Oceanological and Hydrobiological Studies

International Journal of Oceanography and Hydrobiology

Volume 47, Issue 1, March 2018
pages (1-9)

ISSN 1730-413X
eISSN 1897-3191

Predicting chlorophyll-*a* concentrations in two temperate reservoirs with different trophic states using Principal Component Regression (PCR)

by

Kemal Çelik

DOI: **10.1515/ohs-2018-0001**
Category: **Original research paper**
Received: **July 27, 2017**
Accepted: **September 04, 2017**

Faculty of Biology, Balıkesir Universty, 1050 Balıkesir, Turkey

Abstract

Relationships between chlorophyll-*a* (chl-*a*) concentrations and 16 physicochemical variables in temperate eutrophic Çaygören and mesotrophic Ikizcetepeler reservoirs (Turkey) were determined using Principal Component Analysis (PCA). PCA was used to simplify the complexity of relationships between water quality variables. Principal component scores (PCs) were used as independent variables in the multiple linear regression analysis (MLR) to predict chl-*a* in both reservoirs. This procedure is called Principal Component Regression (PCR). In the eutrophic Çaygören Reservoir, chl-*a* was significantly ($p < 0.05$) correlated with nitrite-nitrogen (NO_2), ammonium-nitrogen (NH_4), phosphate (PO_4), total suspended solids (TSS), pH, Secchi disk transparency, total dissolved solids (TDS) and total phosphorus (TP). In the mesotrophic Ikizcetepeler Reservoir, chl-*a* was significantly ($p < 0.05$) correlated with TSS, NO_2, chemical oxygen demand (COD), sulfate (SO_4), TDS, pH and the Secchi disk. In the eutrophic Çaygören Reservoir, six PCs explained 71% of the total variation in the water quality, while in the mesotrophic Ikizcetepeler Reservoir, six PCs explained 75% of the variation. This study has shown that PCR is a more robust tool than direct MLR to simplify the relationships between water quality variables and to predict chl-*a* concentrations in temperate reservoirs with different trophic states.

Key words: Chlorophyll-*a*, eutrophic reservoir, oligotrophic reservoir, principal component analysis

* Corresponding author: *kcelik@balikesir.edu.tr*

DE GRUYTER

The *Oceanological and Hydrobiological Studies* is online at oandhs.ocean.ug.edu.pl

Introduction

The intensified human activities have accelerated the rate of cultural eutrophication of water bodies worldwide (Najar & Khan 2012), resulting in cyanobacterial blooms, reduced oxygen and eventually fish kills (Anita & Pooja 2013). The effects of these activities on water bodies can be estimated by the degree of eutrophication. The eutrophication level is often determined by the chlorophyll-*a* (chl-*a*) content in the water column of aquatic systems (Huszar et al. 2006).

Chl-*a* concentration, a proxy for the abundance of phytoplankton in aquatic systems, depends on the physical and chemical properties of the water column, including the availability of nutrients, light, temperature and grazing pressure as well as interactions between these variables (Jeppesen et al. 2004). Reduction in nutrient loading results in a reduction in phytoplankton biomass and this, in turn, results in better ecological status of water bodies (Phillips et al. 2008).

Statistical methods are widely used in eutrophication studies to explore the interactions between water quality parameters and primary production (Irvine et al. 2015). However, it is difficult to study eutrophication in surface waters when there are a large number of variables. In these cases, researchers face a major difficulty regarding the choice of analytical parameters due to the multicollinearity (Johnson et al. 1982).

One way to overcome this problem is to use principal component analysis (PCA) followed by multiple linear regression (MLR) (Çamdevirena et al. 2005). The process is called principal component regression, PCR, where the principal components (PCs) of PCA are used in MLR. This process reduces the complexity of the multidimensional system by maximizing the variance of component loadings and by eliminating the invalid components (Zhang et al. 2013).

The PCR method has recently been used in various studies to identify factors driving the eutrophication of aquatic systems (Praveena et al. 2011). PCR provides information on the most meaningful variables by reducing the number of latent factors contributing to eutrophication, MLR examines the relationships between a dependent variable and a set of independent variables.

The aim of this study is to apply PCR to data from two temperate freshwater reservoirs with different trophic states. Studies using PCR methodology from this region have so far used data sets either from a single lentic (Çamdevirena et al. 2005) or lotic system (Köklü et al. 2010). The presented work also aims at determining whether the results of PCR application to a eutrophic reservoir are different from those for a mesotrophic reservoir. For this purpose, 17 physicochemical and biological water quality parameters from eutrophic temperate Çaygören and mesotrophic Ikizcetepeler reservoirs were used in PCA, and then the PCs of PCA were used in MLR to predict chl-*a* concentrations in both reservoirs.

Materials and methods

Study area

In Turkey, the most important problem of agriculture is the lack of water for irrigation during summer, because it has a Mediterranean climate that is extremely dry in summer. One of the effective solutions to this problem is the construction of dams on the lotic waters. The Çaygören Reservoir was built in 1971 with the objective to irrigate the Sındırgı and Bigadiç plains. It is also used for power generation. The land cover around the reservoir is mostly black pine (*Pinus nigra*). Sport fishing in the reservoir is also popular. Therefore, the reservoir is important to local and regional economic and ecological sustainability (Arslan & Ergül 2014).

The Çaygören Reservoir is located at 39°172′N, 28°191′E, 55 km southeast of Balıkesir, Turkey (Fig. 1). It lies at 273 m above sea level. It has a maximum depth of 28 m, a length of 4.6 km and a surface area of 9 km^2. The Simav Stream feeds the reservoir (State Water Works 2017).

The Ikizcetepeler Reservoir was built in 1991 for irrigation and drinking water. It is mainly used as a drinking water source of the city of Balıkesir. It is used for irrigation of Pamukçu Plain and for sport fishing.

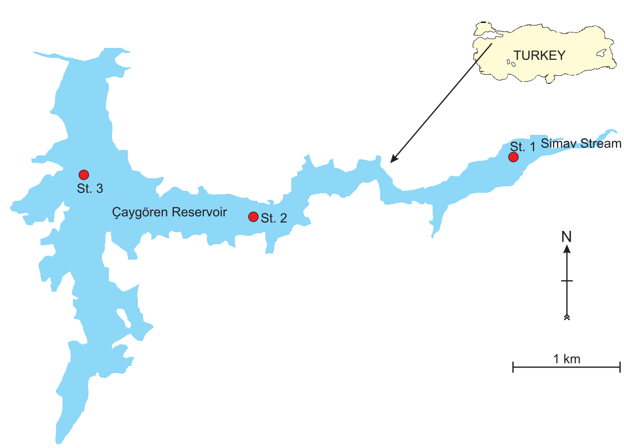

Figure 1

Çaygören Reservoir and the location of sampling stations

Therefore, it is socioeconomically important for the region (Okkan & Karakan 2016).

The Ikizcetepeler Reservoir is located at 39°482'N, 27°939'E, 15 km southwest of Balıkesir, Turkey (Fig. 2). It lies at 175 m above sea level. It has a maximum depth of 25 m, a surface area of 10 km^2 and a length of 6.34 km. The reservoir is mainly fed by the Kile Stream (State Water Works 2017).

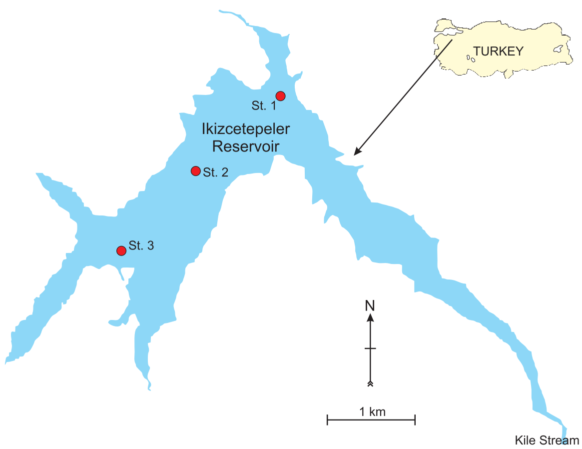

Figure 2

Ikizcetepeler Reservoir and the location of sampling stations

Sampling techniques and methods

Three sampling stations were set in each reservoir. The first station was set near the main inlet of each reservoir, the second station – between the first station and the dam, and the third station – near the dam. Sampling was initiated in February 2007 and ended in January 2009.

Chl-*a*, water temperature, pH, total dissolved solids (TDS), conductivity and oxidation-reduction potential (ORP) were measured in situ using a YSI multi-probe. For alkalinity, total suspended solids (TSS), nitrite-nitrogen (NO_2–N), nitrate-nitrogen (NO_3–N), ammonium-nitrogen (NH_4–N), total nitrogen (TN), total phosphorus (TP), phosphate (PO_4), chemical oxygen demand (COD), biochemical oxygen demand (BOD) and sulfate (SO_4), samples were drawn at 1, 5, 10 and 15 m, then stored in pre-cleaned 250 ml amber glass bottles and transferred to a laboratory for analysis.

Chl-*a* was measured periodically according to the trichromatic method (Arar 1997) from the acetone-extracted samples to check the accuracy of the probe. TP concentrations were determined from non-filtered water as orthophosphate after persulfate-acid hydrolysis at 135°C for 2 hours. TN concentrations were determined from water samples after digestion by the Kjeldahl method (APHA 1995).

TSS were determined by filtering a known volume of water through Whatman 934-AH filters that were pre-rinsed and dried (105°C) and then tared (Okkan & Karakan 2016). PO_4, NO_2–N, NO_3–N, NH_4–N concentrations were determined spectrophotometrically on filtered water. The COD was measured by the Open Reflux Method. BOD was determined by Wrinkle's Azide Modification Titrimetric Method. Secchi disk depth was measured on each sampling date.

Statistical methods

Statistical analyses were carried out using the SPSS (ver. 11.0) software package. Data were log-transformed prior to statistical analysis to meet the requirements of normality for parametric tests. The Kolmogorov-Smirnov normality test was applied to all variables. PCA was performed on 17 water quality parameters to describe their interrelation patterns.

The Kaiser-Meyer-Olkin (KMO) measure of sample adequacy and Bartlett's test of sphericity {with degrees of freedom = 1/2 [p (p – 1)], where p is the number of variables}, were used to verify the applicability of PCA. PC scores of 17 water quality parameters were used as independent variables in MLR to predict the chl-*a* levels (Stevens 1986).

The stepwise variable selection procedure of MLR was used to identify the best predictors of chl-*a*, excluding nonsignificant PC scores in the model. The t-test method was used to test the significance of regression coefficients in the MLR models. This test was used to check if the selected coefficients (PCs) in the regression models were significantly different from each other. Determination coefficient (R^2) values were used as predictive success criteria of the models.

The score value (sk_j) for the jth observation in the kth PC was obtained from the weight of variables in PCs and standardized variables by using the following equation:

where j = 1, 2, …, n is the number of observations,

$$sk_j = t_{1k} z_{1j} + t_{2k} z_{2j} + \ldots + t_{pk} z_{pj}$$

k = 1, 2, …, q is the number of selected PCs, p is the number of independent variables, sk_j is the standardized score value of the jth observation in the kth PC, tpk is the standardized weight of the pth variable in the kth PC, zp_j is the standardized value of the pth variable of the jth observation, calculated from

$$x = \frac{x_p - \overline{x}}{s_x}$$

where x_p is the original value of the pth variable.

Results

χ^2 calculated as 702.7 for the Çaygören Reservoir and 1223 for the Ikizcetepeler Reservoir by Barttlett's sphericity test (d.f. = 136, $p < 0.01$), showing that both data sets have an adequate number of samples for the application of PCA. The residual statistics are given in Table 1 and Table 2 for each reservoir.

Table 1

Residual statistics (n = 17) for Çaygören Reservoir

	Mean	Standard deviation	Minimum	Maximum
Predicted value	1.26	0.66	0.28	2.52
S.E. of predicted value	0.40	0.11	−1.47	1.89
Residual	0.00	1.176	−2.35	2.93
Standard residual	0.00	0.97	−1.91	2.41
Cook's distance	0.06	0.07	0.01	1.61

Table 2

Residual statistics (n = 17) for Ikizcetepeler Reservoir

	Mean	Standard deviation	Minimum	Maximum
Predicted value	1.11	0.53	−0.49	1.81
S.E. of predicted value	0.19	0.08	0.15	0.50
Residual	0.00	0.59	−0.70	1.80
Standard residual	0.00	0.96	−1.15	2.96
Cook's distance	0.08	0.16	0.00	0.57

In both reservoirs, six out of 17 PCs were selected for MLR, which was called the first approach (Tables 3 and 4). In the first approach, the selected PCs explained 71% of the total variation in the Çaygören Reservoir and 75% in the Ikizcetepeler Reservoir (Tables 5 and 6).

Table 3

Descriptive statistics of PCs for Çaygören Reservoir

Principle component	Eigenvalue	Cumulative variance (%)
PC1	4.0	23.8
PC2	2.2	36.8
PC3	1.8	47.7
PC4	1.3	55.8
PC5	1.2	63.4
PC6	1.0	69.4
PC7	0.9	75.1
PC8	0.8	79.7
PC9	0.7	83.7
PC10	0.6	87.4
PC11	0.5	90.3
PC12	0.4	92.6
PC13	0.3	94.6
PC14	0.3	96.3
PC15	0.3	97.8
PC16	0.2	99.1
PC17	0.15	100

Table 4

Descriptive statistics of PCs for Ikizcetepeler Reservoir

Principle component	Eigenvalue	Cumulative variance (%)
PC1	3.9	23.2
PC2	2.5	38.0
PC3	1.9	49.1
PC4	1.6	58.6
PC5	1.4	66.7
PC6	1.2	73.8
PC7	0.9	79.1
PC8	0.8	83.7
PC9	0.7	87.6
PC10	0.5	90.5
PC11	0.4	93.1
PC12	0.3	94.9
PC13	0.3	96.5
PC14	0.3	98.1
PC15	0.2	99.3
PC16	0.1	99.9
PC17	0.01	100

PCA component loadings of the first six PCs for the Çaygören Reservoir are presented in Table 7. The bold-marked loads indicate the highest existing correlation between the variables and the corresponding component. The high values of communalities indicate that the variance was efficiently reflected in the regression analysis.

All 17 variables were included in the six selected PCs. However, only certain variables had significant loads within each PC, such as NO_2–N had the most significant loads in PC1, NH_4–N in PC2, PO_4 and TSS in PC3, pH in PC4, chl-*a* and Secchi disk transparency in PC5 and TDS and TP in PC6.

The component loadings of the first six PCs for the Ikizcetepeler Reservoir are given in Table 8. In this reservoir, TSS and NO_2–N had the most significant loads in PC3, COD and SO_4 in PC4, TDS and pH in PC5 and chl-*a* and Secchi disk depth in PC6.

According to MLR, where all PCs were used as independent variables (called the second approach), only PC scores 7, 10 and 12 were found to have significant linear relationships with chl-*a* in the Çaygören Reservoir (Table 9). In the Ikizcetepeler Reservoir, when all PCs were included in the regression analysis, only PC scores 1 and 12 had significant linear relationships with chl-*a* (Table 10).

As can be seen in Table 5, 71% of the variation in chl-*a* was explained by PC3 in the Çaygören Reservoir. If the five unselected PCs were included in the model, the coefficient of determination would rise to 75%. However, this difference is not statistically significant. Chl-*a* levels were strongly correlated with PC score 3. In other words, the chl-*a* level would be expected to increase as the scores of PC3 increase. Consequently,

Table 5

Results of regression analysis (n = 17) with six PCs for Çaygören Reservoir

Included independent variables	Regression coefficient (bk)	Standard error of bk	Standardized regression coefficient	t	P	R^2 (%)
Constant	1.2	0.34		3.73	0.002	71
Score 3	1.32	0.75	0.4	2.56	0.05	

Table 6

Results of regression analysis (n = 17) with six PCs for Ikizcetepeler Reservoir

Included independent variables	Regression coefficient (bk)	Standard error of bk	Standardized regression coefficient	t	P	R^2 (%)
Constant	1.01	0.18		5.55	0.00	75
Score 1	1.90	0.89	0.501	2.95	0.048	

Table 7

Results of principal component analysis for Çaygören Reservoir

Variables in	Standardized weight of variables in selected PC (t_{ik}; i = 1, 2, …, 16 and k = 1, 2, 3, 4, 5)						Loading of variables (v_{ik})						Communalities
	1	2	3	4	5	6	1	2	3	4	5	6	
pH	−0.49	−0.06	−0.45	4.88	−1.61	−0.74	0.03	−0.05	0.04	**0.48**	−0.05	0.01	0.73
SO_4	−2.21	−3.48	3.37	−2.60	−0.93	2.45	−0.12	0.04	0.32	−0.15	−0.02	0.25	0.73
ORP	1.21	−1.93	−0.44	0.91	1.27	1.80	0.07	-0.28	−0.02	0.05	0.07	0.16	0.49
Secchi	−1.11	−0.88	−2.46	0.62	4.53	−0.42	0.03	0.02	−0.08	0.19	**0.57**	0.11	0.75
Temp	4.51	1.61	−2.32	−0.89	0.89	−0.71	0.29	−0.34	−0.08	−0.03	0.09	0.01	0.85
Chl	−0.17	2.74	0.03	2.81	−4.51	0.97	−0.04	0.06	−0.02	0.29	**−0.57**	0.08	0.78
TN	−1.65	3.43	1.01	−1.37	−1.86	−0.64	−0.04	0.36	0.12	−0.02	−0.05	0.02	0.71
NO_3	−2.21	1.12	−2.65	3.01	0.45	−1.26	−0.05	0.34	−0.09	0.36	0.16	0.02	0.61
NO_2	−4.48	6.69	−0.69	2.42	0.95	0.21	**−0.31**	0.04	−0.09	0.09	0.01	−0.03	0.85
NH_4	−1.95	0.97	−0.82	0.34	−0.19	−1.99	−0.18	**0.44**	−0.03	0.05	0.04	−0.11	0.78
TP	0.21	−0.24	2.62	1.07	−0.31	−4.57	−0.04	0.02	0.12	0.03	−0.08	**−0.44**	0.61
PO_4	−0.06	−0.03	4.59	−1.92	−1.41	−1.80	0.06	0.04	**0.38**	−0.06	−0.06	−0.05	0.59
Cond	2.13	−1.42	−2.28	−3.55	1.38	2.37	0.12	−0.02	−0.17	−0.23	0.07	0.13	0.68
TSS	−1.37	−3.11	4.01	1.65	−2.42	0.52	0.02	0.01	**0.31**	0.18	−0.03	0.12	0.45
BOD	3.19	−4.30	2.25	−2.31	0.54	−0.63	**0.31**	−0.07	0.25	−0.01	0.16	0.09	0.81
TDS	−0.73	−1.08	−0.04	−0.97	−1.29	4.73	0.05	−0.14	0.13	0.02	−0.09	**0.68**	0.71
COD	3.42	−1.72	0.15	1.21	−3.67	0.59	0.25	−0.18	0.01	0.09	−0.25	0.05	0.72

Table 8

Results of principal component analysis for Ikizcetepeler Reservoir

Variables in	Standardized weight of variables in selected PC (t_{ik}; I = 1, 2, …, 16 and k = 1, 2, 3, 4, 5)						Loading of variables (v_{ik})						Communalities
	1	2	3	4	5	6	1	2	3	4	5	6	
pH	−4.36	7.67	3.30	3.93	4.09	4.98	0.02	0.24	0.11	0.12	**0.33**	0.13	0.69
SO_4	−5.16	−4.87	14.36	16.21	3.01	4.71	−0.01	-0.09	0.28	**0.32**	0.00	0.06	0.63
ORP	−1.26	−1.26	1.06	−2.48	1.97	-0.22	−0.13	-0.11	0.09	−0.22	0.07	0.17	0.57
Secchi	−0.08	−0.93	−1.92	2.23	−3.77	−0.22	−0.03	−0.09	−0.19	0.22	**0.33**	**−0.37**	0.74
Temp	3.55	1.86	−3.35	3.09	0.89	0.85	**0.24**	0.08	−0.15	0.14	−0.09	0.04	0.86
Chl	0.33	0.34	0.42	−3.05	4.21	0.42	0.08	0.03	0.04	−0.**33**	−0.01	0.**46**	0.74
TN	3.41	−2.59	0.83	−0.81	0.05	0.019	0.19	−0.14	0.04	−0.04	−0.02	0.03	0.73
NO_3	2.49	−4.29	5.07	−1.13	1.41	2.39	0.17	−0.15	0.18	−0.04	0.31	0.05	0.93
NO_2	−1.84	−0.92	7.68	4.44	2.15	0.53	−0.05	−0.03	0.31	0.17	**−0.35**	0.08	0.74
NH_4	1.76	−3.14	4.44	−0.58	−0.13	1.55	0.17	−0.16	0.22	-0.03	0.27	−0.07	0.94
TP	0.71	4.77	0.67	−2.61	−2.28	0.45	0.07	0.29	0.04	−0.16	0.06	−0.14	0.72
PO_4	0.59	4.99	1.70	−1.96	−2.32	0.54	0.06	0.**29**	0.09	−0.11	−0.02	−0.13	0.69
Cond	5.31	−2.67	−5.70	0.68	1.03	−0.82	0.18	−0.11	−0.23	0.02	−0.11	0.04	0.82
TSS	−0.39	−0.84	−3.54	−1.12	2.96	−0.11	−0.04	−0.07	**−0.30**	−0.09	0.24	0.25	0.58
BOD	−0.67	0.13	−1.17	2.36	3.58	0.72	0.01	0.02	−0.16	0.**32**	−0.07	**0.48**	0.71
TDS	−2.34	1.21	0.48	1.68	2.49	1.37	−0.15	0.13	0.05	0.18	**0.36**	0.27	0.72
COD	−0.67	0.13	−1.17	2.36	3.58	0.72	0.01	0.02	−0.16	0.**32**	−0.07	**0.48**	0.71

Table 9

Results of regression analysis (n = 16) with all PCs for Çaygören Reservoir

Included independent variables	Regression coefficient (bk)	Standard deviation of bk	t	*p*	R^2 (%)
Constant	1.14	0.3	3.79	0.02	76.8
Score 7	0.77	–2.2	1.8	0.049	
Score 10	3.42	1.56	2.2	0.041	
Score 12	0.8	0.5	2.08	0.045	

Table 10

Results of regression analysis (n = 16) with all PCs for Ikizcetepeler Reservoir

Included independent variables	Regression coefficient (bk)	Standard deviation of bk	t	*p*	R^2 (%)
Constant	1.06	0.15	6.97	0.004	80.3
Score 1	0.57	1.2	1.97	0.049	
Score 12	–5.81	1.72	–3.39	0.042	

a total increase in significant variables in PC3, namely, TSS and PO_4 would lead to an increase in the chl-*a* level.

In the Ikizcetepeler Reservoir, 75% of the variation in chl-*a* was explained by PC1. If the five unselected PCs were included in the model, the coefficient of determination would rise to 80.3%. However, this difference is not statistically significant. Chl-*a* levels were strongly correlated with PC score 1. In other words, the chl-*a* level would be expected to increase as the scores of PC1 increase. Consequently, a total increase in significant variables in PC1, namely, water temperature, TN and BOD would lead to an increase in the chl-*a* level.

In the Çaygören Reservoir, six variables that had significant impacts on PC scores 1, 2, 4, 5 and 6, i.e. pH, chl-*a*, TDS, TP, NH_4–N and NO_2–N, were excluded from the regression model of the first approach. However, linear effects of these variables on chl-*a* were partially incorporated in the model, since these variables were also included in PC3. In the Çaygören Reservoir, chl-*a* values of the first approach were calculated by the following formula:

$$\text{Chl-}a = 1.2 + 1.325 \text{ (score 3)}$$

In the Ikizcetepeler Reservoir, eleven variables that had significant impacts on PC scores 2, 3, 4, 5 and 6, i.e. pH, chl-*a*, TDS,TSS, TP, PO_4, Secchi disk transparency, NH_4–N, NO_3–N, NO_2–N and SO_4, were excluded from the regression model of the first approach. However, linear effects of these variables were partially incorporated in the model since they were also included in PC1. In this reservoir, the chl-*a* level of the first approach was calculated by the following formula:

$$\text{Chl-}a = 1.09 + 1.9 \text{ (score 1)}$$

Results of the residual analysis were used to verify the applicability of the assembled regression models in both reservoirs. The existences of influential and outlier observations were checked in the models (Tables 1 and 2). Values of chl-*a* predicted (modeled) using regression models and the observed values are given in Table 11 and 12 for each reservoir.

In the Çaygören Reservoir, 71% of the variation in the water quality parameters was explained by the PC3 scores, while 75% of the variation in the Ikizcetepeler Reservoir was explained by the PC1 scores, implying that chl-*a* could be reliably estimated by using the scores of PCs in the MLR modeling approach.

Chl-*a* was further predicted by the second approach where scores of all PCs were included in the stepwise MLR. In the Çaygören Reservoir, PC scores 7, 10 and 12 were found to be significantly correlated with chl-*a* (Table 9). In the first approach, none of these scores were found to be significantly correlated with chl-*a* ($p > 0.05$), while in this approach they were found to be significantly correlated with chl-*a* ($p = 0.004$).

In the Ikizcetepeler Reservoir, when scores of all PCs were included in MLR (second approach), PC scores 1 and 12 were found to be significantly correlated with chl-*a* (Table 10). In the first approach, PC scores of 12 were not found to be significantly correlated with chl-*a* ($p > 0.05$), while in the second approach they were found to be significantly correlated with chl-*a* ($p = 0.004$).

The main difference between the PCR and the classic MLR is that zero correlations exist between the scores and thus the problem of multi-collinearity is eliminated in PCR. The regression coefficient of PC scores obtained through this approach is presented in Table 9 for the Çaygören Reservoir. The determination coefficient of this model was found to be 76.8% and the following model was used to predict chl-*a*:

$$\text{Chl-}a = 1.14 + 0.77 \text{ (PC score 7)} + 3.42 \text{ (PC score 10)} + 0.8 \text{ (PC score 12)}$$

In the Ikizcetepeler Reservoir, the regression

Table 11

Observed and predicted Chlorophyll-*a* values for Çaygören Reservoir

Observations	Observed values (mg l^{-1})	First approach predicted values (mg l^{-1})	Second approach predicted values (mg l^{-1})
1	1.75	0.7462	1.5680
2	2.04	2.4665	2.0199
3	2.00	1.2119	1.8576
4	1.21	0.9917	1.1429
5	1.26	0.6538	0.9731
6	1.09	0.621	0.8199
7	0.99	1.4281	0.9911
8	1.11	0.5003	1.2811
9	0.82	1.2926	0.6984
10	0.99	0.9912	0.8931
11	1.66	1.5972	1.8327
12	1.23	2.2135	1.2346
13	0.96	0.9209	1.2511
14	5.45	1.8412	5.4829
15	0.55	1.6093	0.392
16	1.01	1.6114	1.1065
17	1.04	1.5	1.34

Table 12

Observed and predicted Chlorophyll-*a* values for Ikizcetepeler Reservoir

Observations	Observed values (mg l^{-1})	First approach predicted values (mg l^{-1})	Second approach predicted values (mg l^{-1})
1	1.62	1.5204	1.420
2	0.99	0.8411	0.790
3	–0.70	0.7428	–0.550
4	0.92	0.6361	0.780
5	0.94	2.0858	0.9100
6	3.34	2.0732	3.200
7	0.69	0.9732	0.6300
8	0.74	0.8894	0.6200
9	0.79	0.9133	0.6800
10	1.16	0.8023	1.000
11	1.71	1.6937	1.500
12	1.17	0.7573	1.110
13	1.10	1.3245	0.980
14	1.27	1.0466	1.150
15	0.95	0.881	0.830
16	1.05	0.559	0.900
17	1.03	0.98	1.100

coefficients of PC scores obtained through the PCR approach are presented in Table 10. The determination coefficient of this model was found to be 80.3% and the following model was used to predict chl-*a*:

$$\text{Chl-}a = 1.06 + 0.57\ (\text{PC score 1}) + 0.8\ (\text{PC score 12})$$

Discussion

In the eutrophic Çaygören Reservoir, six PCs explained 71% of the total variation in the relationships of water quality variables, while in the mesotrophic Ikizcetepeler Reservoir, the same number of components explained 75% of the total variation. These results show that PCR is a suitable tool for showing the relationships between the water quality variables in lakes with different trophic states (Chang et al. 2012).

In the Çaygören Reservoir, all variables were included in the six selected PCs. However, only nine variables (NO_3–N, NO_2–N, NH_4–N, PO_4, TSS, pH, Secchi disk depth, TDS and TP) had significant loads within these PCs. The results showed that phosphorus and nitrogen ions were more effective on primary production than the other factors in the eutrophic

Çaygören Reservoir. The research has shown that in eutrophic lakes nitrogen limitation occurs in a short time, usually in early summer, followed by the appearance of nitrogen-fixing Cyanobacteria, after which the algal community returns to phosphorus limitation (Schindler 2012).

In the mesotrophic Ikizcetepeler Reservoir, all variables were included in the six selected PCs. However, only seven variables (TSS, NO_2–N, COD, SO_4, TDS, pH and Secchi disk depth) had significant loads within these PCs. These results showed that in the mesotrophic Ikizcetepeler Reservoir, mostly water transparency (light availability) was effective on primary production rather than phosphorus and nitrogen. High loadings of TSS and Secchi disk depth into the selected PCs suggest that chl-*a* was mostly controlled by the variation in light intensity in this mesotrophic reservoir. Light intensity is an important factor contributing to seasonal differences in primary production (chl-*a* concentration) in aquatic ecosystems (Karlsson et al. 2009).

In the eutrophic Çaygören Reservoir, when all 16 PC scores were included in the stepwise MLR to predict chl-*a* (the second approach), PC scores 7, 10 and 12 were selected as significant factors. None of these scores were selected in the first approach. In the mesotrophic Ikizcetepeler Reservoir, when all 16 PC scores were included in the regression analysis (the second approach), PC scores 1 and 12 were selected for the prediction of chl-*a*. In the first approach, PC score 12 was not selected. This is due to lower degrees of freedom in the models. Researchers argue that the degree of freedom for stepwise procedures is close to the number of candidate predictors (Hakanson et al. 2003). Since only six PCs were used in the regression analysis in the first approach, the low degree of freedom restricted the selection of parameters, even though they had high loads within the excluded PCs.

In the eutrophic Çaygören Reservoir, the determination coefficient (R^2) was 71% for the first approach's model and 76.8% for the second approach's model. In the mesotrophic Ikizcetepeler Reservoir, the determination coefficient was 75% for the first approach's model and 80.3% for the second approach's model. There is a 5% rise in R^2 in both reservoirs, but this rise was not significant. R^2 values are used as criteria for the predictive success of the regression models (Steyerberg et al. 2001). Generally, univariate simple regression models can yield high R^2 values, but they may obscure the complexity of the natural ecosystems. The predictive power of the present models show that the multidimensional approach for predicting chl-*a* levels is reliable in temperate reservoirs.

The results of this study show that PCR is an appropriate tool for predicting chl-*a* levels in mesotrophic and eutrophic temperate reservoirs. Furthermore, the results showed that by using the PCR approach, it was possible to reduce the number of analytical parameters from 16 to 3 in the eutrophic Çaygören Reservoir and from 16 to 2 in the mesotrophic Ikizcetepeler Reservoir to accurately estimate chl-*a* levels. Finally, the results show that PCR is a more robust tool than direct MLR for predicting chl-*a* concentrations and concurrently showing the difference between trophic states in temperate reservoirs.

Acknowledgements

This study was supported by Balıkesir University Research Foundation (Project No. 2007/18). The author thanks anonymous reviewer(s) for their helpful comments and Tuğba Ongun Sevindik for help in the fieldwork.

References

Anita, B. & Pooja, D. (2013). Water quality guidelines for the management of pond fish culture. *Int. J. Environ Sci.* 3: 1980–2009. DOI: 10.6088/ijes.2013030600019.

APHA. (1995). *Standard Methods for the Examination of Water and Wastewater* (19th ed.). Washington, DC: American Public Health Association.

Arar, E.J. (1997). *Determination of chlorophylls a and b and identification of other pigments of interest in marine and reshwater algae using high performance liquid chromatography with visible wavelength detection. Method 447*. Washington, DC: U.S. Environmental Protection Agency.

Arslan, F. & Ergül, M. (2014). Agricultural activities in the Çaygören Dam Irrigation Area and Sourounding. JASSS 2(1): 171–190. (In Turkish). Retrieved from: http://www.asosjournal.com/DergiTamDetay.aspx?ID=63&Detay=Ozet

Chang, N.B., Yang, Y.J., Daranpob, A., Jin, K.R. & James, T. (2012). Spatiotemporal pattern validation of chlorophyll-*a* concentrations in Lake Okeechobee, Florida, using a comparative MODIS image mining approach. Int. J. Remote. Sens. 33: 2233–2260. DOI: 10.1080/01431161.2011.608089.

Çamdevirena, H., Demir, N., Kanika, A. & Keskin, S. (2005). Use of principal component scores in multiple linear regression models for prediction of Chlorophyll-*a* in reservoirs. *Ecol. Model.* 181: 581–589. DOI: 10.1016/j.ecolmodel.2004.06.043.

Hakanson, L., Malmaeus, J.M., Bodemer U. & Gerhardt, V. (2003). Coefficients of variation for chlorophyll, green algae,

DE GRUYTER

diatoms, cryptophytes and blue-greens in rivers as a basis for predictive modeling and aquatic management. *Ecol. Model.* 169: 179–196. DOI: 10.1016/S0304-3800(03)00269-2.

Huszar, V.L., Caraco, N.F., Roland, F. & Cole, J. (2006). Nutrient-chlorophyll relationships in tropical-subtropical lakes: do temperate models fit? *Biogeochemistry* 79: 239–250. DOI: 10.1007/s10533-006-9007-9.

Jeppesen, E., Jensen, J. P., Søndergaard, M., Fenger-Grøn, M., Bramm, M.E. et al. (2015). Empirical evaluation of the conceptual model underpinning a regional aquatic long-term monitoring program using causal modeling. *Ecol. Indic.* 50: 8–23. DOI: 10.1016/j.ecolind.2014.10.011.

Johnson R.A. & Wichern, D.W. (1982). *Applied Multivariate Statistical Analysis*. Englewood Cliffs: Prentice-Hall Inc.

Karlsson, J., Byström, P., Ask, J., Ask, P., Persson, L. et al. (2009). Light limitation of nutrient-poor lake ecosystems. *Nature* 460: 506–509. DOI: 10.1038/nature08179.

Köklü, R., Sengörür, B. & Topal, B. (2010). Water Quality Assessment Using Multivariate Statistical Methods-A Case Study: Melen River System (Turkey). *Water. Resour. Manag.* 24: 959–978. DOI: 10.1007/s11269-009-9481-7.

Møller, P.H. & Rasmussen, H.U. (2004). Impact of fish predation on cladoceran body weight distribution and zooplankton grazing in lakes during winter. *Freshwater. Biol.* 49: 432–447. DOI: 10.1111/j.1365-2427.2004.01199.x.

Najar, I.A. & Khan, A.B. (2012). Assessment of water quality and identification of pollution sources of three lakes in Kashmir, India, using multivariate analysis. *Environ. Earth. Sci.* 66: 2367–2378. DOI: 10.1007/s12665-011-1458-1.

Okkan, U. & Karakan, E. (2016). Modeling the Effects of Climate Change on Ikizcetepeler Inflows: 2015–2030 Projection. *IMO Teknik Dergi* 450: 7379–7401. (In Turkish). Retrieved from http://dergipark.gov.tr/tekderg/issue/28138/299000

Phillips, G., Pietilainen O.P., Carvalho, L., Solimini, A., Solheim, A.L. et al. (2008). Chlorophyll-nutrient relationships of different lake types using a large European dataset. *Aquat. Ecol.* 42: 213–226. DOI: 10.1007/s10452-008-9180-0.

Praveena, S.M., Kwan, O.I. & Aris, A.Z. (2011). Effects of data pre-treatment procedures on principal component analysis: a case study for mangrove surface sediment datasets. *Environ. Monit. Assess.* 184: 6855–6868. DOI: 10.1007/s10661-011-2463-2.

Schindler, D. (2012). The dilemma of controlling cultural eutrophication of lakes. Proc. R. Soc. Lond. B. Biol. Sci. 7: 4322–4332. DOI: 10.1098/rspb.2012.1032.

State Water Works. (2017). Çaygören *Barajı* (In Turkish). http://www2.dsi.gov.tr/baraj/detay.cfm?BarajID=32

Stevens, J. (1986). *Applied Multivariate Statistics for the Social Science*. New Jersey: Hillsdale.

Steyerberg, E.W., Harrell, F.E. & Habbema, J.D. (2001). Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. *Med. Decis. Making* 21: 45–56. DOI: 10.1177/0272989X0102100106.

Zhang, W., Lou, I.C., Kong, Y., Ung, W.K. & Mok, M. (2013). Eutrophication analysis and principal component regression for two subtropical storage reservoirs in Macau. *Desalin. Water. Treat.* 51: 7331–7340. DOI: 10.1080/19443994.2013.793921.