Correlation between chlorophyll-a and related environmental factors based on Copula in Chaohu Lake, China

Huiqun Ma, Yong Wang

Shandong Electric Power Engineering Consulting Institute Corp, LTD., Jinan, China

*Corresponding author e-mail: mahuiqun@163.com

Abstract. Chlorophyll-a is crucial to characterize the eutrophication status. Return period of Chlorophyll-a is very helpful to the treatment of lakes and reservoirs eutrophication. In this study, the Gumbel-Hougaard Copula method is applied to discern the inherent relationship between chlorophyll-a and environmental variables of Chaohu Lake, China. The result shows that Weibull probability distribution is the preferred marginal distribution for Chl-a, Gamma for TP and TN, Log-normal for SD and CODmn. The sequence of correlation between chlorophyll-a and environmental factors is CODmn, SD, TP and TN. The method proposed here presents an effective tool to analyze the interaction of eutrophic variables in complex water environment system, also could provide reference for integrated management and treatment of lakes and reservoirs.

1. Introduction

In probability theory and statistics, a copula is a multivariate probability distribution for which the marginal probability distribution of each variable is uniform. Copulas are used to describe the dependence between random variables. Copulas have been used widely in both fresh water (Chen 2014, Eric 2014, Jiang 2015, Mojca 2015) and coastal water (Marinella 2015).

At present eutrophication is the most important environmental problem in many lakes and thus brings a tremendous influence on sustainable development of society and economy in lake regions in China (Xu 2014.)

Chlorophyll (Jason 2014) is a green pigment found in almost all plants, algae, and cyanobacteria. It is always used as an estimate of phytoplankton biomass. Nutrient enrichment may give rise to increased phytoplankton biomass, increased frequency and duration of phytoplankton blooms and increased primary production. Many researchers (Anne 2014, Dimitris , 2015, Jiang 2014) have showed that environmental factors are more or less linked to causes or effects of eutrophication, and interactions of these variables determine the variation in physicochemical properties and the growth and change in living organisms over time. Analysis of the effects of environmental factors on chlorophyll-a is helpful to understand the relationship between phytoplankton and environmental variables, and also provides the foundation for water quality assessment.

Most of the methods studying the relation of Chlorophyll-a and other environmental factors that have developed so far are linear or regressive formula in nature. In fact, the interaction of these variables is highly complex. It is a difficult problem, yet to be adequately resolved. This paper introduces frequency analysis into environmental field. Gumbel-Hougaard Copula is used to describe

the relationship between Chlorophyll-a and related environmental factors. Establish the joint distribution of Chlorophyll-a and other environmental factors. Then conditional joint return period of Chlorophyll-a given environment variables is calculated. It shows the contribution to Chlorophyll-a of each environmental variables.

2. Materials and Methods

The correlation of environmental factors can be got form analysis of joint distribution. The Copula method has been developed by Sklar (1959) and others. Central to this method is the determination of the dependence structure that is represented by a Copula. For random variables, X, Y, with their cumulative distribution function s, respectively, as $F_X(x)$ and $F_Y(y)$, their joint distribution function as $F(x, y)$, $F(x, y) = P(X \leq x, Y \leq y)$, let $u = F_X(x)$, $v = F_Y(y)$ one can obtain that

$$x = F_X^{-1}(u) \qquad y = F_Y^{-1}(v) \tag{1}$$

Where F_X^{-1} and F_Y^{-1}, is the inverse function of $F_X(x)$ and $F_Y(y)$, respectively. According to Eq.(1), one obtains

$$C(u, v) = F(F_X^{-1}(u), F_Y^{-1}(v)) \tag{2}$$

Where $C(u, v)$ is the dependence structure of variables, X, Y. According to $F(x, y)$ and $F_X(x)$ and $F_Y(y)$, $C(u, v)$ can be obtained. According to $C(u, v)$ and $F_X(x)$ and $F_Y(y)$, $F(x, y)$ can be obtained too. Thus, these Copulas can not be applied for arbitrary dependence structures between the variables and should be adopted appropriately according to the dependence characteristics.

2.1. Joint distribution of Chlorophyll-a and other environmental factors

Gumbel-Hougaard Copula is in common use in hydrologic field. This paper also uses it to describe the dependence between Chlorophyll-a and other environmental factors:

$$C(u, v) = \exp(-((-\ln u)^\theta + (-\ln v)^\theta)^{1/\theta}) \quad \theta \in [1, +\infty] \tag{3}$$

Where

$$\varphi(t) = (-\ln t)^\theta, \tau = 1 - \theta^{-1} \tag{4}$$

Determine Kendall's τ from observations as:

$$\tau_N = \binom{N}{2}^{-1} \sum_{i<j} sign[(x_i - x_j)(y_i - y_j)] \tag{5}$$

$$sign = \begin{cases} 1 & (x_i - x_j)(y_i - y_j) \geq 0 \\ -1 & (x_i - x_j)(y_i - y_j) < 0 \end{cases} \quad i, j = 1, 2, \text{L}, N \tag{6}$$

Where N is the number of observations; τ_N is the estimate of τ from observations.

2.2. Conditional joint return periods
Conditional joint return period is an important factor in hydrology field. Though it, the contribution to Chlorophyll-a of each environmental variables can be got.

The conditional distribution using the Copula method can be expressed. Let X and Y be random variables with $u = F_X(x)$, $v = F_Y(y)$. As an example, the conditional distribution function of X given $Y = y$ can be expressed by the Copula method as

$$
\begin{aligned}
F(X \leq x | Y = y) &= C_\theta(u | V = v) \\
&= \lim_{\Delta v \to 0} \frac{C_\theta(u, v + \Delta v) - C_\theta(u, v)}{\Delta v} \\
&= \frac{\partial}{\partial v} C_\theta(u, v) | V = v
\end{aligned}
\tag{7}
$$

Similarly, an equivalent formula for the conditional distribution function for *Y given X = x* can be obtained.

Thus, conditional joint return period can be got:

$$
T(X \leq x | Y = y) = \frac{1}{1 - F(X \leq x | Y = y)}
\tag{8}
$$

In the next section, conditional joint return period of Chlorophyll-a given environment variables is calculated, by which the influence of environmental factors affecting to

3. Results

Chaohu Lake is a lake located at the juncture of Chaohu and Hefei cities in Anhui Province, China. It is the largest lake in Anhui and one of the five largest freshwater lakes in China. Laoshan Island is situated within the lake. About 5 million people live near the lake, and use it for irrigation, transportation and fishing. Heavy use of the lake in recent years has led to eutrophication and silting. Due to China's rapid economic growth, the lake is now one of China's most polluted lakes. Location of Chaohu Lake can be seen in figure 1.
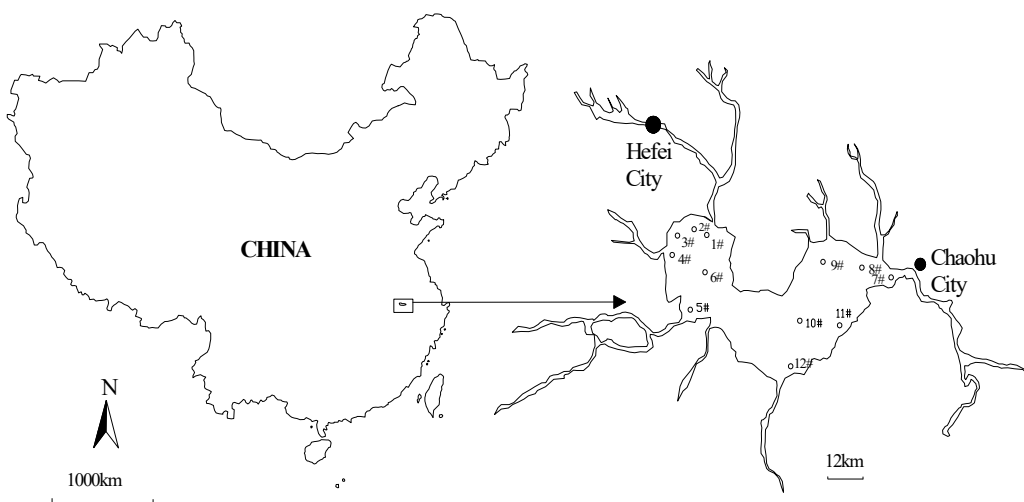
Figure 1 Location of Chaohu Lake in China

In China, the Ministry of Water Resources of China and State Environmental Protection Administration of China have professionals whose primary responsibility is to collect water quality data. These data are subject to a variety of national standards and professional standards, such as those of the State Environmental Protection Administration of China (SEPAC, 2002) and of Ministry of Water Resources of China (MWRC, 1999), etc., which assure and control the quality of the data. These standards also regulate sampling, monitoring, inspecting, analyzing, and experimenting of water quality data in a comprehensive manner. There are 12 monitoring sites in Chaohu Lake. The data used here is the average month value of these twelve sites from January 2000 to December 2007. TP, TN, CODmn, SD and Chlorophyll-a are used in this section. They are often chosen in eutrophication assessment. Statistics information of these data is given in Table 1.

Table 1 Statistics information of environmental variables

Index	SD(m)	CODmn(mg/L)	TN(mg/L)	TP(mg/L)	Chlorophyll-a(mg/m^3)
Mean	0.46	5.15	2.23	0.20	28.14
Standard Deviation	0.50	1.15	1.13	0.10	55.46
Maximum Value	3.65	8.09	6.79	0.57	447.69
Minimum Value	0.11	1.15	1.04	0.08	0.72

The marginal distributions used were: normal, gamma, Logistic, weibull and lognormal distribution. The parameters of these distributions were estimated by the maximum likelihood method. The root mean square error (RMSE) of marginal distribution and plotting position formula was used to measure the goodness-of-fit of the distribution. RMSE can be seen in table 2.

Table 2 Root mean square error values of different marginal distribution

Index	Normal	Lognormal	weibull	Gamma	Logistic
SD	0.2365	**0.1738**	0.2209	0.2006	0.1776
CODmn	0.0538	**0.0484**	0.0520	0.0503	0.0583
TN	0.0844	0.4354	0.0688	**0.0591**	0.0636
TP	0.0715	3.5725	0.0578	**0.0505**	0.0666
Chlorophyll-a	0.1732	0.5646	**0.0612**	0.0676	0.0917

Table 2 shows that lognormal distribution was the preferred marginal distribution for SD and CODmn, gamma distribution for TN and TP, and, weibull distribution for Chlorophyll-a. Although it is different of marginal distribution, Copula method is suitable for coupling them. According to formula (4)-(6), Kendall τ and parameter θ can be got, listed in table 3.

Table 3 Kendall τ and parameter θ of different environmental factors

Index	SD	CODmn	TP	TN
Kendall τ	0.27	0.49	0.40	0.59
θ	1.38	1.97	1.66	2.43

Kendall τ is a well-known measure of dependence based on ranks. Through table 3, it can be seen that the ranks of dependence of Chlorophyll-a and environmental factors are: TN, CODmn, TP and SD. One thing should be noted that, in our instinct understand, Chlorophyll-a and SD should be weak negative correlations. But in this study it is weak positive correlations. The reason may be from two aspects: (1) the data array is not long enough; (2) the data used in this study is average data and the value is very centralized.

In order to study the correlation between Chlorophyll-a and different environmental factors, lake eutrophication assessment grade standard was used in this section, seen in table 4.

Table 4 Grade standards of lake eutrophication

Index	SD(m)	CODmn(mg/L)	TN(mg/L)	TP (mg/ m^3)	Chlorophyll-a(mg/m^3)
Very poor	27	0.12	0.02	1.0	0.26
Poor	8	0.48	0.08	4.6	1.60
Middle	2.40	1.80	0.31	23	10
Eutrophic	0.73	7.10	1.20	110	64
Heavy eutrophic	0.40	14	2.30	250	160
Very eutrophic	0.12	54	9.10	1250	1000

From table 1 and table 4, it can be seen that the average value of Chlorophyll-a and CODmn is between the grade of middle and eutrophic. TN, TP and SD is between the grade of eutrohic and heavy eutrophic. So the conditional value of environmental factors is as follows:

(1) TN, TP and SD choose the value of eutrohic and heavy eutrophic;

(2) CODmn chooses the value of middle and eutrophic.

Figure 2 shows the conditional joint return period of Chl-a given environmental variables.
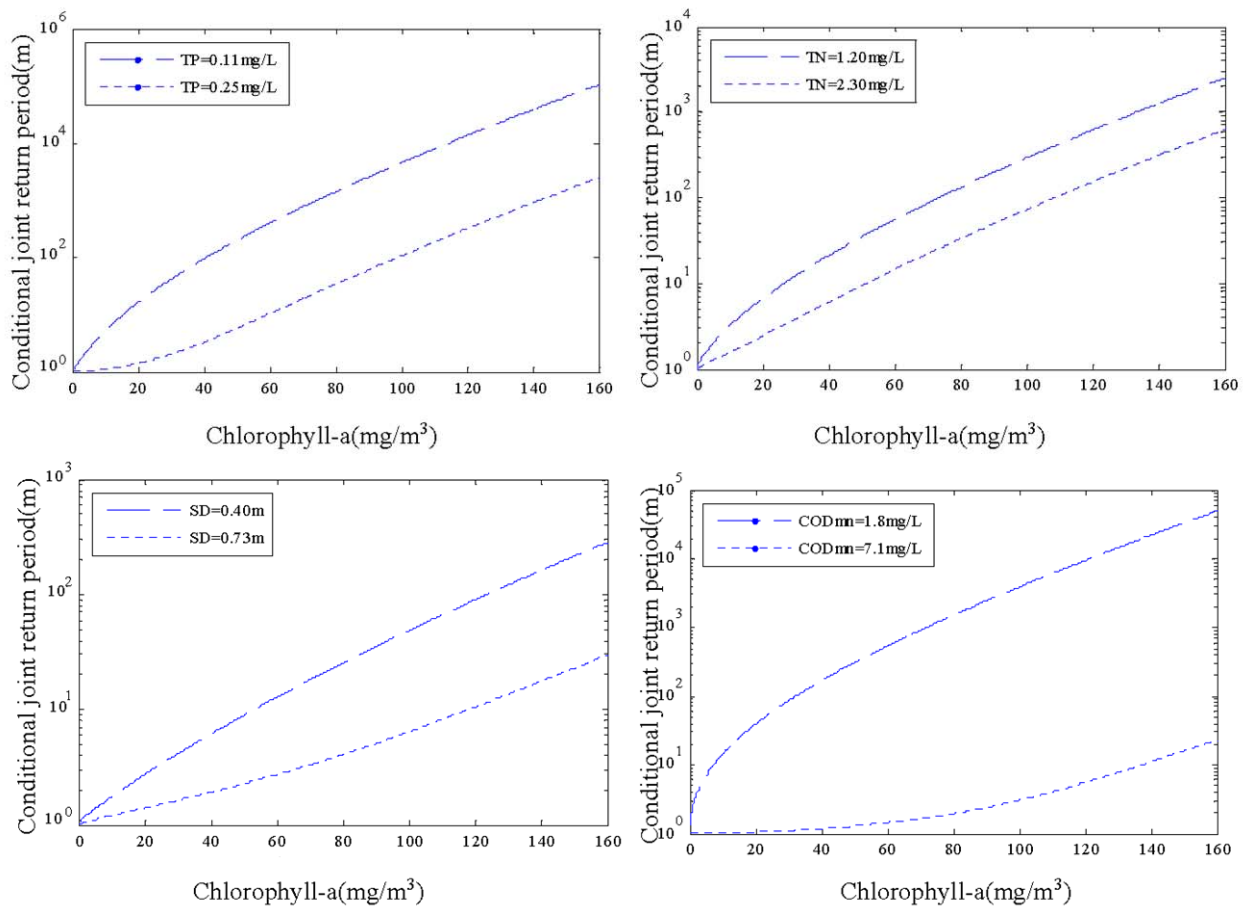
Figure 2 Conditional joint return period of Chl-a given environment variables

4. Discussions

(1) If SD=0.4m, the conditional joint period of Chlorophyll-a reaching eutrophic grade is about 1 year, reaching heavy eutrophic grade is about 24years; If SD=0.73m, the conditional joint period is about 0.25 year and 2.5 years respectively;

5

(2) If CODmn=1.8mg/L, the conditional joint period of Chlorophyll-a reaching eutrophic grade is about 55 year, reaching heavy eutrophic grade is about 4094 years; If CODmn=7.1mg/L, the conditional joint period is about 0.125 year and 2 years respectively;

(3) If TN=1.2mg/L, the conditional joint period of Chlorophyll-a reaching eutrophic grade is about 5.5 year, reaching heavy eutrophic grade is about 207 years; If TN=0.25mg/L, the conditional joint period is about 1.5 years and 51 years respectively;

(4) If TP=0.11mg/L, the conditional joint period of Chlorophyll-a reaching eutrophic grade is about 4 year, reaching heavy eutrophic grade is about 8682 years; If TP=0.25mg/L, the conditional joint period is about 1 year and 203 years respectively.

If each environmental factor reaches eutrophic grade respectively, conditional joint period of Chlorophyll-a is 1 year, 0.125 year, 5.5 years and 4 years corresponding to SD, CODmn, TN and TP. In other words, according to the effect to Chlorophyll-a, the sort order is CODmn, SD, TP and TN.

This paper exploring introduces frequency analysis into entrophication research. Copulas are used to characterize the inherent relationship between chlorophyll-a and environmental variables. Application of Chaohu Lake in China shows the method proposed here presents an effective tool to analyze the interaction of eutrophic variables in complex water environment system, also could provide reference for integrated management and treatment of lakes and reservoirs.

References

[1]   Anne P., Aafke M., Ton M. et al. 2014 Unraveling the relationships between freshwater invertebrate assemblages and interacting environmental factors. *Freshwater science,* 33(4), 1148-1158.

[2]   Chen P., Singh V., Guo S. 2014 Measure of correlation between river flows using the Copula-Entropy method. *Journal of Hydrological Enginnering*, 18(2013), 1591-1606.

[3]   Dimitris S., Heiko B., András Z. et al. 2015 Assessment of ecophysiology of lake shore reed vegetation based on chlorophyll fluorescence, field spectroscopy and hyperspectral airborne imagery. *Remote Sensing of Environment*, 157(2015), 72-81.

[4]   Eric P., Anne-Catherine F., Jacques B. et al.2014 Copula models for frequency analysis what can be learned from a Bayesian perspective? *Advances in Water Resources*, 63 (2014), 91–103.

[5]   Jason D., Mark E., Moninya R. et al. 2014 Relative impact of seasonal and oceanographic drivers on surface chlorophyll a along a Western Boundary Current. *Progress in Oceanography*, 120 (2014), 340–351.

[6]   Jiang C.,Xiong L., Xu C. et al. 2015 Bivariate frequency analysis of nonstationary low-flow seriesbased on the time-varying copula. *Hydrological Processes*, 2015, 29(6), 1521–1534.

[7]   Jiang Y., He W., Liu W. et al. 2014 The seasonal and spatial variations of phytoplankton community andtheir correlation with environmental factors in a large eutrophic Chinese lake (Lake Chaohu). *Ecological Indicators*, 40(2014),58-67.

[8]   Marinella M., Alberto L., Renata A. 2014 Coastal flooding, A copula based approach for estimating the joint probability of water levels and waves. *Costal Engineering*, 97(2015), 37-52.

[9]   Mojca S. , Nejc B., Mitja B. 2015 Bivariate flood frequency analysis using the copula function, A case study of the Litija station on the Sava River. *Hydrological Processes*, 29(2), 225-238 .

[10]   Sklar. 1959 Fonctions de répartition à n dimensions et leurs marges. *Publishing Institute of Statistical University of Paris*, 1959 (8), 229 – 231.

[11]   Xu S., Peng Q, Li M. 2014 Morphospecies and genospecies ofMicrocystisduring bloomsin eutrophic Lake Taihu (China) in autumn. *Biochemical Systematics and Ecology*, 57(2014), 322-327.