Joint improvement of river water quality indicators based on a multivariate joint probability distribution of the discharge and water quality

Yang Liu, Shengle Cao, Xi Zhang, Fuzhen Li and Xitong Li

ABSTRACT

Based on the multivariate joint probability distribution of the discharge and water quality indicators, this paper analysed the occurrence probabilities and improvement probabilities of combinations of water quality indicators under different discharge conditions and then presented a method for calculating the optimal discharge to seek a balance between the discharge dispatch and water quality improvement. The method was used to construct the relationship curve between the discharge and joint improvement probability used by a copula function and then calculate the critical point on the curve. The proposed method was applied to the Yi River Basin above Gegou Station with data composed of the discharge and main pollution indicators (NH_3-N and COD_{Mn}) from 1982 to 2015. The results showed that the trivariate joint probability distribution can more reasonably reveal the statistical characteristics of different combinations of discharge and water quality indicators. Furthermore, the optimal discharges and the corresponding improvement probabilities that improved NH_3-N and COD_{Mn} to different grades were calculated. The calculation method took the interdependence of multiple water quality indicators into account, thereby providing a more reasonable method for using discharge dispatch data to improve the river water quality.

Key words | copula, joint improvement, multiple indicators, optimal discharge, water quality

Yang Liu
Shengle Cao (corresponding author)
Xi Zhang
Fuzhen Li
Xitong Li
School of Civil Engineering,
Shandong University,
Jinan 250061,
China
E-mail: *cao_shengle@163.com*

INTRODUCTION

With the acceleration of urbanization in China, the deterioration of the quality of river water has become increasingly serious. One of the main water pollution control technologies used to improve the water quality is a reasonable discharge dispatch system (Jang *et al.* 2012; Qian *et al.* 2013; Shokri *et al.* 2014). However, the determination of the optimal discharge and the quantitative relationship between the discharge dispatch and the improvement in the water quality need to be further explored, especially with regard to the joint effects of discharge and water quality indicators.

Most of the studies conducted on the relationships between the water quantity and water quality follow two main approaches: mechanism models and non-mechanism models. The former takes into account the relationship of the mutual transformation of pollution factors and the coupling of hydrodynamic modules to simulate and predict concentration changes in pollution factors (Cox 2003; Wang *et al.* 2015; Yu *et al.* 2016). Nevertheless, many interdependent factors affect both the discharge and the water quality; consequently, the calibration of the parameters and the construction of a model are relatively complex. In contrast, non-mechanism models avoid the above-mentioned problems to some extent (Sun *et al.* 2011; Phan *et al.* 2016). Commonly used non-mechanism models include nonlinear algorithms (Liu *et al.* 2015), statistical models (Bonansea *et al.* 2015; Liu *et al.* 2016), Bayesian models and artificial neural network (ANN) models (Chitsazan *et al.*

doi: 10.2166/nh.2018.196

2015). However, most non-mechanism models require that the variables obey the same marginal distribution or be transformed to the same distribution; this not only leads to the loss of information regarding the original data but also does not conform to the fact that variables such as the discharge and water quality obey different marginal distributions. Thus, most of the previous studies have some limitations on the occurrence probability of water quality combination events (Faruk 2010; Zhu *et al.* 2012). A copula can connect different univariate marginal distributions without changing the dependencies of variables (Zegpi & Fernandez 2010; Dai *et al.* 2014), i.e. it can construct the joint probability distribution of any univariate marginal distribution while retaining the information included in the initial data during the transformation process. Due to these advantages, copulas are widely used in multivariate hydrological analyses and calculations (Rauf & Zeephongsekul 2014). Their application is also becoming increasingly common in the risk analysis of floods (Kao & Chang 2012; Bezak *et al.* 2014; Li & Zheng 2016; Balistrocchi & Bacchi 2017) and the analysis of drought characteristics in watersheds (Hao & AghaKouchak 2013; Yusof *et al.* 2013; Bazrafshan *et al.* 2015; Li *et al.* 2017). In recent years, copulas have also been used to analyse the joint probability of the water quantity and water quality (Zhang *et al.* 2011; Wu *et al.* 2015). In summary, there have been few applications of copulas to the joint analysis of discharge and water quality; those few studies mainly focused on the bivariate joint distribution of discharge and individual indicators. Moreover, investigations of the multivariable joint distribution are relatively rare, and those focused on determining the optimal discharge for improving multiple water quality indicators are even scarcer.

Therefore, this paper presents novel research insomuch that the copula function is not only used to establish the multi-dimensional joint probability distribution of the water quantity and water quality but is also combined with the marginal probability distribution function to deduce the relationship curve between the discharge and the joint probability. Furthermore, the critical point of the improvement in the water quality is analysed from a statistical perspective, thereby extending the application scope of the copula function. The objective of this research is to construct a multivariate joint probability distribution of the discharge and multiple water quality indicators based on copula functions and then to use the joint probability distribution to construct a relationship curve between the discharge and joint improvement probability to analyse the comprehensive improvement of the river water environment under different discharge conditions. Moreover, the optimal discharge is suggested to seek a balance between the discharge dispatch and water quality improvement, providing a more scientific and reasonable basis for river water quality management.

METHODS

Multivariate joint probability distribution

The theory of the copula was proposed by Sklar (1959), who stated that any d-dimensional joint cumulative distributions defined in R can be connected by using a copula function. The Archimedean copula, which is commonly used in hydrology (Schumann 2017), can be generally defined as follows:

$$C(u_1, u_2, \ldots, u_d) = \varphi^{-1}[\varphi(u_1) + \varphi(u_2) + \ldots + \varphi(u_d)] \quad (1)$$

where $\varphi(\sim)$ is the generation function, φ^{-1} is the inverse function of φ, and u_d is the marginal distribution of a d-dimensional random variable.

Common symmetric Archimedean copulas include the Clayton, Gumbel–Hougaard (G-H), Frank and Ali–Mikhail–Haq (AMH) varieties. Their specific function forms and parameter ranges are shown in Table 1.

The establishment of a joint distribution based on a copula is divided into the following two steps:

1. Selection of the marginal distributions of hydrological variables. Using the maximum likelihood method, the parameters of marginal distributions are estimated. Then, the proper marginal distributions are chosen using two goodness-of-fit tests: the Kolmogorov–Smirnov (K-S) test and the root mean squared error (RMSE) test.
2. Determination of the parameter values θ of the multivariate joint distribution based on copula functions. The parameters of the multivariate joint distributions are

Table 1 | Forms and parameter ranges of four symmetric Archimedean copula functions

	$C(u_1, u_2, \ldots, u_d)$	Ranges of θ
Clayton	$\left[\left(\sum_{i=1}^{d} u_i^{-\theta}\right) - d + 1\right]^{-1/\theta}$	$\theta > 0$
G-H	$e^{-\left[\sum_{i=1}^{d}(-\ln u_i)^{\theta}\right]^{1/\theta}}$	$\theta \geq 1$
Frank	$-\frac{1}{\theta}\ln\{1 + [\prod_{i=1}^{d}(e^{-\theta u_i} - 1)]/(e^{-\theta} - 1)^{d-1}\}$	$\theta \in R \cap \theta \neq 0$
AMH	$[(1-\theta)\prod_{i=1}^{d} u_i]/\{\prod_{i=1}^{d}[1-\theta(1-u_i)] - \theta\prod_{i=1}^{d} u_i\}$	$0 < \theta < 1$

estimated by the two-stage estimation method and the inference functions for margins (IFM) (Joe 2005).

Calculation of the optimal discharge

Construction of the relationship curve between the discharge and joint probability

To improve the river water quality gradually, the corresponding recovery criteria of various water quality indicators are constructed according to the different conditions of the river water quality. The management targets of different water quality indicators, which can be obtained from the 'Environmental Quality Standards for Surface Water' (GB3838-2002), are denoted by T_{mj}, where m is the number of water quality indicators ($m = 1, 2, \ldots, n$), and j represents the different management targets ($j = 1, 2, \ldots, 5$) In addition, the marginal probability corresponding to T_{mj} can be obtained and recorded as u_{mj}. Then, the marginal probability of the discharge sequence Q_i, where i is the serial number of the discharge sequence ($i = 1, 2, \ldots n$), is arranged in descending order and recorded as u'_{qi}. The new $m+1$ dimensional matrix A_j is constructed by u'_{qi} and u_{mj} as shown in Equation (2):

$$A_j = \begin{bmatrix} u'_{q1} & u_{1j}\cdots & u_{mj} \\ u'_{q2} & u_{1j}\cdots & u_{mj} \\ \vdots & \vdots & \vdots \\ u'_{qn} & u_{1j}\cdots & u_{mj} \end{bmatrix} \tag{2}$$

Combined with the parameter θ of the selected copula function, the probability C'_{ij} can be obtained. Then, the relationship curve $Q_i \sim C'_{ij}(Q, T_1, T_2, \cdots, T_m)$ can be drawn and recorded as $Q \sim C'_j$, which denotes changes in the probability with an increase in the discharge when T_m reaches the jth grade.

Determination of the critical point on the relationship curve

The critical point on the curve $Q \sim C'_j$ refers to a catastrophe point of the change rate of C'_{ij} with an increase in Q_i. To calculate the critical point on the curve, whether a critical point exists along the curve must be clarified. It is difficult to deduce the relationship between Q_i and C'_{ij} directly because it is relatively complex. Therefore, the relationships between C'_{ij} and u_{qi} and between u_{qi} and Q_i are deduced individually to clarify the relationship between Q_i and C'_{ij}.

1. Relationship between u_{qi} and C'_{ij}. The Frank copula, which is taken as an example here for the sake of explanation, is expressed as:

$$C'_{ij} = -\frac{1}{\theta} \ln\left[\frac{(e^{-\theta u_{qi}} - 1)(e^{-\theta u_{1j}} - 1)(e^{-\theta u_{2j}} - 1)\cdots(e^{-\theta u_{mj}} - 1)}{(e^{-\theta} - 1)^m} + 1\right] \tag{3}$$

Following the simplification of the copula:

$$\begin{aligned} & e^{-\theta C'_{ij}} = k(e^{-\theta u_{qi}} - 1) + 1 \\ & u_{qi} = C'_{ij} + \frac{1}{\theta}\ln k \end{aligned} \tag{4}$$

where C'_{ij} is the occurrence probability of the jth grade under Q_i conditions, $k = ((e^{-\theta u_{1j}} - 1)(e^{-\theta u_{2j}} - 1)\cdots(e^{-\theta u_{mj}} - 1))/(e^{-\theta} - 1)^m$, u_{mj} is the marginal probability of the mth indicator at the jth grade, and θ is the parameter of the copula function.

Equation (4) constitutes a linear relationship between u_{qi} and C'_{ij}. Similarly, the relationships between u_{qi} and C'_{ij} for the other three functions in Table 1 can be deduced to be linear relationships.

2. Relationship between Q_i and u_{qi}. According to the most commonly used hydrological distribution function in China (Luo & Song 2014), the probability density function (PDF) of Q_i can generally be selected from the lognormal distribution, gamma distribution or P-III distribution. Furthermore, u_{qi} is a cumulative distribution function (CDF) of Q_i. Three types of CDFs are shown in Table 2.

Taking the lognormal distribution as an example, the monotonicity of the integrand is analysed first. It can be proven that t^{-1} and $e^{(-(\ln t-\mu)^2)/2\sigma^2}$ are monotonically increasing functions when $t \geq 0$ and that their values are all greater than 0, and thus, the integrand $f(t)$ obtained by multiplying these two formulas is also an increasing function. The second derivative $u''_{qi} = f'(t) \geq 0$ can be obtained, following which u_{qi} can be proven to be a convex function. Similarly, the relationship between Q_i and u_{qi} is also a convex function when Q_i is subjected to a gamma or P-III distribution.

In summary, the relationship curve of $Q \sim C'_j$ is a monotonically increasing convex function. Therefore, there exists a point where the change in the growth rate along the curve is a maximum, i.e. a critical point. When the discharge exceeds this critical point, the combined effect of multiple water quality indicators will not obviously improve with an increase in the discharge. Therefore, the discharge corresponding to the critical point can be regarded as the balance point between an improvement in the water quality and the discharge dispatch; this critical discharge is defined as the optimal discharge for the joint improvement of multiple water quality indicators. The relationship curve of $Q \sim C'_j$ can be fitted by MATLAB, and the critical point can be calculated by a curvature method, slope method or another approach (Gippel & Stewardson 1998). The formulas of the curvature method and slope method are given in Equations (5) and (6), respectively:

$$|\rho| = \frac{\left|\partial C'/\partial u_q^2\right|}{\left[1 + (\partial C'/\partial u_q)^2\right]^{3/2}} \tag{5}$$

$$|k| = \frac{\partial C'}{\partial u_q} \tag{6}$$

where ρ is the curvature of the curve, and k is the slope of the curve.

STUDY AREA

The Yi River is one of the most important rivers in Shandong Province. Gegou Station, which is located at the coordinates 110°28′ east longitude and 35°21′ north latitude in Gegou Village, Linyi City, China, is a controlled hydrological station located in the upper reaches of the Yi River.

Table 2 | Three CDFs and ranges of the parameters

	CDFs	Ranges of parameters
Lognormal	$u_{qi} = \frac{1}{\sigma\sqrt{2\pi}}\int_0^{Q_i} t^{-1}\, e^{(-(\ln t-\mu)^2)/2\sigma^2}dt$	$Q_i \geq 0, \mu > 0$
Gamma	$u_{qi} = \frac{1}{b^a\Gamma(a)}\int_0^{Q_i} t^{a-1}e^{(-t)/b}dt$	$Q_i \geq 0, a > 0, b > 0$
P-III	$u_{qi} = \frac{1}{b^a\Gamma(a)}\int_0^{Q_i} (t-a_0)^{a-1}e^{(-(t-a_0))/b}dt$	$Q_i \geq a_0, a > 0,$ $b > 0, 0 < a_0 < t$

This study mainly examines the Yi River Basin above Gegou Station. The Yi River extends 110 km above Gegou Station with a watershed area of 5,500 km^2, and the terrain is dominantly hilly. In recent years, along with increasing population growth and economic development, non-point source pollution in the basin has polluted the water quality of the upper reaches of the Yi River and has become increasingly serious. In a few months, the water quality has exceeded a grade of V; as a consequence, the aquatic ecosystem in the basin has been severely damaged. Three reservoirs have been built within the basin, namely, the Tianshan Reservoir, Bashan Reservoir and Andi Reservoir. Thus, it is necessary to improve the water quality of the river by employing reasonable discharge dispatch through the use of water storage within these three reservoirs. The drainage map of the Yi River Basin above Gegou Station is shown in Figure 1.

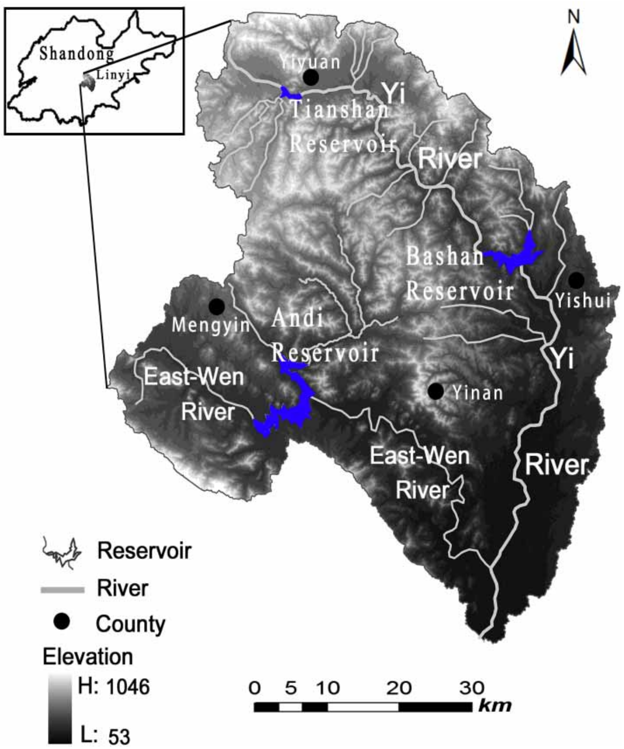

Figure 1 | Drainage map of the Yi River Basin above Gegou Station.

The studied data were collected from Gegou Station, which has monitored the discharge and water levels since 1952 and provides complete water quality data records beginning in 1970. To ensure the validity and representativeness of the data, this paper chose the discharge and water quality data of Gegou Station from 1982 to 2014 (water quality data are recorded every other month) for a total of 153 months. This paper collected a total of six water quality indicators, namely, the chemical oxygen demand (COD), permanganate index (COD_{Mn}), biochemical oxygen demand (BOD), ammonia nitrogen (NH_3-N), total phosphorus (TP) and total nitrogen (TN), from Gegou Station. The ratio of each measured concentration to the grade III value of each water quality indicator is shown in Figure 2. The grade III values of the six water quality indicators are obtained from the 'Environmental Quality Standards for Surface Water' (GB3838-2002). Figure 2 shows that the excess ratios of the BOD, NH_3-N, COD_{Mn} and TN concentrations are

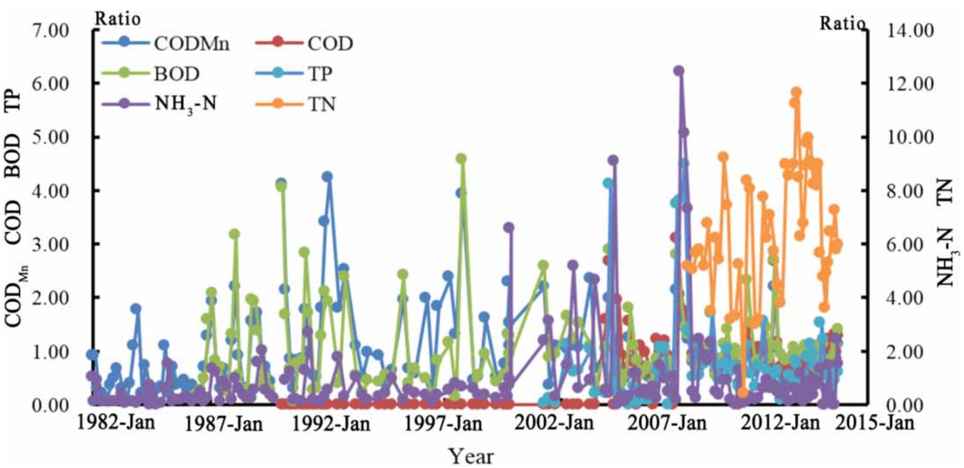

Figure 2 | Ratio of each water quality indicator to the grade III value of the 'Environmental Quality Standards for Surface Water'.

significantly greater; however, the length and continuity of the data of the BOD and TN are not as good as those of the data of the NH_3-N and COD_{Mn}. Therefore, this paper selected the NH_3-N and COD_{Mn} as the research indicators.

RESULTS AND DISCUSSION

Multivariate joint distribution of Q, NH_3-N and COD_{Mn}

Establishment of the marginal distribution and joint probability distribution

1. Establishment of the marginal distribution. The parameters of the normal, lognormal, gamma and P-III distributions of Q, NH_3-N and COD_{Mn} are estimated by the maximum likelihood method, as shown in Table 3. The results the of K-S and RMSE tests are given in Table 4, which shows that the lognormal distribution is the most appropriate for the Q, NH_3-N and COD_{Mn} indicators at Gegou Station on the Yi River.
2. Establishment of the joint probability distribution. Combined with the sample empirical frequency, the K-S and RMSE results are calculated, as shown in Table 5. The results show that the trivariate joint distribution of the set (Q, NH_3-N, COD_{Mn}) is in accordance with the Clayton copula function, and the bivariate joint distributions of the set (Q, NH_3-N) and the set (Q, COD_{Mn}) are in accordance with the G-H copula function and AMH copula function, respectively.

Table 3 | Statistical parameters of the marginal distributions of Q, NH_3-N and COD_{Mn}

	Q	NH_3-N	COD_{Mn}
Normal	$\mu = 24.774$, $\sigma = 38.191$	$\mu = 0.948$, $\sigma = 1.171$	$\mu = 5.473$, $\sigma = 3.799$
Lognormal	$\mu = 2.382$, $\sigma = 1.262$	$\mu = -0.760$, $\sigma = 1.119$	$\mu = 1.487$, $\sigma = 0.672$
Gamma	$\mu = 0.726$, $\sigma = 34.122$	$\mu = 0.835$, $\sigma = 1.136$	$\mu = 2.499$, $\sigma = 2.189$
P-III	$\alpha = 0.299$, $\beta = 79.780$, $a_0 = 0.947$	$\alpha = 0.264$, $\beta = 3.603$, $a_0 = -0.002$	$\alpha = 1.291$, $\beta = 3.514$, $a_0 = 0.932$

In addition, sample empirical distributions are compared with the theoretical distributions of the sets (Q, NH_3-N), (Q, COD_{Mn}) and (Q, NH_3-N, COD_{Mn}), as shown in Figures 3 and 4, for which the correlation coefficients R^2 are all relatively high. This result illustrates that the selected copulas can appropriately describe the multivariate probability distributions of Q, NH_3-N and COD_{Mn}.

Analysis of the joint probability distribution of the discharge and water quality

With the established multivariate joint probability distribution of Q, NH_3-N and COD_{Mn}, their corresponding contour plots and contour surfaces are shown in Figures 5 and 6, respectively; note that the coordinate axes of Figure 6 use a logarithmic coordinate. Moreover, some typical joint probabilities of Q, NH_3-N and COD_{Mn} are given in Table 5.

Figure 5 shows the contour plots of the sets (Q, NH_3-N) and (Q, COD_{Mn}). When the value of the bivariate joint probability distribution is between 0.1 and 0.4, the contour plot is more intensive than that between 0.6 and 0.9. This reflects that the degrees of change in the joint probabilities for both NH_3-N and COD_{Mn} present a decrease with an increase in the discharge.

Furthermore, as seen from Figure 6, when the value of the trivariate joint probability distribution is between 0.1 and 0.2 or

Table 4 | Goodness-of-fit test results for the marginal distribution determination

	Normal		Lognormal		Gamma		P-III	
	K-S	RMSE	K-S	RMSE	K-S	RMSE	K-S	RMSE
Q	0.2828	0.1669	0.0617	**0.0287**	0.1453	0.0831	0.2426	0.1216
NH_3-N	0.2900	0.1776	0.0427	**0.0177**	0.1344	0.0751	0.3751	0.1945
COD_{Mn}	0.1565	0.0883	0.0479	**0.0247**	0.0717	0.0352	0.0771	0.0365

Note: The numbers in bold denote the minimum RMSE, and the K-S value is 0.10995 at a 5% significance level in this study.

Table 5 | The parameter values and goodness-of-fit tests of the copulas

	Q, NH_3-N			Q, COD_{Mn}			Q, NH_3-N, COD_{Mn}		
	θ	K-S	RMSE	θ	K-S	RMSE	θ	K-S	RMSE
Clayton	–	–	–	–	–	–	0.8861	0.0865	0.0369
G-H	**0.8919**	**0.0551**	**0.0215**	0.6239	0.0574	0.0220	0.9219	0.1658	0.0862
Frank	1.6756	0.0832	0.0310	1.7545	0.0689	0.0236	0.3592	0.1537	0.7980
A-M-H	0.8085	0.0702	0.0218	**1.3839**	**0.0488**	**0.0214**	0.4901	0.1235	0.8341

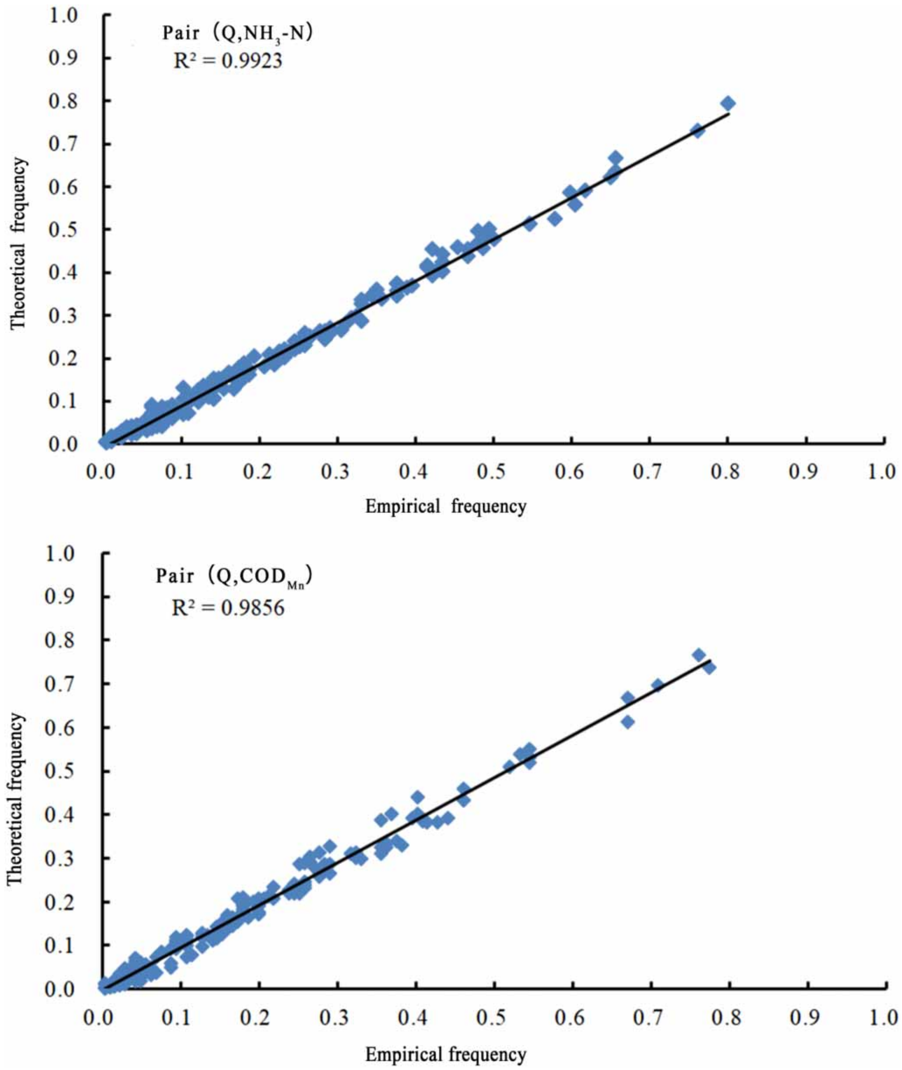

Figure 3 | Comparison between the empirical frequency and theoretical frequency of the sets (Q, NH_3-N) and (Q, COD_{Mn}).

between 0.7 and 0.9, the contour surface has a larger spacing, and Q, NH_3-N, COD_{Mn} each change more than that spacing value between 0.3 and 0.6, indicating that a trivariate joint probability from 30 to 60% is the most sensitive to the discharge. Combining Tables 6–8, the degree of increase in the probability does not always increase as the discharge increases.

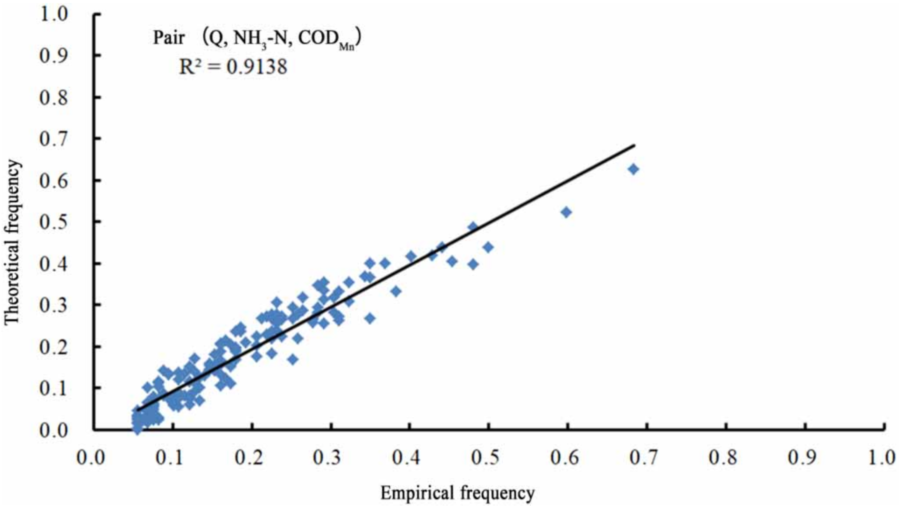

Figure 4 | Comparison between the empirical frequency and theoretical frequency of the set (Q, NH_3-N, COD_{Mn}).

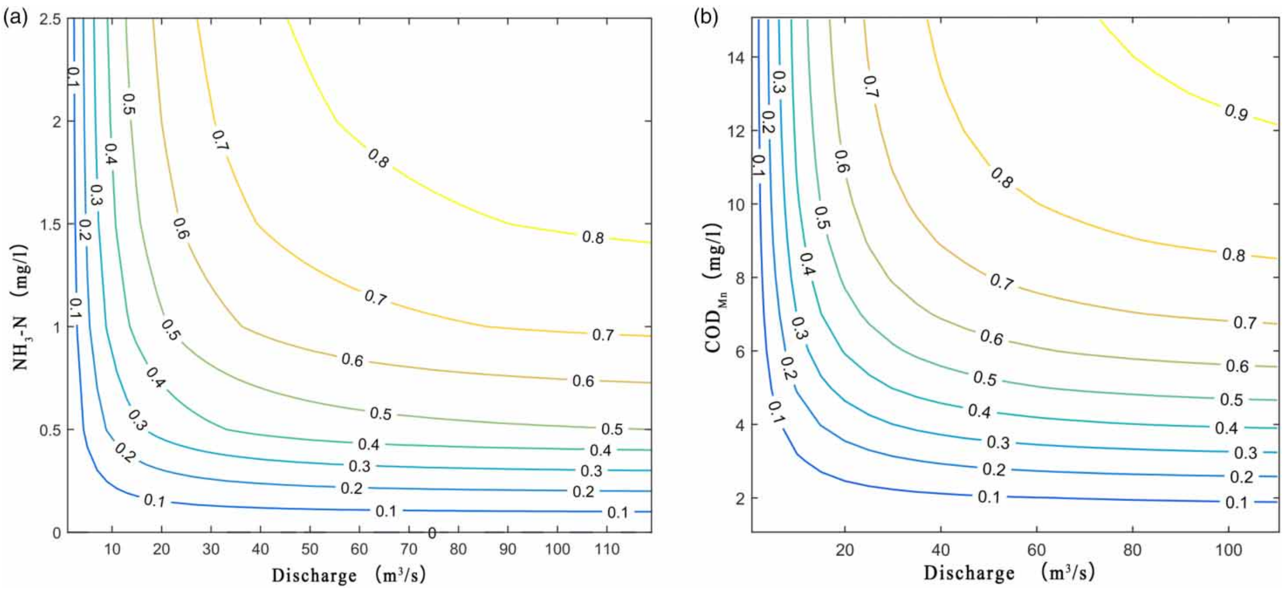

Figure 5 | Contour plots of the bivariate joint probability distributions: (a) contour plot of Q and NH_3-N; (b) contour plot of Q and COD_{Mn}.

Application analysis of the joint improvement of the water quality based on the relationship curve

Construction of the relationship curve among Q, NH_3-N and COD_{Mn}

Based on the improvement probability of an individual water quality indicator, the $Q \sim C_j'$ curves of recovering NH_3-N and COD_{Mn} to grades of III, IV and V are constructed separately, and the fitting function of each curve is computed by MATLAB, as shown in Figure 7. For the joint improvement probability of multivariate water quality indicators, the $Q \sim C_j'$ curve of NH_3-N and COD_{Mn} simultaneously recovered to grade III is denoted as (III, III), that of NH_3-N recovering to grade III while COD_{Mn} recovers to grade IV is denoted as (III, IV), and that of NH_3-N recovering to grade IV while COD_{Mn} recovers to grade III

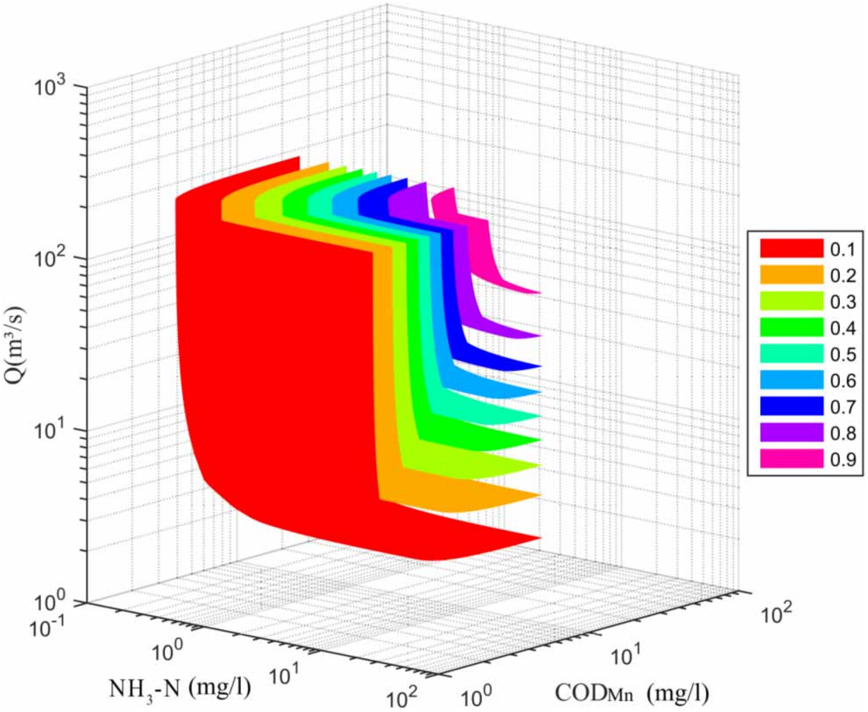

Figure 6 | Contour surface of the trivariate joint probability distribution of Q, NH_3-N and COD_{Mn}.

Table 6 | Joint probability of Q and NH_3-N

	Discharge (m³/s)				
NH_3-N (mg/L)	**10**	**20**	**40**	**60**	**80**
0.5	0.2203	0.3302	0.4223	0.4608	0.48093
1.0	0.3302	0.4873	0.6152	0.6676	0.69470
1.5	0.3826	0.5606	0.7031	0.7607	0.79016
2.0	0.4116	0.6006	0.7504	0.8103	0.84070

Table 7 | Joint probability of Q and COD_{Mn}

	Discharge (m³/s)				
COD_{Mn} (mg/L)	**10**	**20**	**40**	**60**	**80**
4	0.1488	0.2436	0.3355	0.3767	0.39836
6	0.2594	0.4063	0.6947	0.0000	0.62116
10	0.3898	0.5812	0.7372	0.7992	0.83020
15	0.4473	0.6532	0.8147	0.8774	0.90846

is denoted as (IV, III). These curves are each constructed individually, thereby representing equal and non-equal weights for the recovery of the set (NH_3-N, COD_{Mn}), and the fitting function of each curve is computed by MATLAB, as shown in Figure 8.

Table 8 | Joint probability of Q, NH_3-N and COD_{Mn}

		Q (m³/s)				
NH_3-N (mg/L)	**COD_{Mn} (mg/L)**	**10**	**20**	**40**	**60**	**80**
0.5	4	0.2233	0.2657	0.2898	0.2978	0.3014
	6	0.2764	0.3423	0.3821	0.3955	0.4018
	10	0.3104	0.3944	0.4471	0.4653	0.4738
	15	0.3206	0.4106	0.4677	0.4875	0.4968
1.0	4	0.2609	0.3192	0.3539	0.3655	0.3709
	6	0.3345	0.4329	0.4965	0.5186	0.5291
	10	0.3843	0.5168	0.6077	0.6405	0.6562
	15	0.3997	0.5440	0.6450	0.6817	0.6994
1.5	4	0.2735	0.3378	0.3766	0.3897	0.3958
	6	0.3549	0.4665	0.5404	0.5665	0.5790
	10	0.4109	0.5644	0.6732	0.7131	0.7324
	15	0.4284	0.5966	0.7186	0.7639	0.7858
2.0	4	0.2793	0.3466	0.3874	0.4012	0.4076
	6	0.3645	0.4828	0.5620	0.5902	0.6036
	10	0.4237	0.5878	0.7061	0.7499	0.7711
	15	0.4423	0.6227	0.7559	0.8059	0.8302

Analysis of the improvement probability and the optimal discharge

According to the bivariate and trivariate joint probability $Q \sim C_j'$ curves, the typical improvement probability values and the corresponding discharges of each individual water quality indicator and of the multiple water quality indicators are given in Tables 9–11.

Taking the $Q \sim C_j'$ function (Equation (9)) with the curvature method and slope method (Equations (5) and (6), respectively), the optimal discharge can be obtained. Accordingly, the optimal discharges under the conditions in which the set (NH_3-N, COD_{Mn}) is improved to grades of (III, III), (IV, IV), (V, V), (III, IV), (III, V), (IV, V) and (IV, III) are separately shown in Table 12.

Tables 11 and 12 give the joint improvement probabilities of the water quality and its corresponding discharge under various recovery scenarios of the water quality, providing a variety of choices for managers to make water dispatch schemes. Those managers can therefore choose an appropriate discharge according to the storage capacity of a reservoir and the pollution of a river.

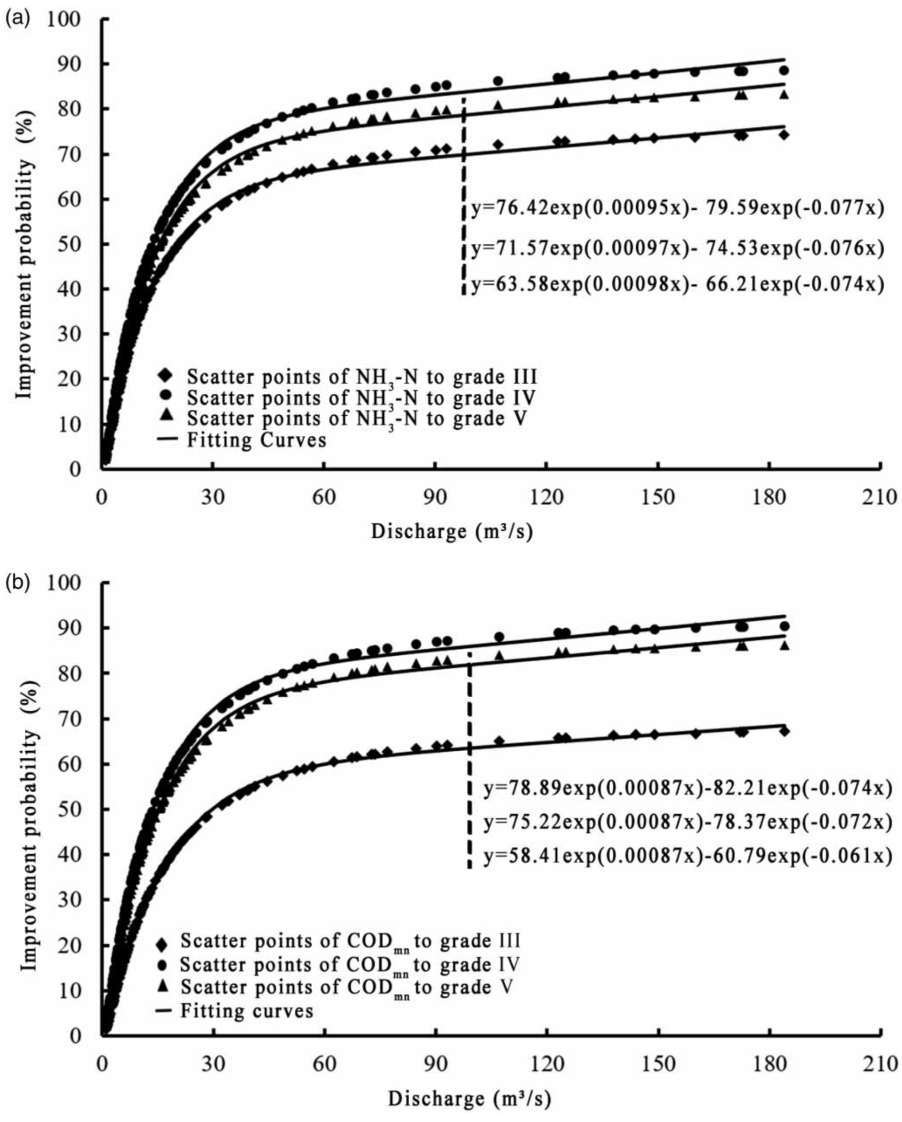

Figure 7 | $Q \sim C_j'$ curves of the sets (Q, NH_3-N) and (Q, COD_{Mn}) and the corresponding fitting functions.

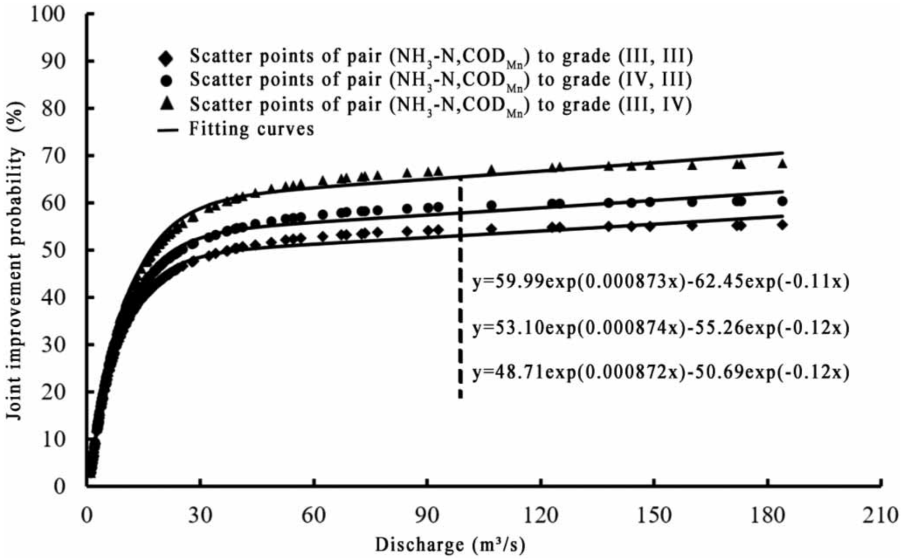

Figure 8 | $Q \sim C_j'$ curves of the set (Q, NH_3-N and COD_{Mn}) and the corresponding fitting function.

Table 9 | Improvement probability and corresponding discharge of Q and NH_3-N

NH_3-N	40.0%	45.0%	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%
III	13.50	16.30	20.00	24.95	33.15	50.50	98.80	175.00	–
IV	11.00	13.05	15.65	18.80	22.85	28.4	38.15	59.60	115.00
V	9.77	11.65	13.70	16.00	19.30	23.20	28.60	38.10	59.60

Table 10 | Improvement probability and corresponding discharge of Q and COD_{Mn}

COD_{Mn}	40.0%	45.0%	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%
III	19.00	23.40	30.00	39.40	59.70	125.00	–	–	–
IV	10.95	13.15	14.96	18.25	12.50	26.65	33.10	46.65	76.60
V	9.78	11.65	13.60	16.10	18.80	22.80	26.65	35.50	48.70

Table 11 | Joint improvement probability and corresponding discharge of Q, NH_3-N and COD_{Mn}

(NH_3-N, COD_{Mn})	40.0%	45.0%	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%
(III, III)	15.00	21.90	38.90	138.00	–	–	–	–	–
(IV, IV)	9.80	12.00	15.10	19.20	24.40	35.80	64.50	–	–
(V, V)	8.90	10.80	13.00	16.20	20.10	25.40	28.60	52.60	108.10
(III, IV)	10.75	13.40	16.80	22.30	34.03	93.10	160.00	–	–
(III, V)	10.10	12.30	14.90	19.20	25.30	39.80	100.60	170.00	–
(IV, V)	9.20	11.00	13.10	15.80	19.30	24.50	33.15	59.60	131.50
(IV, III)	12.70	17.25	24.75	42.90	138.20	–	–	–	–

Table 12 | Comparison of the optimal discharges between multiple water quality indicators and single water quality indicators

Indicators	Recovery grade	Optimal discharge (m^3/s) Curvature	Slope	Corresponding probability (Curvature)
NH_3-N	III	25.40	22.76	0.5413
	IV	27.56	23.50	0.6332
	V	28.30	24.15	0.6826
COD_{Mn}	III	27.90	22.32	0.4825
	IV	29.75	24.55	0.6667
	V	32.30	24.76	0.7241
(NH_3-N, COD_{Mn})	(III, III)	20.45	17.65	0.4393
	(IV, IV)	23.80	20.03	0.5902
	(V, V)	26.74	21.75	0.6454
	(III, IV)	21.78	19.14	0.5342
	(III, V)	22.35	19.65	0.5668
	(IV, V)	24.18	21.88	0.6352
	(IV, III)	20.81	16.60	0.4671

Discussion

Bivariate and trivariate joint probability distributions have been constructed based on copula functions, and the typical joint probability values of Q, NH_3-N and COD_{Mn} have been given. Under the condition of NH_3-N listed in Table 6, the occurrence probability is 0.6152 when Q is less than 40 m^3/s and NH_3-N is less than 1.0 mg/L. Under the condition of COD_{Mn} in Table 7, the occurrence probability is 0.6947 when Q is less than 40 m^3/s and COD_{Mn} is less than 6.0 mg/L. In contrast, under the conditions of NH_3-N and COD_{Mn} in Table 8, the occurrence probability is 0.4965 when Q is less than 40 m^3/s, NH_3-N is less than 1.0 mg/L and COD_{Mn} is less than 6.0 mg/L. These results show that the trivariate joint probability of the set (Q, NH_3-N, COD_{Mn}) is evidently less than the bivariate joint probability of the sets (Q, NH_3-N) and (Q, COD_{Mn}). Therefore, if only the bivariate joint probability distributions are considered, it is difficult to reflect the correct statistical characteristics of the discharge and water quality. On the contrary, the trivariate joint probability distribution comprehensively considers Q, NH_3-N and COD_{Mn} and can more reasonably illustrate the joint probabilities of different combinations of the discharge and water quality quantitatively.

Based on the results of the improvement probabilities of Q, NH_3-N and COD_{Mn} in Tables 9–11, respectively, additional discharge is evidently needed when the improvement probability is increased at a rate of 5%. Moreover, the calculation of an individual indicator can obtain only the improvement probability of that indicator while ignoring the improvement probability of other indicators; consequently, this approach cannot reflect the characteristics of the overall improvement of the main pollutants in a given water environment. For instance, according to the relationship between the discharge and a single indicator, under the condition in which the discharge is 20 m^3/s, the improvement probability of only NH_3-N recovering to grade III is 50%, which reflects only the improvement probability in which only NH_3-N exists in the water environment, and it does not simultaneously reflect the improvement of coexisting pollution indicators and their interactions. However, the joint improvement probability of multiple indicators is analysed through the statistical characteristics of the indicators and the discharge, which can help predict the joint improvement of several important indicators throughout the water environment.

Figures 7 and 8 and Table 12 show that the trivariate improvement probability is smaller than the bivariate probability despite the fact that the optimal discharge for improving multiple water quality indicators is less than that of a single indicator. The main reason is that there are certain synergistic and antagonistic effects among the pollution factors in the water system; as a result, the joint improvement of multiple water quality indicators is more difficult to achieve than the improvement of a single indicator. In general, multiple pollution factors coexist within a river, and thus, the joint improvement of multiple water quality indicators can reflect the relationships between the discharge and multiple indicators more reasonably.

Meanwhile, Table 12 illustrates that the optimal discharge calculated by the curvature method is higher than that calculated by the slope method. Therefore, the value calculated by the curvature method is adopted for the sake of safety.

CONCLUSIONS

In this study, the trivariate joint probability distribution of Q, NH_3-N and COD_{Mn} is established to indicate the occurrence probability of water quality indicators under different discharge conditions. Then, based on the trivariate joint probability distribution, a method for calculating the optimal discharge for the joint improvement of water quality indicators is proposed. The main conclusions are as follows:

1. Because the trivariate joint probability distribution of the set (Q, NH_3-N, COD_{Mn}) can better reflect the improvement of the water quality with multiple coexisting pollutants, this approach is more effective and comprehensive than the bivariate joint probability distribution of the sets (Q, NH_3-N) and (Q, COD_{Mn}) in estimating the combinations of water quality indicators under different discharge conditions.
2. Based on the trivariate joint probability distribution, the existence of a critical point along the $Q \sim C_j'$ curve is proven. Then, the joint improvement probability and the optimal discharges of multiple water quality indicators under different discharge conditions are calculated.
3. The methods that can be employed to fit the relationship curve of $Q \sim C_j'$ and identify the critical point on the curve are flexible and varied. Therefore, the selection of the proper method needs to be based on the specific situation of the river. Meanwhile, due to the limited availability of data, only a few water quality indicators are selected (i.e. only NH_3-N and COD_{Mn}). Therefore, in the future, the monitoring and accumulation of water quality indicator data should be strengthened to better reveal the relationships between the discharge and multiple water quality indicators.

ACKNOWLEDGEMENTS

This study was funded by the Non-profit Industry Financial Program of MWR: Key Technology Research and Demonstration Project of Water Ecological Civilization Construction (No. 201401003).

REFERENCES

Balistrocchi, M. & Bacchi, B. 2017 Derivation of flood frequency curves through a bivariate rainfall distribution based on copula functions: application to an urban catchment in northern Italy's climate. *Hydrol. Res.* **48** (3), 749–762.

Bazrafshan, J., Nadi, M. & Ghorbani, K. 2015 Comparison of empirical copula-based joint deficit index (JDI) and multivariate standardized precipitation index (MSPI) for drought monitoring in Iran. *Water Resour. Manage.* **29** (6), 2027–2044.

Bezak, N., Mikoš, M. & Šraj, M. 2014 Trivariate frequency analyses of peak discharge, hydrograph volume and suspended sediment concentration data using copulas. *Water Resour. Manage.* **28** (8), 2195–2212.

Bonansea, M., Ledesma, C., Rodriguez, C. & Pinotti, L. 2015 Water quality assessment using multivariate statistical techniques in Río Tercero Reservoir, Argentina. *Hydrol. Res.* **46** (3), 377.

Chitsazan, N., Nadiri, A. A. & Tsai, T. C. 2015 Prediction and structural uncertainty analyses of artificial neural networks using hierarchical Bayesian model averaging. *J. Hydrol.* **528**, 52–62.

Cox, B. A. 2003 A review of currently available in-stream water-quality models and their applicability for simulating dissolved oxygen in lowland rivers. *Sci. Total Environ.* **314–316**, 335–377.

Dai, Q., Han, D., Rico-Ramirez, M. A. & Islam, T. 2014 Modelling radar-rainfall estimation uncertainties using elliptical and Archimedean copulas with different marginal distributions. *Hydrol. Sci. J.* **59** (11), 1992–2008.

Environmental Quality Standards for Surface Water GB3838-*2002*. State Environmental Protection Administration of China/General Administration of Quality Supervision, Inspection and Quarantine of the PRC, Beijing, China.

Faruk, D. O. 2010 A hybrid neural network and ARIMA model for water quality time series prediction. *Eng. Appl. Artif. Intel.* **23** (4), 586–594.

Gippel, C. J. & Stewardson, M. J. 1998 Use of wetted perimeter in defining minimum environmental discharges. *Regulated Rivers Res. Manage.* **14** (1), 53–67.

Hao, Z. & AghaKouchak, A. 2013 Multivariate standardized drought index: a parametric multi-index model. *Adv. Water Resour.* **57**, 12–18.

Jang, C. S., Chen, S. K. & Kuo, Y. M. 2012 Establishing an irrigation management plan of sustainable groundwater based on spatial variability of water quality and quantity. *J. Hydrol.* **414** (2), 201–210.

Joe, H. 2005 Asymptotic efficiency of the two-stage estimation method for copula-based models. *J. Multivar. Anal.* **94** (2), 401–419.

Kao, S. C. & Chang, N. B. 2012 Copula-based flood frequency analysis at ungauged basin confluences: Nashville, Tennessee. *J. Hydrol. Eng.* **17** (7), 790–799.

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

Li, Y., Chen, C. & Sun, C. 2017 Drought severity and change in Xinjiang, China, over 1961–2013. *Hydrol. Res.* **48** (5), 1343–1362.

Liu, X., Tu, Q., Hua, Z., Huang, W., Xing, L. & Zhou, Y. 2015 Multi-parameter identification of a two-dimensional water-quality model based on the Nelder–Mead Simplex algorithm. *Hydrol. Res.* **46** (5), 711.

Liu, X., Teubner, K. & Chen, Y. 2016 Water quality characteristics of Poyang Lake, China, in response to changes in the water level. *Hydrol. Res.* **47** (S1), 238–248.

Luo, W. S. & Song, X. Y. 2014 *Engineering Hydrology and Water Conservancy Calculation*, 2nd edn. China Water & Power Press, Beijing.

Phan, T. D., Smart, J. C. R., Capon, S. J., Hadwen, W. L. & Sahin, O. 2016 Applications of bayesian belief networks in water resource management. *Environ. Model. Softw.* **85**, 98–111.

Qian, L., Liu, Y. & Chao, J. Y. 2013 The current situation and development trend of China joint regulating of water quality and water quantity. *Environ. Sci. Technol.* **36** (6 L), 484–487.

Rauf, U. F. A. & Zeephongsekul, P. 2014 Analysis of rainfall severity and duration in Victoria, Australia using non-parametric copulas and marginal distributions. *Water Resour. Manage.* **28** (13), 4835–4856.

Schumann, A. 2017 Flood safety versus remaining Risks-Options and limitations of probabilistic concepts in flood management. *Water Resour. Manage.* **31** (10), 3131–3145.

Shokri, A., Haddad, O. B. & Mariño, M. A. 2014 Multi-objective quantity-quality reservoir operation in sudden pollution. *Water Resour. Manage.* **28** (2), 567–586.

Sklar, A. 1959 N-dimensional distribution functions and their margins. *Publ. Inst. Stat. Univ. Paris* **8**, 229–231.

Sun, Z. B., Wang, B. L., Ji, H. F., Huang, Z. Y. & Li, H. Q. 2011 Water quality prediction based on probability-combination. *Chin. Environ. Sci.* **31** (10), 1657–1662.

Wang, S. W., Qi, S. Q., Yu, D. D., Zhang, Y. W. & Wan, L. H. 2015 Forecast and evaluation of water environment quality based on WASP model: a case study on Harbin section of Songhua River. *J. Nat. Disasters* **24** (1), 39–45.

Wu, S. F., Zhang, X., Shao, Q. X. & Christopher, J. G. 2015 Impact of river discharge on water quality combination events under different scenarios: a case of Bengbu Sluice in Huai River Basin. *J. Basic Sci. Eng.* **23** (4), 669–679.

Yu, S., He, L. & Lu, H. W. 2016 An environmental fairness based optimisation model for the decision-support of joint control over the water quantity and quality of a river basin. *J. Hydrol.* **535**, 366–376.

Yusof, F., Hui-Mean, F., Suhaila, J. & Yusof, Z. 2013 Characterisation of drought properties with bivariate copula analysis. *Water Resour. Manage.* **27** (12), 4183–4207.

Zegpi, M. & Fernandez, B. 2010 Hydrological model for urban catchments – analytical development using copulas and numerical solution. *Hydrol. Sci. J.* **55** (7), 1123–1136.

Zhang, X., Ran, Q. X., Xia, J. & Song, X. Y. 2011 Jointed distribution function of water quality and water quantity based on Copula. *J. Hydr. Eng.* **42** (4), 483–489.

Zhu, L., Li, H. E. & Li, J. K. 2012 Study on the response relationship between water quality and quantity of heavy pollution river in the arid and semi-arid areas. *Acta Sci. Circumstant.* **32** (10), 2617–2624.

First received 7 December 2017; accepted in revised form 29 May 2018. Available online 21 June 2018