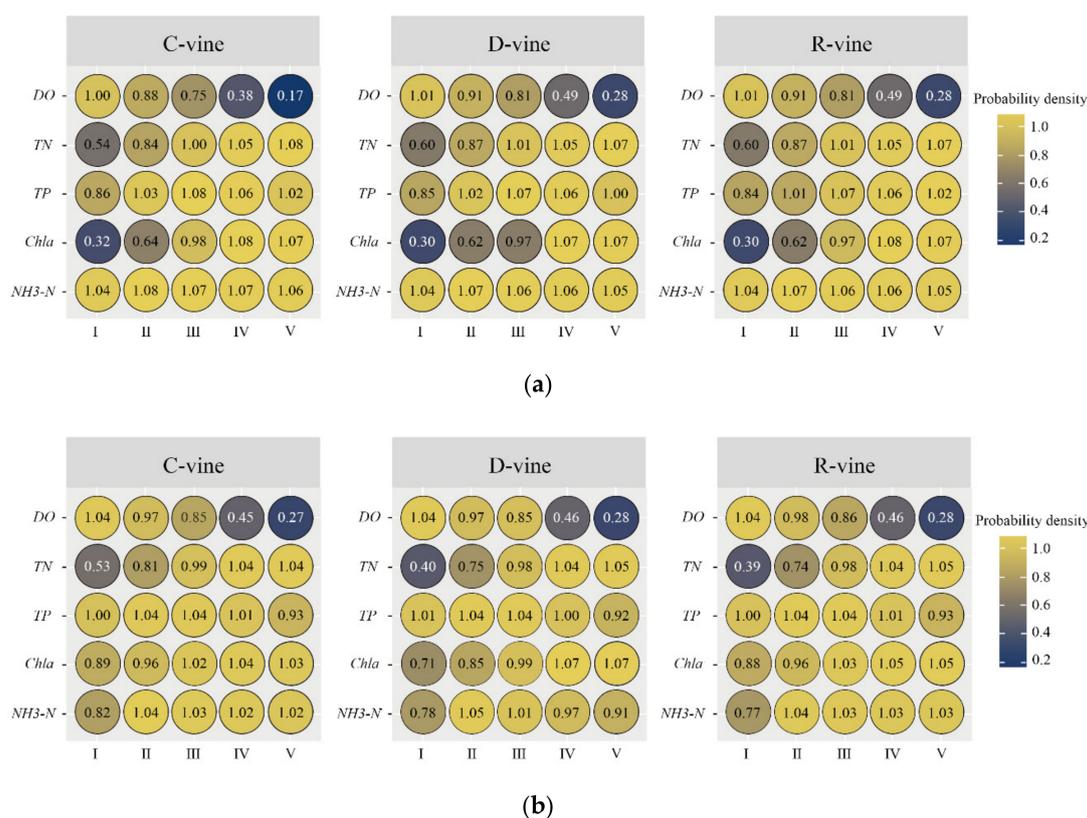


Figure 7. The risk probability density in different working conditions in the period of B-IRSN (a) and A-IRSN (b) by using C-, D-, and R-vine, respectively. When the indicators are in the same standard limits, the higher the probability density of a certain variable, the greater the impact on joint water pollution risk. The roman numbers represent the risk probability density.

The quality of ponds around Nanyang station needs to reach Class III; therefore, we define the water pollution risk when the quality of ponds around Nanyang station is worse than Class III water, namely Class IV, and Class V water. In the same standard limits, the higher the probability density of a certain variable, the greater the influence on the change rate of joint distribution, that is, the joint water pollution risk is more sensitive to it. Therefore, we consider the sensitive factors with less decreasing of the probability density after the operation of the IRSN as the key risk factors. From Figure 7, the probability densities of each variable except DO are higher in class IV and class V than other conditions. This means that with the possible exception of DO, other water quality variables are more sensitive to joint risk of water pollution. Comparing the probability densities before and after the operation of the IRSN, we find that the probability densities of TN, TP, Chla, and NH3-N decrease to some extent in class IV and class V in the period of A-IRSN, which indicates that the joint risk sensitivity of TN, TP, Chla, and NH3-N decreased after the operation of the IRSN. Through the comparison of the probability density of TN, TP, Chla, and NH3-N after the operation of the IRSN, it is noticeable that the probability density of TP drops markedly and it fell to below 1 in Class V water, which was the smallest figure compared with that in other Class water. It could be argued that the joint risk sensitivity of TP has an obvious change and it becomes the most insensitive factor among TN, TP, Chla, and NH3-N after the operation of the IRSN. Accordingly, TP was excluded and TN, Chla, and NH3-N were considered as the key risk factors.

In addition, we used the same method to calculate the risk probability densities for the full sample period and plotted the time-series graph of risk probability density changes by taking C-vine as an example (Figure 8). The figure facilitates to judge observations intuitively and clearly according to the color changes. By observing the color changes, it is clear that the probability densities of each indicator

decrease apparently and change smoothly after the operation of the connection project. Furthermore, through the analysis of the whole time series, one can readily observe how the risk probability density of a certain water quality variable is changing in time so that we can summarize the regularity of the variation trend for water quality variables and determine the impact on the risk probability density of water quality indicators caused by a certain water transfer project, which could provide useful reference for water quality managers. For example, via Figure 8, it is concluded that the operation of the IRSN improves the water quality and the achievements are generally obvious.

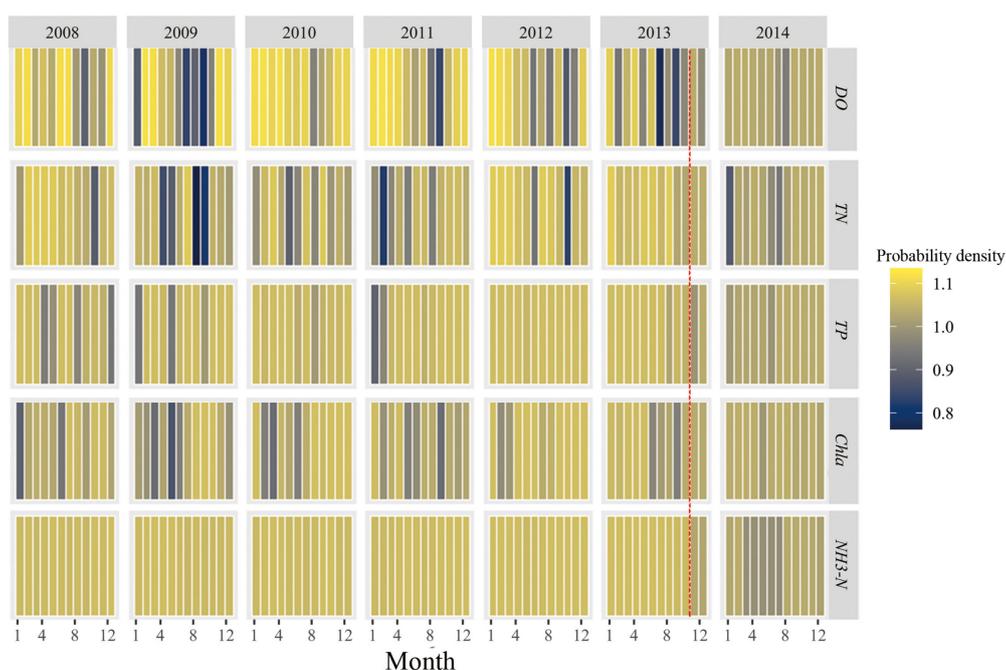


Figure 8. Time-series graph of risk probability density change for the full sample period. Taking the red line as the demarcation line, the model was divided into two periods—before (January 2008–October 2013, the left of the red line) and after (November 2013–December 2014, the right of the red line) the operation of the IRSN. The probability density is used to specify the probability of the random variable falling within a particular range of values. Through our calculation of the risk probability density for full sample period, the probability density of water quality indicators ranges from 0.76 to 1.13.

4. Discussion

In previous research, cross-coupling water quality models, which generally require an enormous amount of detailed data, are the main method to assess water pollution. However, when we are confronted with a limited dataset, the relatively poor prediction accuracy of these models limits their application. Whereas the statistical modeling for multiple variables analysis as the other method to identify the key indicators of water quality requires relatively less data [50].

Vine Copula is a commonly multiple variables analysis model, which can be used in modeling the water quality data with more potential flexibility and adaptability over alternative multivariate copulas, particularly when the data is limited [30]. From a statistical point of view, the monthly data from 2008 to 2014 in this article are too few to fit the multivariate model. However, from the analysis result, Vine Copula gave a great performance in characterizing the dependence structure of multiple variables. In fact, Vine Copula is an essential and commonly used method for the analysis of extreme data or even missing data in hydrology [51–53]. It should be noted that we just take the water quality data from the one measurement location due to the insufficient data available. That is to say, this is rather a methodological paper. Actually, one of the goals of our study is proving how the Vine Copula-based modeling proposed in this paper can be used to identify the key water risk

factors, particularly in a small water body. We just take the ponds around Nanyang station as an example to demonstrate the flexibility and efficiency of Vine Copula in modeling the water quality data and identifying the key water quality indicators. However, this could equally be used in several measurement locations. According to the identification results, the Vine Copula-based modeling can perform well in characterizing the dependence structure of multiple water quality variables, and successfully identified the key indicators in the ponds around Nanyang station using the limited water quality data.

Apparently, the previous evaluation results show that AIC, BIC, Log-likelihood, and the Vuong test lead to a consistent selection (Tables 6 and 7). It seems like C-vine could fit the data better before the operation of the IRSN, while D-vine has a greater performance after the operation of the IRSN. Next, we further discuss the cause for this result. It can be found that the impact of each water quality indicator on joint water pollution risk is quite different before the operation of the IRSN, and *DO* is the one that has the greatest impact on joint risk in most cases (Figure 8). But the impact of water quality factors on joint risk tend to nearly stable after the operation of the IRSN, that is, the leading role of *DO* is no longer significant. This further confirms that the C-vine is applicable to fit a multi-variable with a key variable that controls interactions in the data, and the D-vine is suitable especially when variables are relatively independent. The advantage of the D-vine copula to fit the variables with unclear dependence, which is a feature that the C-vine copula models do not have. As for the R-vine, it has a greater advantage for dealing with the massive volume and high-dimension of data. This is why the previous evaluation results occurred.

As two popular classes of vine copula structures, C- and D-vine copula have been widely used in various fields. Actually, when facing the high-dimension data, the pivotal elements and independent variables may exist at the same time. R-vine copulas weaken the properties of conditional independence and enhance the flexibility and universality of the model, which make R-vines have the ability to fit a more sophisticated series of multiple dependencies even in larger dimensions. However, R-vines are seldom applied by the enormous number of possible R-vine tree sequences. The MST-PRIM algorithm, which is used to construct R-vine copula in this paper, can successfully remove this limitation. We recognize this, and in our future work, R-vine Copula-based modeling will be given priority to identify the high-dimension water quality variables in IRSN projects.

5. Conclusions

In this paper, we proposed a novel method based on vine copula to identify the key indicators of water pollution risk before and after the operation of the IRSN. Through this study, the major conclusions are follows:

1. A feasible assessment system for water pollution risk was established to identify the key indicators of water pollution risk in IRSN. The system achieved great performance in characterizing the dependence structure of multiple water quality variables. Our assessment system has comparative advantages over cross-coupling water quality models that require an enormous amount of detailed data.
2. Taking the IRSN of ponds around Nanyang station as an example, models based on three classes of vine copulas (C-, D-, and R-vine) were respectively utilized to identify the risk of water quality indicators before and after the operation of the connection project. The sensitivity of five risk indicators including *DO*, *TN*, *TP*, *Chla*, and *NH₃-N* was analyzed, and the results showed that *TN*, *Chla*, and *NH₃-N* should be considered as primary risk factors after the operation of the connection project.
3. By comparing the advantages and prediction accuracy of the C-, D-, and R-vine, the different adaptive circs among them were deduced. The C-vine is applicable to fit a multi-variable with a pivotal element. D-vines might actually be more beneficial when we do not want to assume the existence of a key variable that governs the dependencies. The R-vine has a greater advantage when it comes to dealing with the massive volume and high-dimension data. In addition, using

the R-vine Matrix (RVM) to express the vine copula structures is without doubt a concise and effective method. We will consider R-vine Copula-based method when there are high-dimension water quality variables in future work.

Author Contributions: Conceptualization, C.Z.; methodology, R.Y. (Ruolan Yu) and R.Y. (Rui Yang); formal analysis, R.Y. (Ruolan Yu); writing—Original draft preparation, R.Y. (Ruolan Yu); writing—Review and editing, C.Z. and M.Š.; visualization, N.K.-K. and G.S.; supervision, C.Z.; funding acquisition, C.Z. and G.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (No.2018YFC0407203).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Dou, M.; Shi, Y.X.; Yu, L.; Li, J.Q.; Jia, R.P. Optimization of connecting schemes for urban river networks based on graph theory: A case study of Xuchang section of Qingying River. *J. Hydraul. Eng.* **2020**, *51*, 664–674.
- Dai, J.Y.; Wu, S.Q.; Wu, X.F.; Lv, X.Y.; Sivakumar, B.; Wang, F.F.; Zhang, Y.; Yang, Q.Q.; Gao, A.; Zhao, Y.H.; et al. Impacts of a large river-to-lake water diversion project on lacustrine phytoplankton communities. *J. Hydrol.* **2020**, *587*, 124938. [[CrossRef](#)]
- Welch, E.B.; Barbiero, R.P.; Bouchard, D.; Jones, C.A. Lake trophic state change and constant algal composition following dilution and diversion. *Ecol. Eng.* **1992**, *1*, 173–197. [[CrossRef](#)]
- Everard, M.; Sharma, O.P.; Vishwakarma, V.K.; Khandal, D.; Sahu, Y.K.; Bhatnagar, R.; Singh, J.K.; Kumar, R.; Nawab, A.; Kumar, A.; et al. Assessing the feasibility of integrating ecosystem-based with engineered water resource governance and management for water security in semi-arid landscapes: A case study in the Banas catchment, Rajasthan, India. *Sci. Total Environ.* **2018**, *612*, 1249–1265. [[CrossRef](#)]
- Zhang, X.L.; Zou, R.; Wang, Y.L.; Liu, Y.; Zhao, L.; Zhu, X.; Guo, H.C. Is water age a reliable indicator for evaluating water quality effectiveness of water diversion projects in eutrophic lakes? *J. Hydrol.* **2016**, *542*, 281–291. [[CrossRef](#)]
- Dou, M.; Cui, G.T.; Zuo, Q.T.; Wang, C.; Mao, C.C.; Xu, Y.F. Character analysis of river and lake system interconnection. *China Water Resour.* **2011**, *16*, 17–19.
- Yang, W.; Zhang, L.P.; Zhang, Y.J.; Li, Z.L.; Xiao, Y.; Xia, J. Developing a comprehensive evaluation method for Interconnected River System Network assessment: A case study in Tangxun Lake group. *J. Geogr. Sci.* **2019**, *29*, 389–405. [[CrossRef](#)]
- He, L.; Wang, J.Y.; Li, H.C.; Feng, L.Y.; Wang, Y.X.; Wu, S.Q. Study on the Interconnected River System Network for high-quality development. *China Water Resour.* **2020**, *10*, 11–15.
- Natalia, K.K.; Maria, S.; Zhang, C.; Pronin, M. Zooplankton functional traits as a tool to assess latitudinal variation in the northern-southern temperate European regions during spring and autumn seasons. *Ecol. Indic.* **2020**, *117*, 106629.
- Mansi, T.; Sunil, K.S. Allocation of weights using factor analysis for development of a novel water quality index. *Ecotoxicol. Environ. Saf.* **2019**, *183*, 109510.
- Chen, X.; Wang, Y.H.; Cai, Z.C.; Zhang, M.H.; Ye, C. Response of the nitrogen load and its driving forces in estuarine water to dam construction in Taihu Lake, China. *Environ. Sci. Pollut. Res. Int.* **2020**, in press. [[CrossRef](#)] [[PubMed](#)]
- Liu, S.G.; Lou, S.; Kuang, C.P.; Huang, W.R.; Chen, W.J.; Zhang, J.L.; Zhong, G.H. Water quality assessment by pollution-index method in western Bohai Sea, China. *Mar. Pollut. Bull.* **2011**, *62*, 2220–2229. [[CrossRef](#)] [[PubMed](#)]
- Noori, N.; Kalin, L. Coupling SWAT and ANN models for enhanced daily streamflow prediction. *J. Hydrol.* **2016**, *533*, 141–151. [[CrossRef](#)]
- Noori, N.; Kalin, L.; Lsik, S. Water quality prediction using SWAT-ANN coupled approach. *J. Hydrol.* **2020**, *590*, 125220. [[CrossRef](#)]
- Wu, J.; Tian, X.G.; Tang, Y.; Zhao, Y.J.; Hu, Y.D.; Fang, Z.L. Application of Analytic Hierarchy Process-Grey Target Theory Systematic Model in Comprehensive Evaluation of Water Environmental Quality. *Water Environ. Res.* **2010**, *82*, 633–641. [[CrossRef](#)] [[PubMed](#)]

16. Yang, R.; Wu, S.Q.; Gao, X.P.; Zhang, C. A Vine Copula-based study on identification of multivariate water environmental risk under different connectivity of rivers and lakes. *J. Hydraul. Eng.* **2020**, *51*, 606–616.
17. Schepsmeier, U. Efficient information based goodness-of-fit tests for vine copula models with fixed margins: A comprehensive review. *J. Multivar. Anal.* **2015**, *138*, 34–52. [[CrossRef](#)]
18. Niu, J.Y.; Wu, Z.N.; Feng, P. Combined risk assessment method of water environment system. *Syst. Engineering—Theory Pract.* **2012**, *32*, 2097–2103.
19. Renard, B.; Lang, M. Use of a Gaussian copula for multivariate extreme value analysis: Some case studies in hydrology. *Adv. Water Resour.* **2007**, *30*, 897–912. [[CrossRef](#)]
20. Huang, K.D.; Ye, L.; Chen, L.; Wang, Q.S.; Dai, L.; Zhou, J.Z.; Singh, V.P.; Huang, M.T.; Zhang, J.H. Risk analysis of flood control reservoir operation considering multiple uncertainties. *J. Hydrol.* **2018**, *565*, 672–684. [[CrossRef](#)]
21. Gao, C.; Xu, Y.P.; Zhu, Q.; Bai, Z.X.; Liu, L. Stochastic generation of daily rainfall events: A single-site rainfall model with Copula-based joint simulation of rainfall characteristics and classification and simulation of rainfall patterns. *J. Hydrol.* **2018**, *564*, 41–58. [[CrossRef](#)]
22. Li, H.S.; Wang, D.; Singh, V.P.; Wang, Y.K.; Wu, J.F.; Wu, J.C.; Liu, J.F.; Zou, Y.; He, R.M.; Zhang, J.Y. Non-stationary frequency analysis of annual extreme rainfall volume and intensity using Archimedean copulas: A case study in eastern China. *J. Hydrol.* **2019**, *571*, 114–131. [[CrossRef](#)]
23. Tosunoglu, F.; Onof, C. Joint modelling of drought characteristics derived from historical and synthetic rainfalls: Application of Generalized Linear Models and Copulas. *J. Hydrol. Reg. Stud.* **2017**, *14*, 167–181. [[CrossRef](#)]
24. Nabaei, S.; Sharafati, A.; Yaseen, Z.M.; Shahid, S. Copula based assessment of meteorological drought characteristics: Regional investigation of Iran. *Agric. For. Meteorol.* **2019**, *276*, 10761. [[CrossRef](#)]
25. Wu, P.Y.; You, G.J.; Chan, M.H. Drought analysis framework based on copula and Poisson process with nonstationarity. *J. Hydrol.* **2020**, *588*, 125022. [[CrossRef](#)]
26. Wang, W.S.; Li, Y.Q. Copula Assessment Method and Its Application for Eutrophication of Lake Water Quality Assessment. *Adv. Eng. Sci.* **2011**, *43*, 39–42.
27. Zhang, Y.; Dou, M.; Li, G.Q. The analysis of joint risk probability of eutrophication based on copula function. *Acta Sci. Circumstantiae* **2018**, *38*, 4204–4213.
28. Zhang, X.; Ran, Q.X.; Xia, J.; Song, X.Y. Jointed distribution function of water quality and water quantity based on Copula. *J. Hydraul. Eng.* **2011**, *42*, 483–489.
29. Xu, C.; Feng, M.Q. Joint risk of water quantity and quality in water sources of water diversion project. *J. Northwest A&F Univ.* **2016**, *44*, 228–234.
30. Daneshkhah, A.; Remesan, R.; Chatrabgoun, O.; Holman, L.P. Probabilistic modeling of flood characterizations with parametric and minimum information pair-copula model. *J. Hydrol.* **2016**, *540*, 469–487. [[CrossRef](#)]
31. Joe, H.; Li, H.; Nikoloulopoulos, A.K. Tail dependence functions and vine copulas. *J. Multivariate Anal.* **2010**, *101*, 252–270. [[CrossRef](#)]
32. Bedford, T.; Cooke, R.M. Probability density decomposition for conditionally dependent random variables modeled by vines. *Ann. Math. Artif. Intell.* **2001**, *32*, 245–268. [[CrossRef](#)]
33. Bedford, T.; Cooke, R.M. Vines—A new graphical model for dependent random variables. *Ann. Stat.* **2002**, *30*, 1031–1068. [[CrossRef](#)]
34. Aas, K.; Czado, C.; Frigessi, A.; Bakken, H. Pair-copula constructions of multiple dependence. *Insur. Math. Econ.* **2009**, *44*, 182–198. [[CrossRef](#)]
35. Cekin, E.S.; Pradhan, A.K.; Tiwari, A.K.; Gupta, R. Measuring co-dependencies of economic policy uncertainty in Latin American countries using vine copulas. *Q. Rev. Econ. Financ.* **2020**, *76*, 207–217. [[CrossRef](#)]
36. Montes, R.; Heredia, E. Multivariate environmental contours using C-vine copulas. *Ocean Eng.* **2016**, *118*, 68–82. [[CrossRef](#)]
37. Arya, F.K.; Zhang, L. Copula-based markov process for forecasting and analyzing risk of water quality time series. *J. Hydrol. Eng.* **2017**, *22*, 1–12. [[CrossRef](#)]
38. Ni, L.L.; Wang, D.; Wu, J.F.; Wang, Y.K.; Tao, Y.W.; Zhang, J.Y.; Liu, J.F.; Xie, F. Vine copula selection using mutual information for hydrological dependence modeling. *Environ. Res.* **2020**, *186*, 109604. [[CrossRef](#)]
39. Dissmann, J.; Brechmann, E.C.; Czado, C.; Kurowicka, D. Selecting and estimating regular vine copulae and application to financial return. *Comput. Stat. Data Anal.* **2013**, *59*, 52–69. [[CrossRef](#)]
40. Tung, Y.K.; Mays, L.W. Risk models for flood levee design. *Water Resour. Res.* **1981**, *17*, 833–841. [[CrossRef](#)]

41. Sukcharoen, K.; Leatham, D.J. Hedging downside risk of oil refineries: A vine copula approach. *Energy Econ.* **2017**, *66*, 493–507. [[CrossRef](#)]
42. Martey, E.N.; Okine, N.A. Analysis of train derailment severity using vine copula quantile regression modeling. *Transp. Res. Part C Emerg. Technol.* **2019**, *105*, 485–503. [[CrossRef](#)]
43. Czado, C.; Schepsmeier, U.; Min, A. Maximum likelihood estimation of mixed C-vines with application to exchange rates. *Stat. Model.* **2012**, *12*, 229–255. [[CrossRef](#)]
44. Brechmann, E.C. Truncated and Simplified Regular Vines and Their Applications. Master's Thesis, Technische Universitaet Muenchen, Munich, Germany, 2010.
45. Yu, W.H.; Yang, K.; Wei, Y.; Lei, L.K. Measuring Value-at-Risk and Expected Shortfall of crude oil portfolio using extreme value theory and vine copula. *Phys. A Stat. Mech. Its Appl.* **2018**, *490*, 1423–1433. [[CrossRef](#)]
46. Xu, H.; Sang, G.Q.; Yang, L.Y.; Yan, F.J.; Liu, Y.C. Temporal and spatial distribution characteristics of water quality in Nansi Lake in recent ten years. *Trans. Oceanol. Limnol.* **2019**, *2*, 47–52.
47. Supper, H.; Irresberger, F.; Wei, G. A comparison of tail dependence estimators. *Eur. J. Oper. Res.* **2020**, *284*, 728–742. [[CrossRef](#)]
48. Yao, C.Z.; Sun, B.Y. The study on the tail dependence structure between the economic policy uncertainty and several financial markets. *N. Am. J. Econ. Financ.* **2018**, *45*, 245–265. [[CrossRef](#)]
49. Wang, X.; Zang, N.; Liang, P.Y.; Cai, Y.P.; Li, C.H.; Yang, Z.F. Identifying priority management intervals of discharge and TN/TP concentration with copula analysis for Miyun Reservoir inflows, North China. *Sci. Total Environ.* **2017**, *609*, 1258–1269. [[CrossRef](#)]
50. Genest, C.; Favre, A.C. Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resour. Res.* **2007**, *43*, 5275. [[CrossRef](#)]
51. Karamouz, M.; Taheriyoun, M.; Seyedabadi, M.; Nazif, S. Uncertainty based analysis of the impact of watershed phosphorus load on reservoir phosphorus concentration. *J. Hydrol.* **2015**, *521*, 533–542. [[CrossRef](#)]
52. Aissia, M.A.; Chebana, F.; Ouarda, T.B. Multivariate missing data in hydrology- Review and applications. *Adv. Water Resour.* **2017**, *110*, 299–309. [[CrossRef](#)]
53. Vuong, Q.H. Ratio tests for model selection and non-nested hypotheses. *Econometrica* **1989**, *57*, 307–333. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).