Desalination and Water Treatment
www.deswater.com
doi: 10.1080/19443994.2013.793921

51 (2013) 7331–7340
November

Taylor & Francis
Taylor & Francis Group

Eutrophication analyses and principle component regression for two subtropical storage reservoirs in Macau

Weiying Zhang[a], In Chio Lou[a,*], Yijun Kong[b], Wai Kin Ung[b], Kai Meng Mok[a]

[a]*Faculty of Science and Technology, Department of Civil and Environmental Engineering, University of Macau, Av. Padre Tomás Pereira, Taipa, SAR, Macau, P.R. China*
Email: iclou@umac.mo
[b]*Laboratory and Research Center, Macao Water Co. Ltd., SAR, Macau, P.R. China*

Received 18 March 2012; Accepted 12 February 2012

ABSTRACT

Eutrophication analyses of two subtropical storage reservoirs in Macau, the Special Administrative Region of China, namely Main Storage Reservoir (MSR) and Sai Pa Van Reservoir (SPVR), were performed in this study. Totally, 17 monthly water parameters including, five hydrological parameters (precipitation, imported volume, exported volume, water level, and hydraulic retention time), four physical parameters (temperature, pH, turbidity, and conductivity), seven chemical parameters (dissolved oxygen, Ammonium, nitrite, nitrate, total nitrogen (TN), orthophosphate (PO_4^{3-}), and total phosphorus (TP)), and one biological parameter (phytoplankton abundance) were sampled and monitored in 2010. The correlation analysis and principle component regression (PCR), that is, principle component analysis (PCA) followed by multiple linear regression (MLR), were used to simplify the complexity of the relationships and to predict the phytoplankton abundance levels as well. The eutrophication analyses results showed that both reservoirs were in eutrophic status with the trophic state indices of 58–72 for MSR and 51–71 for SPVR, respectively. Phytoplankton abundance in both reservoirs were found to be linearly correlated with turbidity, temperature, and TP, while anti-correlated with conductivity, TN, nitrate, TN/TP, and water level. The PCA showed that three PCs with Eigen value over one, can explain 84.6% of total variation of the water parameters in MSR, while only two PCs can explain 70.8% for SPVR. The MLR models can be used for predicting phytoplankton abundance in the reservoirs with the predictive power of 0.90 in MSR, while only of 0.67 in SPVR.

Keywords: Reservoirs; Phytoplankton abundance; Eutrophication analysis; Principle component regression

1. Introduction

Freshwater algal blooms has emerged as a challenging issue in water management due to the increases in the occurrence and severity, and the production of toxin by fresh water algae (also called phytoplankton) poses a significant threat to drinking water safety. Blooms occur in eutrophication of water bodies, which are the results of an excess of nutrients, and under appropriate environmental conditions phytoplankton population can rapidly increase or

*Corresponding author.

1944-3994/1944-3986 © 2013 Balaban Desalination Publications. All rights reserved.

accumulate. Thus, an understanding of trophic level of water bodies and factors which induce the blooms of phytoplankton is important to safeguard water resources.

Carlson's trophic state index (TSI) [1] is the most commonly used index to describe trophic level of lakes or reservoirs, which depends on the parameters of phosphorus, chlorophyll, and Secchi depth. Four classes such as, oligotrophic, mesotrophic, eutrophic, and hypereutrophic class with the corresponding TSI of < 30–40, 40–50, 50–70, $70–100^+$, respectively are defined, from low to high primary productivity. Eutrophic and hypereutrophic lakes or reservoirs are susceptible to algal bloom. Previous laboratory studies [2–4] have indicated, that the growth of phytoplankton is influenced by a number of environmental variables such as nutrients, light, temperature, pH, conductivity, turbidity, stable conditions, flow rates and water levels, and the complex interaction of these variables leads to the development of algal blooms. However, the combination of such variables that triggers and sustains an algal bloom is not well understood and it is impossible to attribute algal blooms to any specific environmental variable. Besides, the species and the concentration of phytoplankton have been showed to have varied response and dynamics to different environmental variables [5–7]. Hence, an important challenge for environmental engineers and scientists are to search for an effective method to understand the interaction and behaviors of the variables involved in the multidimensional complex processes, for monitoring water resource, as well as forecasting the phytoplankton abundance.

Multiple regression analysis is the most widely used methodology for expressing the dependence of a response variable using a couple of independent variables. Though there is some evident success in many applications, such as water resource and air quality studies, the regression approach does face serious difficulties when the independent variables are correlated with each other [8]. High correlation or multicollinearity between the independent variables in a regression equation can make it difficult to correctly identify the most important contributors to a physical process. One of the solutions that receive growing interest is PCR, that is, principle component analysis (PCA) followed by multiple linear regression (MLR). The PCR analysis is to use an orthogonal transformation, to convert a set of observations of possibly correlated variables into a set of values of uncorrelated variables called PCs, thus reducing the complexity of multidimensional system by maximization of component loadings variance and elimination of invalid components. PCA have been used alone or in combination with other methods, such as MLR, to model aquatic environmental and ecological processes including algal blooms problem, particularly in predicting the chlorophyll-a and algae population in reservoirs [9–11]. From these studies, only the PCs with Eigen values greater than one were selected for MLR, which can explain the high percentage of total variation of the environmental variables in PCA. It is followed by the MLR to check if the chlorophyll a, cyanobacteria abundance, or microcystin concentrations could be explained by environmental variables, and to use for further prediction.

In this study, we present the temporal variations of 17 water parameters with the corresponding phytoplankton population in the subtropical reservoirs in Macau, and examine the significant water parameters that affect the population dynamics. The studied MSR and SPVR were the main storage units utilized for drinking water and experienced the algal bloom problem in the recent years. To understand the water parameters that cause the blooming, the TSI of the reservoir was estimated for determining the trophic level classification, and combination of correlation analysis and PCA were performed, to simply the complexity the water parameters, after which the PCS (principle component scores) were used as independent variable in the MLR for predicting the phytoplankton abundances of both the reservoirs in Macau.

2. Materials and methods

2.1. Study areas

Macau is situated 60 km southwest of Hong Kong, and experiences a subtropical seasonal climate that is greatly influenced by the monsoons. The difference of temperature and rainfall between summer and winter are significant though not great. MSR, located in the east part of Macau peninsula, is the biggest reservoir in Macau with the capacity of about 1.9 million m^3 and the water surface area of 0.35 km^2, while SPVR, located in the center of Coloane peninsula, is the other drinking water reservoir with the capacity of about 0.45 million m^3 and the water surface area of 0.12 km^2. They both are pumped storage reservoirs that receive raw water from the West River of the Pearl River network, and can provide water supply to the whole areas of Macau for about one week. MSR and SPVR are particularly important, as the temporary water source during the salty tide period when high salinity concentration is caused by the intrusion of sea water to the water intake location. In recent years, there were reports (Macao Water Co. Ltd., unpublished data) that the reservoirs have been experiencing some problems of algal blooms and the situation appeared to be worsening.

2.2. Field sampling

Location in the inlet of the reservoirs were selected (Fig. 1) for sampling. Samples were collected in duplicate monthly from January to December 2010 at 0.5 m from the water surface. All samples were kept in ice boxes and transported to the laboratory for further analysis. A total of 17 water quality parameters, including hydrological, physical, chemical, and biological parameters, were monitored monthly. Precipitation was obtained from Macau Meteorological Center (http://www.smg.gov.mo/www/te_smgmail. php). The imported and exported volumes were recorded by the inlet and outlet flow meters, and the water levels were read by the ruler in the reservoirs, based on which the hydraulic retention time (HRT) can be calculated. Temperature was measured *in situ* with a mercury thermometer. The pH value was determined at the laboratory with a pH meter (DKKTOA, HM-30R), turbidity was measured using a turbidity meter (HACH, 2100 N IS), and conductivity was measured with an EC meter (DKKTOA, CM-30R). Dissolved oxygen (DO), Ammonium, nitrite, nitrate, total nitrogen (TN), PO_4^{3-}, and total phosphorus (TP) were measured according to the standard methods [12]. The phytoplankton samples were fixed using 5% formaldehyde and transported to laboratory for microscopic counting. Trophic status was assessed using Carlson's Trophic Status Index (TSI) that is based on TP concentration and chlorophyll a concentration [1] according to the following equations. The overall TSI of the reservoir was estimated by taking the average value of TSI (Chl) and TSI (TP).

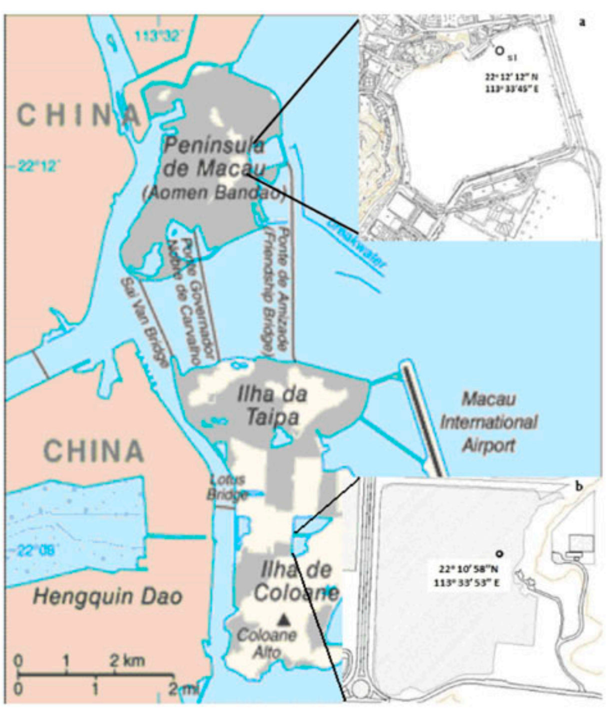

Fig. 1. Sampling location for the MSR (a) and SPVR (b).

$$\mathrm{TSI\ (Chl)} = 10(6 - \frac{2.04 - 0.68 \ln \mathrm{Chl}}{\ln 2}; \ \mathrm{TSI\ (TP)} = 10(6 - \frac{\ln \frac{48}{\mathrm{TP}}}{\ln 2})$$

2.3. Statistical analysis and PCR

Statistical analyses were carried out using PASW 19 software package (SPSS Inc.). Logarithmic transformation was applied to phytoplankton abundance data. The TN/TP ratio was calculated and taken as one individual variable. Besides, due to the limited available data (12 observations for each variable) for PCA, only those variables that can highly potentially explain from the mechanisms the change of phytoplankton population were used for statistical analyses Based on the preliminary data analysis, totally 11 water parameters including the phytoplankton abundance were selected. Correlation analysis was conducted to identify water parameters which were significantly correlated with phytoplankton abundance. Except for phytoplankton population, the 10 parameters with complete data set were accessed with Kaiser–Meyer–Olkin (KMO) measure of sample adequacy and Bartlett's test of Sphericity (χ^2 with degrees of freedom = 1/2 [$p(p - 1)$]) was used to verify the applicability of PCA [13,14]. Only parameters with communalities great than 0.5 were used for analysis.

By using the PCA method, the input variables were changed into PCs that are independent and linear compound of input variables, instead of direct using of input variables, thus the information of input variables will present with minimum losses in PCs. In this study, PCA was performed on these water parameters to rank their relative significance and to describe their interrelation patterns, as well as onto the phytoplankton population levels. The PCS of the selected water parameters were used as independent variables in the MLR to check if the occurrences of phytoplankton population could be explained by water parameters, as well as to predict the phytoplankton abundance.

In MLR model, only the PCS of the PCs with Eigen value greater than one were used as independent variables for predicting the phytoplankton abundance. Application of PCS in MLR models for prediction was described by Çamdevýren et al. [9] and Draper and Smith [15].

Regression model in matrix form can be shown as:

$$Y = XC + e \quad (1)$$

where Y is the response matrix, C is the regression coefficient matrix, and e is the fitting error matrix. By solving Eq. (1) for C we get,

$$C = (X'X)^{-1}(X'Y) \quad (2)$$

where X' is the transpose of X. To solve this problem, the multicollinearity between independent variables with PCA was removed. MLR analysis of phytoplankton abundance on the PC scores was performed using stepwise variable selection procedures to identify the best predictors of phytoplankton abundance. Nonsignificant score values were excluded from the model by stepwise method. The *t*-test method was used in testing the regression coefficients. Determination coefficient (R^2) was used as the standard criterion of predictive success which was widely used in ecological modeling.

3. Results and discussion

3.1. Trophic status and environmental parameters measurement

The values of chlorophyll a were in the range of 62.54–71.78 mg/L, with the average of 67.68 mg/L and standard deviation of 3.45 mg/L. TSI was used as an indicator for evaluating the eutrophication status of the reservoirs. TSI (Chl), TSI (TP), and TSI (overall) were calculated as 63–72 and 58–68, and 58–72, respectively, for MSR, and 62–71, 51–64 and 51–71, respectively, for SPVR, indicating that both the reservoirs were categorized as the reservoirs in eutrophic and hypereutrophic status, which are consistent with our observation that the reservoirs are susceptible to the algal blooms. The indicator can be used not only as a predictive tool in reservoir management programs, but also as a valid scientific tool for investigations where an objective standard of trophic state is necessary. The details of various water parameters affecting the algae population dynamics were described as below.

3.1.1. Hydrological parameters

The hydrological parameters over the whole year of 2011 were described in Fig. 2. The Macau rainfall in 2010 was 2,064 mm (Fig. 2(b, d)) with two peaks in June and September which accounted for more than 50% of the total amount, while nearly no rainfall in the dry season from October to March. If the evaporation amount of about 1,000 mm (data from Macao Water Co. Ltd.) was excluded, the net precipitation for the whole year was approximately 1,000 mm, that is, 350,000 m^3 for MSR and 120,000 m^3 for SPVR. In addition, it was observed in the Fig. 2(a, c) that the imported and exported volumes in 2010 were about 4.9 million m^3 and 8.1 million m^3, respectively for MSR, while 2.7 and 4.5 million m^3, respectively for SPVR. It has to be noted that there are nearly no imported water and large amount of exported water in July for MSR and in June for SPVR, which are because the Water Company attempted to maximize the utilization of raw water in the reservoirs, and thus replace it with the newly imported water, hopefully reducing the possibility of algal bloom problem occurred in the coming summer. Due to the low amount of exported water in June and July, the HRT in MSR was extremely high (>200 days), which can be classified as long HRT reservoir. SPVR belongs to short HRT reservoir with the average HRT of about 26 days. The water levels of MSR and SPVR were maintained at 3.3–5.2 m and 2–3.1 m respectively, with the low levels in summer and high levels in winter and spring.

Peak rainfall can create hydraulic turbulence that significantly affects the compositions and the dominant species of phytoplankton [16]. However, the amount of peak rainfall in Macau only accounted for less than 10% of effective volume of the reservoir, meaning that the effect of rainfall can be negligible. In addition, the water quality data from the upstream (Pearl River Water Resource Commission, unpublished data), showed that the nutrient and phytoplankton abundance are low, suggesting that the imported and exported water may not have strong correlation with phytoplankton abundance in the reservoirs. Thus, only HRT and water level of the hydrological parameters were included in the PCA study.

3.1.2. Physical parameters

The change of temperature, pH, turbidity, and conductivity for both reservoirs were showed in Fig. 3. Combined with phytoplankton population data (Fig. 5), our results agreed with previous study [6] that temperature strongly affects growth rates and metabolisms of phytoplankton. Conductivity was highest in January (~630 us/cm) for MSR, while in February (~553 us/cm) for SPVR, and both decreased linearly until July, and the value was maintained at

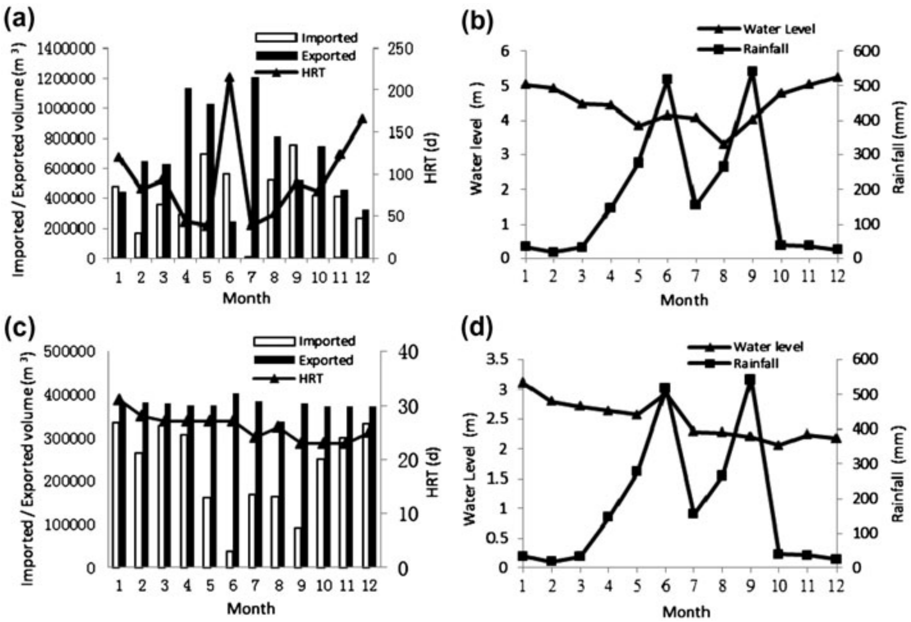

Fig. 2. Variation of hydrological parameters of MSR (a–b) and SPVR (c–d).

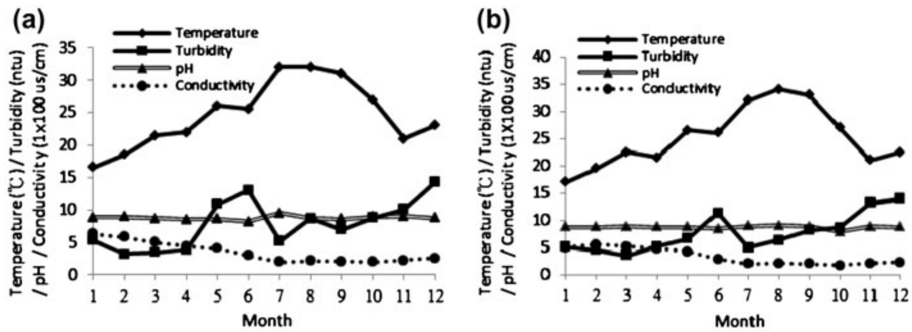

Fig. 3. Variation of physical parameters of MSR (a) and SPVR (b).

about 200 us/cm till the end of the year. In contrast to conductivity, turbidity was low in spring, and fluctuated for the remaining time of the year. These parameters would be included in statistical analysis, except for the pH that showed the low variance and was omitted for further study.

3.1.3. Chemical parameters

Fig. 4 showed the change of DO, N, and P concentrations for both reservoirs. It was interesting to note that for both reservoirs the TN and nitrate concentrations (Figs. 4(a) and 4(c)) were high in the spring, and dramatically decreased in April. The decrease continued to July and then maintained at low level in fall and winter. The difference there was one peak of TN appeared in October in SPVR (Fig. 4(c)) while did not happen in MSR (Fig. 4(a)). The TP, PO_4^{3-}, and DO concentrations kept much stable, with only a little high for TP in June for both reservoirs, though MSR generally had higher TP, and PO_4^{3-} than SPVR. Algal blooms are the results of the excessive of nutrient, mainly nitrogen and phosphorus concentrations [17]. TN [18] or TP [19] can be the control factor on the growth of phytoplankton and lead to change of composition.

In the study, the TP concentration fluctuated and seems to be not correlated to phytoplankton population (Table 1). Compared to the concentrations of nitrate and TN, the concentration of ammonium and nitrite (the immediate oxidation stage of ammonium to nitrate) were much lower, and the DO data showed low variability for the whole year. Thus the parameters of ammonium and DO were omitted in the PCA.

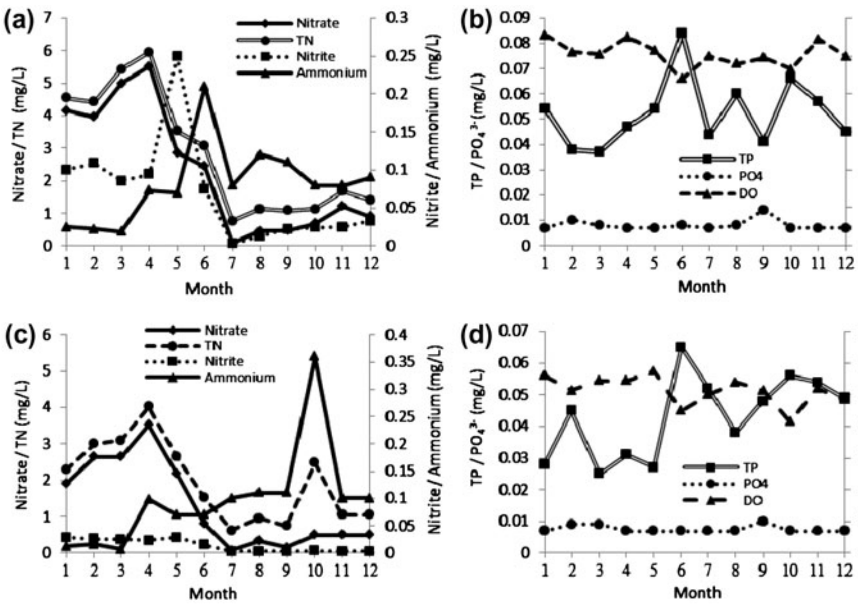

Fig. 4. Variation of chemical parameters of MSR (a and b) and SPVR (c and d).

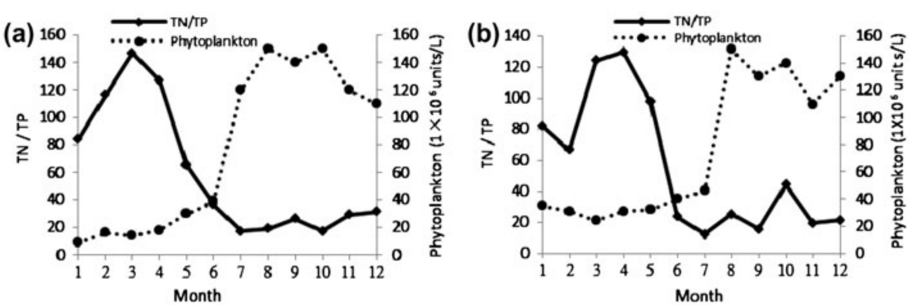

Fig. 5. Variation of biological parameter and TN/TP ratios of MSR (a) and SPVR (b).

3.1.4. Biological parameter and TN/TP ratio

The phytoplankton abundances in both reservoirs were maintained at low levels, less than 40 million cells/L for the first half an year, and dramatically increased to 120 million cells/L in July for MSR (Fig. 5(a)) or in August for SPVR (Fig. 5(b)). The high levels of algae population were kept to the end of year. The seasonal variation suggested that temperature was an important factor that causing the bloom. It was also found that the phytoplankton abundances increased with the decreasing TN/TP ratio. The TN/TP ratio has been reported to affect the dominant species of algae. High TN/TP (20–50) favors growth of green algae while low TN/TP (5–10) prefers blue algae [20].

3.2. Statistical analysis

3.2.1. Correlation analysis

The correlations between every two of water parameters were showed in Table 1 for MSR and SPVR. Phytoplankton abundances were correlated with temperature in MSR and turbidity in SPVR, while anti-correlated with TN, Nitrate, TN/TP ratio, and conductivity for both reservoirs. These implied that the nutrients were consumed in the rapid growth of phytoplankton abundances during the bloom. In addition, the phytoplankton abundance in SPVR were found to be anti-correlated with nitrite, HRT and water level, and a weakly correlated with TP or PO_4^{3-}, suggesting that phosphorus was not the control factor

Table 1
Correlation results between phytoplankton and water parameters in MSR and SPVR

		Turbidity	Temperature	Conductivity	Nitrate	TN	PO_4^{3-}/nitrite***	TP	TN/TP	HRT	Water level
Log_{10} Phytoplankton	MSR	0.496	**0.748****	**−0.967****	**−0.942****	**−0.944****	0.144	.202	**−.880****	−.020	−.243
	SPVR	**0.643***	0.445	**−0.852****	**−0.788****	**−0.667***	**−0.875****	.504	**−.731****	**−.717***	**−.811****

*Correlation is significant at the 0.05 level (2-tailed). **Correlation is significant at the 0.01 level (2-tailed).
***PO_4^{3-} was selected for MSR, while nitrite was chosen for SPVR due the high correlation coefficient in the corresponding reservoirs.

on the algal bloom. It can be observed that there was a significant amount of multicollinearity among the water variables, judging by the correlation coefficients between them. The presence of high multicollinearity leads to inappropriateness when applying the MLR later, which should be solved by PCA. This can be achieved only to select a small number of variables or components that would explain a high enough percentage (approximately 80% in this study) of the total variation in the water variables.

3.2.2. Principle component regression

3.2.2.1. Principle component analysis. The PCA was performed by using the selected 10 variables' complete datasets. The value of KMO was 0.610 (MSR) and 0.639 (SPVR), respectively, both above the criteria value of 0.6 [13]. The value of χ^2, calculated as 131.24 (MSR) and 129.048 (SPVR), respectively, with P-value less than 0.0005 by Barlett's Test of Sphericity test, indicating that the analysis was applicable [13].

Table 2 showed, that the first three components can explain 84.65% variation of the data variation in MSR. The screen test suggested only three components with the Eigen values greater than one to be retained, in which all the 10 environmental variables were included. PC1 was mainly composed of physical parameters and nitrogen source. PC2 was influenced by hydrological parameters. PC3 was defined as the phosphorus source (Table 3).

However, in SPVR only the first two components can explain 83.17% (Table 2) variation of the data variation. The screen test suggested two components with the Eigen values greater than one to be retained, in which all the 10 environmental variables were included. PC1 was mainly composed of physical/hydrological parameters and nitrite. PC2 was influenced by nutrient source and turbidity (Table 3).

3.2.2.2. Multiple linear regression. The main objective of this section was to select a subset of the water variables that provides the best prediction equation for the modeling of phytoplankton abundance by using the multiple regression method. The qualified, selected, independent variables should be those with high loadings, associated with each of the PCs included in the regression equation that had high coefficients of regression. The MLR results of $\log_{10}$ (phytoplankton) using the standardized values of PCS related to the results of PCA are summarized in Table 4. The high values of communalities indicated that the variances were efficiently reflected in the regression analysis (Table 3).

Table 2
Descriptive statistics of PCs in MSR and SPVR

		PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
MSR	Eigen value	5.017	2.284	1.164	.893	.297	.181	.121	.029	.015	.000
	Cumulative %	50.167	73.010	84.645	93.576	96.549	98.355	99.566	99.851	99.999	100.00
SPVR	Eigen value	7.084	1.233	0.868	0.404	0.196	0.093	0.061	0.048	0.009	0.004
	Cumulative %	70.837	83.171	91.846	95.886	97.843	98.775	99.387	99.867	99.958	100.00

In MSR, all 10 variables were included in the three selected PCs, showing that 90.9% of variation in phytoplankton abundance can be explained by PCS1 that defined a variation of 50.2% in water parameters. If the two omitted scores were included in the model, the determination can increase to 92.3% (not shown here), though the difference was not statistically significant. Other parameters, water level, HRT, TP, and PO_4^{3-}, which have significant impacts in PC2 and PC3, were excluded from the regression model. However, linear effects of these parameters on phytoplankton abundance were partially incorporated in PC1 of the predictive model written as Log_{10} (phytoplankton) = 7.693–0.097(PCS1) (Table 4).

Similar to the analysis in MSR, all of the 10 variables in SPVR were included in the two selected PCs, showing that 66.8% of variation in phytoplankton abundance can be explained by PCS1 that defined a variation of 70.8% in water parameters. If the omitted score was included in the model, the determination can increase to 71.8% (not shown here), though the difference was not statistically significant. Other parameters, turbidity, TP, TN/TP, nitrate, and TN were excluded from the regression model, which have significant impacts in PC2, while their linear effects on phytoplankton abundance had been partially incorporated in PC1 of the predictive model written as $\log_{10}$ (phytoplankton) = 7.781–0.074 (PCS1) (Table 4).

Table 3
Results of PCA ($\lambda > 1$) in MSR and SPVR

	Variables	Component			Communalities
		PC1	PC2	PC3	
MSR	Nitrate	**0.976**	0.065	0.032	0.958
	TN	**0.974**	0.033	0.041	0.951
	TN/TP	**0.950**	0.016	−0.207	0.946
	Conductivity	**0.916**	0.168	−0.109	0.878
	Temperature	**−0.729**	−0.645	−0.039	0.950
	Turbidity	**−0.607**	0.329	0.577	0.810
	Water Level	0.237	**0.873**	−0.120	0.832
	HRT	−0.174	**0.778**	0.366	0.770
	TP	−0.296	0.046	**0.850**	0.812
	PO_4^{3-}	−0.245	−0.022	**−0.706**	0.559
SPVR	Temperature	**−0.895**	−0.009		.801
	Water level	**0.837**	0.272		.774
	HRT	**0.774**	0.402		.760
	Nitrite	**0.750**	0.609		.933
	Conductivity	**0.744**	0.635		.957
	Turbidity	−0.043	**−0.884**		.783
	TP	−0.232	**−0.842**		.763
	TN/TP	0.492	**0.806**		.892
	Nitrate	0.646	**0.697**		.903
	TN	0.552	**0.668**		.751

Table 4
Results of MLR analysis for three PCs ($\lambda > 1$, stepwise method) in MSR and two PCs ($\lambda > 1$, stepwise method) in SPVR

	Included independent variables	Regression coefficient (B)	Std. Error of B	Std. regression coefficient (Beta)	t	Sig.	R^2
MSR	Constant[a]	**7.693**	0.043		178.442	0.000**	0.909
	PCS1	**−0.097**	0.010	−0.953	−9.967	0.000**	
SPVR	Constant[a]	**7.781**	0.056		137.783	0.000**	0.668
	PCS1	**−0.074**	0.016	−0.817	−4.481	0.001**	

Stepwise (Criteria: Probability-of-*F*-to-enter <= .050, Probability-of-*F*-to-remove >= .100).
[a]Predictors: (Constant), PC1 and dependent Variable: log10 phytoplankton.
**Correlation is significant at the 0.01 level (2-tailed).

Table 5
Comparison of predicted values and measured data of Log_{10} (phytoplankton) in MSR and SPVR

Month	MSR		SPVR	
	Observation	3PCs, stepwise	Observation	2PCs, stepwise
1	6.94	7.163488917	7.54	7.318817182
2	7.20	7.135373703	7.49	7.469520176
3	7.15	7.062720866	7.38	7.561773375
4	7.26	7.095977133	7.49	7.586620357
5	7.48	7.652459398	7.51	7.653316584
6	7.59	7.841722547	7.60	7.696268651
7	8.08	8.129864588	7.66	8.046944822
8	8.18	8.136882746	8.18	8.022269962
9	8.15	8.086553984	8.11	8.097495639
10	8.18	8.070912715	8.15	8.056516864
11	8.08	7.92024181	8.15	7.944106396
12	8.04	8.019801278	8.11	7.918349991

The higher regression coefficient for predicting phytoplankton abundance in MSR than in SPVR indicated that the PCR is more successful in MSR to be applied in water quality monitoring. The MLR-based predicted values and the corresponding monthly measured data of $\log_{10}$ (phytoplankton abundance) for both reservoirs were showed for comparison in Table 5.

4. Conclusion

It is well known that algal blooms in freshwater system are the most common water pollution problem, due to the discharge of excessive nutrient into water bodies. In this study, eutrophication analyses of two major raw water reservoirs in Macau were performed and their algae populations were predicted using statistical methods, correlation analyses, and principle component regression. Seventeen monthly water quality parameters were sampled and monitored in Macau MSR and SPVR that are experiencing algal bloom problems in recent years. Eutrophication analysis was performed to estimate the TSIs, and PCR was applied to predict the phytoplankton abundances based on the 10 selected parameters from correlation analysis. The results showed that the TSIs of MSR were estimated as 58–72 for MSR and 51–71 for SPVR, indicating that both the reservoirs were in eutrophic–hypereutrophic status that is susceptible to algal blooms. In MSR, the phytoplankton abundance was affected by temperature with linear regression coefficient of 0.75, and anti-correlations were found between phytoplankton abundance and TN ($R^2 = 0.94$), nitrate ($R^2 = 0.94$), TN/TP ratio ($R^2 = 0.88$), or conductivity ($R^2 = 0.97$), implying that the nutrients were consumed in the rapid growth of phytoplankton abundance during the bloom. On the other hand, the phytoplankton abundance in SPVR was affected by turbidity with linear regression coefficient of 0.64, and anti-correlations were found between phytoplankton abundance and TN ($R^2 = 0.67$), nitrate ($R^2 = 0.79$), nitrite ($R^2 = 0.88$), TN/TP ratio ($R^2 = 0.73$), conductivity ($R^2 = 0.85$), HRT ($R^2 = 0.72$), and water level ($R^2 = 0.81$), meaning that the high concentrations of nutrients were consumed in the rapid growth of phytoplankton abundance during the blooms. Besides, considering the weak correlation of phytoplankton population to TP, $R^2 < 0.21$ for MSR and $R^2 < 0.51$ for SPVR, it is highly probably that TN, instead of TP, was the control factor for the bloom. Compared to SPVR with lower prediction of 0.67 for phytoplankton abundance, MSR has higher regression coefficients of 0.91. The PCA can be used to simplify the complexity of algal bloom problem by identifying the dominant water parameters (nitrogen source, TN/TP ratio, and physical and hydrological parameters) that cause blooming, and the MLR using PCS as input was successful in eliminating multicolinearity problem, to remove indirect effect and number of water parameters, for predicting the phytoplankton population.

This study is a preliminary research into the factors contributing to algal blooms in Macau Reservoirs. However due to the complex nonlinear relationship between water variables and phytoplankton abundance, the PCR method may not be good enough for predicting the phytoplankton population, particularly for SPVR, considering the prediction power is low. Further studies based on nonlinear models, such as artificial neural network, support vector machine and relevance vector machine, need to be undertaken to better understand the mechanism of the algal bloom, as well as principle factors affecting the algae population. ANN is a well suited method with self-adaptability, self-organization and error tolerance [21], and SVM has advantages of only requirement of a small amount of samples, high degree of prediction accuracy and long prediction period by using kernel function to solve the nonlinear problems. It reduces the generalization ability, and the complexity of algorithm. Thus it will be helpful for us to forecast the algae population given the water variables in the reservoirs, and later to develop the water quality monitoring and reservoir management programs.

Acknowlegements

This research was supported by Fundo para o Desenvolvimento das Ciencias e da Tecnologia (FDCT), under grant No. 016/2011/A and Research Committee at University of Macau under grant No. MRG002/LIC/2012. The authors are very grateful to all the technical staff from the Research & Laboratory Center and the Operations Department in Macao Water Supply Co. Ltd., who provided full supports to this research.

Reference

[1] R.E. Carlson, A trophic state index for lakes, Limnol. Oceanogr. 22(2) (1977) 361–369.

[2] H.W. Paerl, J. Huisman, Climate: Blooms like it hot Science, 320(5872) (2008) 57–58.

[3] T.W. Davis, D.L. Berry, The effects of temperature and nutrients on the growth and dynamics of toxic and non-toxic strains of Microcystis during cyanobacteria blooms, Harmful Algae 8(5) (2009) 715–725.

[4] Y. Li, J. Ma, Z. Yang, I. Lou, Influence of non-point source pollution on water quality of Wetland Baiyangdian, China, Desalin. Water Treat. 32(1–3) (2011) 291–296.

[5] L. Håkanson, J.M. Malmaeus, Coefficients of variation for chlorophyll, green algae, diatoms, cryptophytes and blue-greens in rivers as a basis for predictive modelling and aquatic management, Ecol. Modell. 169(1) (2003) 179–196.

[6] M. Moreira-Santos, A.M. Soares, A phytoplankton growth assay for routine *in situ* environmental assessments, Environ. Toxicol. Chem. 23(6) (2004) 1549–1560.

[7] E. Briand, N. Escoffier, Spatiotemporal changes in the genetic diversity of a bloom-forming Microcystis aeruginosa (cyanobacteria) population, ISME J. 3(4) (2009) 419–429.

[8] S.A. Abdul-Wahab, C.S. Bakheit, S.M. Al-Alawi, Principle component and multiple regression analysis in modeling of ground-level ozone and factors affecting its concentrations, Environ. Model. Softw. 20 (2005) 1263–1271.

[9] H. Çamdevýren, H.N. Demýr, A. Kanika, S. Keskyn, Use of principal component scores in multiple linear regression models for prediction of Chlorophyll-a in reservoirs, Ecol. Modell. 181(4) (2005) 581–589.

[10] S.H. Te, K.Y.H. Gin, The dynamics of cyanobacteria and microcystin production in a tropical reservoir of Singapore, Harmful Algae 10(3) (2011) 319–329.

[11] W. Zhang, I. Lou, W.K. Ung, Y. Kong, K.M. Mok, Eutrophication in Macau Main Storage Reservoir, 12th International Conference on Environmental Science and Technology, Rhodes island, Dodecanisos, Greece, September 8–10 (2011) 1114–1121.

[12] APHA, Standard methods for the examination of water and wastewater. A.W.W.A.a.W.E.F., American Public Health Association, 1995.

[13] J. Pallant, I. Chorus, Toxic cyanobacteria in water, SPSS Survival Manual (2007).

[14] J. Stevens, Applied multivariate statistics for the social science, New Jersey, Hill Sdale, USA, 1986, pp. 515.

[15] N.R. Draper, H. Smith, Applied regression analysis, third ed., John Wiley and Sons Inc., New York, pp. 706 1998.

[16] A.M. Geraldes, M.J. Boavida, Limnological variations of a reservoir during two successive years: One wet, another dry, Lakes Reservoirs: Res. Manage. 9(2) (2004) 143–152.

[17] D.W. Schindler, Evolution of phosphorus limitation in lakes, Science 195(4275) (1977) 260–262.

[18] C. Nalewajko, T.P. Murphy, Effects of temperature, and availability of nitrogen and phosphorus on the abundance of Anabaena and *Microcystis* in Lake Biwa, Japan: An experimental approach, Liminology 2 (2001) 45–48.

[19] B.K. Basu, F.R. Pick, Longitudinal and seasonal development of planktonic chlorophyll a in the Rideau River, Ontario, Can. J. Fish. Aquat. Sci. 52 (1995) 804–815.

[20] N.G. Bulgakov, A.P. Levich, The nitrogen to phosphorus ratio as a factor regulating phytoplankton community structure, Archiv fur Hydrobiologia 146 (1999) 3–22.

[21] M.E. Kim, T.S. Shon, K.S. Min, H.S. Shin, Forecasting performance of algae blooms based on artificial neural networks and automatic observation system, Desalin. Water Treat. 38(1–3) (2012) 293–301.