Journal of Oceanology and Limnology
Vol. 38 No. 6, P. 1771-1786, 2020
https://doi.org/10.1007/s00343-019-9223-5

Establishment of a diatom-total phosphorus transfer function for lakes on the Songnen Plain in northeast China*

SUI Fengyang[1], ZANG Shuying[1, **], FAN Yawen[2, **], LU Xinxin[2], HUI Hongkuan[3]

[1] *Key Laboratory of Remote Sensing Monitoring of Geographic Environment, College of Heilongjiang Province, Harbin Normal University, Harbin 150025, China*

[2] *Key Laboratory of Biodiversity of Aquatic Organisms, Harbin Normal University, Harbin 150025, China*

[3] *Qilu Normal University, Jinan 250200, China*

Received Sep. 7, 2019; accepted in principle Oct. 11, 2019; accepted for publication Oct. 18, 2019

Abstract To understand the relationship between planktonic diatoms and environmental variables in the lakes on the Songnen Plain, northeast (NE) China, we investigated water quality and planktonic diatoms from 71 sampling points in 27 lakes, based on which Canonical Correspondence Analysis (CCA) was conducted. The result show that planktonic diatoms displayed certain responses to environment gradients, and the total phosphorus (TP) explained the maximum variation of planktonic diatom species among the 15 environmental variables, suggesting that TP was the most important and significant environmental parameter affecting the distribution of diatom species. In addition, we established a diatom-total phosphorus transfer (DTPT) function, of which component 5 of the weighted averaging partial least squares regression (WA-PLS) was chosen to and compared with a series of weighted average regression (WA) models and WA-PLS models. According to the jackknifing statistical test, the component 5 of WA-PLS models provided a lower root-mean-square error prediction (RMSEP=0.202) and a higher correlation coefficient between observation and prediction (R^2_Jack=0.759). After deletion of three outliers, the root-mean-square error prediction of the DTPT function was reduced (RMSEP=0.169) while the correlation coefficient between observation and prediction was increased (R^2_Jack=0.823). Therefore, this DTPT function performed better than other regional TP models in the world. However, it remains demanding for expanding the background dataset to improve the prediction ability of the model.

Keyword: Songnen Plain; diatom; total phosphorus; transfer function

1 INTRODUCTION

Lakes are sensitive aquatic ecosystems (Witak et al., 2017). The lakes on the Songnen Plain constitute the main part of those in the Northeast Plain of China, where the landscape characteristics are closely related to their geographical environment. These lakes are characteristic of small size, shallow lake water, gentle slope, deep and thick modern sediments, and high salinity, and they are formed due to recent crustal subsidence, low terrain, poor drainage, and river shift (Wang and Dou, 1998). In recent years, increasing salinization and water pollution because of the combined effects of surface runoff recharge, discharge methods, climate, topography, and human disturbance have posed great threats to the lakes with relatively fragile ecosystems. According to some studies (Wei et al., 2001; Xiao et al., 2011; Zhou et al., 2016; Dai et al., 2018), the potassium permanganate index, total phosphorus (TP) content, and total nitrogen (TN) content in the lakes have generally been higher on the Songnen Plain in recent years, leading to an intensification in eutrophication, thus resulting in the deterioration of lake water quality along with changes in the aquatic ecosystem structure in this area. In view of the reality of the ecological environment of the lakes on the Songnen Plain, it is an urgent need for us

* Supported by the National Natural Science Foundation of China (Nos. 41571199, 41030743, 31870187)

** Corresponding authors: zsy6311@163.com; fanyaw@163.com

to study the mechanism of water quality change in this area and to explore the relationship between the biological community and the environment, so as to reveal the change trend of the ecological environment of the lakes in this area.

At present, the history of water quality monitoring for most lakes in the region is too short to reflect the natural changes in the past (Xiao et al., 2011; Dai et al., 2018). However, lake sediments can provide abundant biological and physicochemical information that can be used to reconstruct the history of past changes in lakes and their basins (Battarbee, 1999; Dong et al., 2006). Among them, diatom fossils are regarded as the most important biological indicators in sediment core (Liu et al., 2012). Diatoms which are characterized by abundant species, wide distribution, and rapid reproduction are able to respond quickly to environmental changes, are especially suitable for revealing the stories of environmental events. After the diatoms die, their siliceous valves are preserved in bottom of the lake. Due to their strong dissolution resistance, they are preserved well in sediments. Therefore, diatom fossils can provide a wealth of environmental information for different geological periods. With the support of dating technique, the fossil diatom assemblages can be used to quantitatively reconstruct a variety of environmental variables (Qi and Li, 2004).

Over the past two decades, regional diatom-environment transfer functions based on weighted average have been established consecutively in parts of Africa, Europe, Australia, America, and Asia, with temperature (Wunsam and Schmidt, 1995; Bigler and Hall, 2002; Krawczyk et al., 2017; Luostarinen et al., 2017), water level (Duthie et al., 1996; Zhuang et al., 2014; Ma, 2018), pH (Enache and Prairie, 2002; Chen et al., 2008; Finkelstein et al., 2014), salinity (Fritz et al., 1991; Battarbee et al., 2002; Huang et al., 2013; Xu, 2018), and nutrients (TP, TN) as the main environmental variables for quantitative reconstruction (Andrén et al., 2017; Craig, 2018; Liu et al., 2018). In most of these studies, the modern datasets were created based on the diatoms collected from surface sediments, although some were created based on benthic diatoms in a minority of studies (Winter and Duthie, 2000; Denicola et al., 2004; Huang et al., 2013). By comparing diatom assemblages in surface sediments with recent planktonic diatoms in Lago Maggiore, Marchetto and Musazzi (2001) found that one sediment sample might be a good estimate of the planktonic diatoms composition (Marchetto and Musazzi, 2001). The study carried out by Jones and Juggins (1995) demonstrated that many benthic species did not respond directly to nutrient concentrations in water (Jones and Juggins, 1995). Reavie et al. (2014) investigated the phytoplankton and water quality in the Laurentian Great Lakes, and established the diatom-total phosphorus transfer function using phytoplankton diatoms, and then, when the transfer function was applied to a sediment core of Lake Superior, the former TP content was effectively reconstructed (Reavie et al., 2014). In some studies, it had also been found that plankton species accounted for a large proportion in modern datasets established based on diatom assemblages in surface sediments (Miettinen, 2003; Tammelin et al., 2017). Therefore, the modern background dataset was created directly by using the species of planktonic diatoms and the corresponding environmental factors in the present study.

The Songnen Plain lake group is one of the lake regions with the highest lake density in China (Wang and Dou, 1998). But so far, there have been few reports on the research work of the quantitative reconstruction of the paleoenvironment based on diatoms in this area. The content of this paper was constructed according to the main environmental problems faced by the Songnen Plain lake group, namely, the aggravation of salinization and the eutrophication of the lakes in this area due to intensive human activity (Yao et al., 2010; Xiao et al., 2011; Dai et al., 2018). Based on the establishment of a regional modern diatom-water environment background dataset, through multivariate statistical analysis, the relationship between diatoms and environmental factors was studied, the most significant environmental variable (TP) affecting diatom species was extracted, and then the diatom-total phosphorus transfer function was established, which laid a foundation for the quantitative reconstruction of historical nutrition level and the study on eutrophication mechanism for the lakes in this area in the future.

2 MATERIAL AND METHOD

2.1 Overview of sampling points

The Songnen Plain is located in the northern part of the Northeast Plain in China and is formed as a result of erosion and alluvial sedimentation from the Nenjiang and Songhua Rivers. It is bounded by the Songliao watershed in the south, adjoining to the

Lesser Khingan Mountains in the north, bordering the eastern mountains and the Greater Khingan Mountain Range in the east and west, respectively. The southwestern part of Heilongjiang Province, the central and western parts of Jilin, and the eastern part of Inner Mongolia (eastern foot of Greater Khingan Range Mountain) are under its administrative jurisdiction. The climate in this area is a temperate continental semi-humid and semi-arid climate (Yao et al., 2010), cold and dry in winter, warm and humid in summer. Its average annual precipitation is 400–500 mm, and the precipitation in June-September accounts for 70%–80% of annual total precipitation (Wang and Dou, 1998; Yao et al., 2010).

The lake group on the Songnen Plain is a low-lying lake group in the mid-temperate sub-humid region of China, covering a lake region area of about 42 750 km^2, a lake area of 2 570 km^2, and a lake coverage rate of 6% (Yao et al., 2010; Xiao et al., 2011). The lakes in this area are mainly distributed along the Songhua River, the Nenjiang River and its tributaries. During the flood season, river water naturally flows into the lake as supplement source. The existence of lake waters plays a regulation role in the surrounding climate, so that the climate around the lake area is relatively humid (Xiao et al., 2011).

2.2 Sample collection and preparation

From June to August 2008, the phytoplankton and water samples in the lakes on the Songnen Plain were collected in the field. A total of 27 lakes were surveyed, each lake was visited on a single day. Seventy-one sampling points were distributed based on water area, topography and surrounding environment for these 27 lakes, in which 1–5 sampling points were set for lakes with an area less than 100 km^2 and 5–10 sampling points were set for lakes with an area of 100–500 km^2, S1 to S71 represents the 71 sampling points respectively in the paper (Fig.1). These lakes are all shallow lakes with a depth of 1–4 m in the wet season, among which Chagan Lake has the largest area of 307.0 km^2 with average depth of 2.5 m, while the rest are mostly small and medium-sized lakes. Many of the lakes were rich in aquatic plants, with swampy meadows or reed marshes along the lakes at the time of sampling in 2008.

The samples were collected from each site at a water depth of 0.5 m with water samplers, pH value, water temperature (WT), and conductivity of the 71 sampling points were measured in situ with a portable water quality analyzer. The water samples for the measurement of environmental parameters were packed into polypropylene bottles with a volume of 5 L and stored on site in an incubator (with pre-frozen ice packs) for refrigeration. Eight main ions (K^+, Na^+, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, CO_3^{2-} and HCO_3^-), TN, and TP were measured at Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences. The water samples were titrated with calibrated dilute hydrochloric acid to obtain HCO_3^-, CO_3^{2-} of the samples, filtered with 0.45-μm GF/F filter membrane, and the filtrate was used for the determination of cations (K^+, Na^+, Ca^{2+}, Mg^{2+}) and anions (Cl^-, SO_4^{2-}). TP content was determined by the ammonium molybdate spectrophotometry. TN content was determined by the alkaline potassium persulfate spectrophotometry. The sampling and analysis methods of chemical parameters followed the relevant provisions of the *Monitoring and Analysis of Water and Wastewater* (SEPA, 2002). Total dissolved solids (TDS) was expressed by the sum of 8 major ion concentrations (Wetzel, 2001; Yang, 2004), and total alkalinity (TA) was expressed in the sum of CO_3^{2-}+ HCO_3^- concentration (Huang et al., 2000).

The samples for the analysis of planktonic diatoms were preserved using Lugol's solution (15-mL Lugol's solution was added into per 1 000-mL water sample). For quantitative analyses of phytoplankton communities, the mixed water was collected at each site and precipitated for 24–36 h, the final volume was set to 30 mL of concentrated sediments, and then part of each sample was made to be permanent slides through acid treatment and centrifugation (Zhang and Huang, 1991; Hu and Wei, 2006). The diatoms were observed, identified, and counted under an Olympus (BH) optical microscope (×100 oil immersion lens and ×10 eyepiece). Diatom cells with a damaged area not exceeding 1/4 in the field of view were identified and counted. The number of diatom valves counted in each sample was not less than 300. Finally, the relative abundance of each species was recorded as a percentage. Identification of taxa followed the taxonomic guide from Krammer and Lange-Bertalot (Krammer and Lange-Bertalo, 1997a, b, 2004a, b), as well as the taxonomic records from several websites (Kociolek et al., 2019; Potapova et al., 2019).

Moreover, a pilot experiment about the sedimentary diatoms in the study area had been conducted. A 40-cm long columnar core was collected from Chagan Lake in the research area, preliminary identification of diatoms in the sedimentary core revealed that more than 90% of the diatom species in the sedimentary

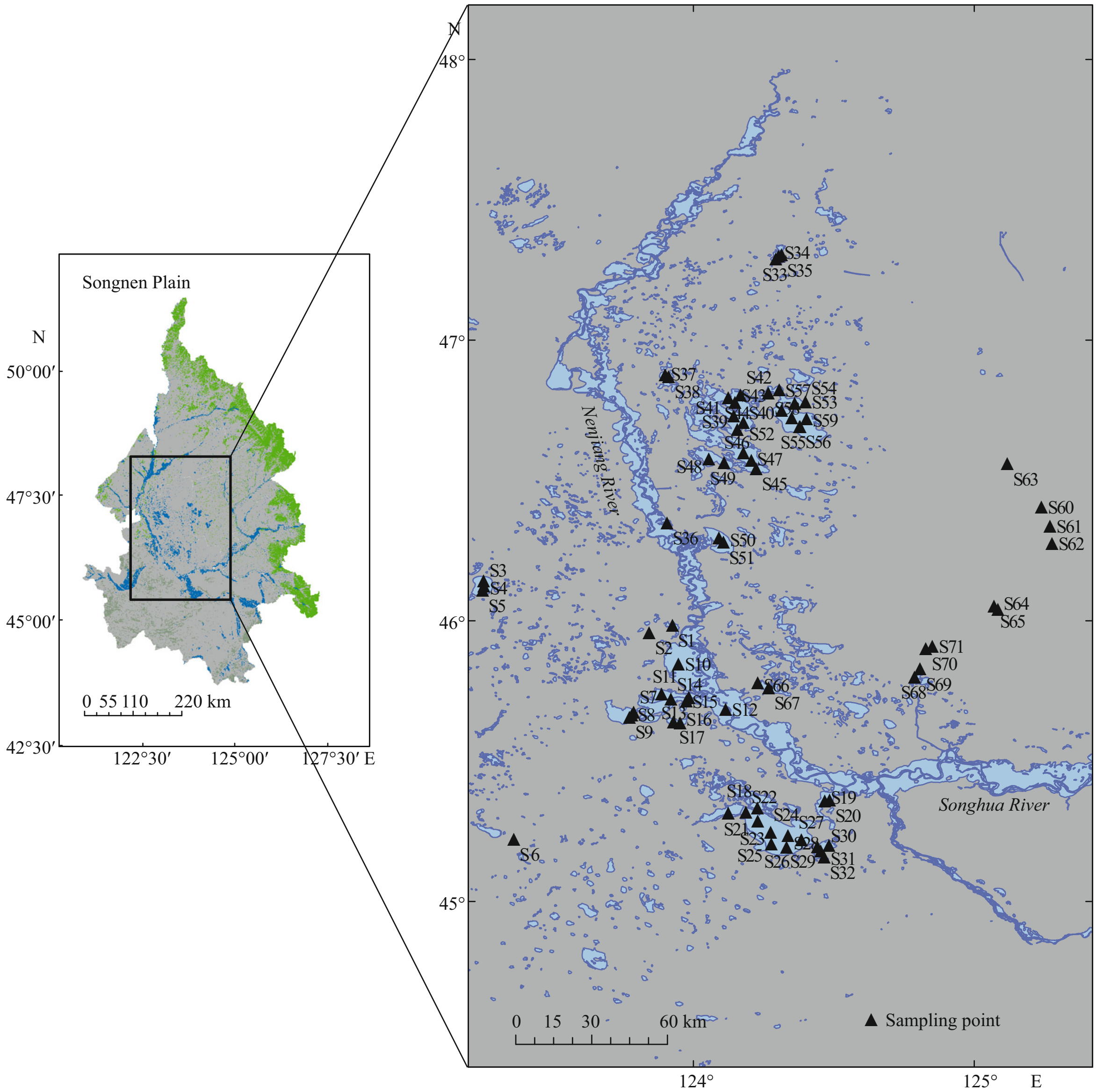

Fig.1 Distribution diagram of sampling points of the lakes on the Songnen Plain

core were found in the dataset of plankton diatoms collected in 2008. This suggested that it was feasible to establish the background dataset with plankton diatoms.

2.3 Data processing

2.3.1 Data standardization

The data was standardized before analysis. The data on diatom species which at least appeared at 2 sampling points and presented at more than 1% in at least one sample were selected, to which content square root transformation should be made; the environmental parameters, except pH, were lg(x+1) transformed to form the documents required for statistical analysis (Dong et al., 2006).

2.3.2 Canonical Correspondence Analysis (CCA)

CCA is a gradient analysis and weighted average regression technique based on ecological theory, which can reflect the relationship between species, sampling points and environmental variables on the coordinate axis (Ter Braak, 1987). In the present study, the analysis was carried out according to the following steps: (1) detrended Correspondence Analysis (DCA) was used to derive the unimodal response of the species (standard-deviation units: SD). If the value of SD is greater than 3, then CCA analysis is performed (Ter Braak and Prentice, 1988); (2) a preliminary CCA analysis on a dataset including all environmental variables was conducted, and those environmental variables with variance inflation factor

(VIF) greater than 20 (Ter Braak, 1988) were selectively deleted; (3) a series of constrained Canonical Correspondence Analyses (cCCA) constrained only to one environmental variable at a time was performed to test the marginal effect of each environment variable (Ter Braak and Verdonschotp, 1995). The importance of the environmental variables in the order of the variance in the individual interpretation of the species data was sequenced, and the significance of the interpretation was test by using the Monte Carlo technique (P<0.05) (Ter Braak and Šmilauer, 1998). Meanwhile, the ratio of the first eigenvalue ($\lambda1$) to the second eigenvalue ($\lambda2$) was calculated. The environment variable with $\lambda1/\lambda2>0.5$, showing its significance and importance, could be used to establish the transfer function (Winter and Duthie, 2008); (4) finally, Detrended Canonical Correspondence Analysis (DCCA) constrained only to one environmental variable at a time was performed to determine whether the dataset was appropriate for the establishment of nonlinear or linear regression models, and the response pattern of the diatoms to the variable was determined according to the effective gradient length of the environment variable on the first axis. When the gradient length is ≥2 SD, the nonlinear weighted average model is used to establish the diatom-environment indicator transfer function, or otherwise the linear model is applied (Ter Braak and Juggins, 1993; Birks, 1995).

All of the above sequencing analysis was performed by using the CANOCO software (Version 4.5) for Windows 4.5 (Ter Braak and Šmilauer, 1998).

2.3.3 Establishment of the transfer function

The weighted average regression method (WA) (Birks et al., 1990) is mainly used for the nonlinear response mode. In this method, since the initial derivation values are underestimated after being weighted twice in the calculation, they should be restored in order to obtain the accurate derivation values. According to the different ways of regression and whether the weights of the species tolerance are reduced or not, it can be divided into Classical WA and Inverse WA; and the Classical WAtol and Inverse WAtol with weights reduction of the species tolerance. Another more complicated method is the most prevalent weighted averaging partial least squares regression (WA-PLS) (Ter Braak and Juggins, 1993) which improves the derivation ability of the transfer function by multiple extraction and optimization of the residual information amount from the previous function based on the weighted average regression. In contrast, WA is relatively simple in terms of theoretical concept and algorithm, and WA-PLS could be regarded as an optimization and progression of WA method (Zhuang et al., 2004). However, the models created by the WA are not necessarily poorer than those created by WA-PLS. The derivation ability of a model is mainly measured by the root-mean-square error of prediction (RMSEP) and the correlation coefficient (R^2) between the observed values and the predicted values. The former indicates the actual derivation prediction error of the transfer function, and the lower the value, the stronger the function's derivation ability; the latter indicates the fitting degree between the analog value and the observed value, and the higher the value, the stronger the function derivation ability. The above statistics were mainly tested by jackknifing, which is commonly used in diatom quantification now (Dong et al., 2006). The above transfer functions were created and tested by using the program of C2 version 1.6.6 (Juggins, 2003).

Table 1 Measurements of water environment of 71 sampling points

Environmental variable	Maximum	Minimum	Mean±SD
Conductivity (μs/cm)	5 522	253	1 396.77±1 244.39
TP (mg/L)	3.30	0.05	0.39±0.60
TN (mg/L)	8.14	0.38	3.92±1.80
WT (°C)	29.35	20.81	24.73±1.87
pH	9.81	7.80	8.92±0.37
Cl^- (mg/L)	942.60	2.65	97.31±150.03
SO_4^{2-} (mg/L)	297.76	1.64	62.73±71.95
Ca^{2+} (mg/L)	69.47	4.07	18.41±11.52
Mg^{2+} (mg/L)	42.04	7.00	19.39±7.99
K^+ (mg/L)	13.23	1.35	4.85±2.40
Na^+ (mg/L)	1 228.09	23.37	265.36±264.28
CO_3^{2-} (mg/L)	333.39	2.98	76.65±89.87
HCO_3^- (mg/L)	1 567.06	100.91	549.16±366.37
TDS (mg/L)	4 312.9	221.1	1 093.85±869.58
TA (mg/L)	1 869.9	146.4	625.81±443.80

3 RESULT

3.1 Characteristics of water environment

The statistical values (maximum, minimum and average) of the water environmental variables at the 71 sampling points on the Songnen Plain lake group are shown in Table 1. The conductivity of the study area ranged from 253 to 5 522 μs/cm. Among the eight major ions, the average value of K^+ was the lowest, 4.85 mg/L, and the average values of HCO_3^- and Na^+ were higher, 549.16 and 265.36 mg/L respectively, appearing in the typical soda-type water features. TDS and TA were calculated from the eight major ions, so their variation range was related to the content of eight major ions in each sample point. In the nutritional indexes, there were some gradient changes in TP and TN, and the contents of TP and TN in the 71 sampling points were 0.05–3.30 and 0.38–8.14 mg/L respectively, in accordance with the National Surface Water Environmental Quality Standard of the People's Republic of China (Wu and Zhang, 2009); if taking TP as a reference, the water qualities of the study area were all above Class III (≤0.05 mg/L), and if taking TN as a reference, the water qualities were all above Class I (≤0.2 mg/L). To sum up, the water qualities of the study area belonged to a mesotrophic to eutrophic type. The WT was the only physical variable, which varied from 20.81 to 29.35°C, the variation range of WT was not very large since all samples were taken in summer; and the pH value reflected a weak alkaline environmental condition of the study area as a whole.

3.2 Distribution of diatom species

According to the taxonomic guide of Krammer and Lange-Bertalot, 168 diatom taxa were identified in 71 sampling points of the lakes on the Songnen Plain, including 144 species, 21 varieties, and 3 forms; 88 diatom taxa (appeared at least 2 sampling points and its content was more than 1% in at least one sample) were used to establish the modern diatoms dataset.

The species with a relative abundance of more than 5% were defined as the dominant species (Pokras and Molfino, 1986), and the species that were dominant species at least one sampling point in the modern diatom dataset were selected, to study their assemblage changes. The samples were arranged from high to low in terms of TP content (Fig.2), with the decrease of TP content, the diatom assemblage changed. The obvious assemblage change occurred at S52, taking S52 as a boundary, the TP content above S52 was 0.33–3.30 mg/L, the dominant species of diatoms included *Aulacoseira granulata* (Ehrenberg) Simonsen, *Cyclotella meneghiniana* Kützing, and *Nitzschia palea* (Kützing) W. Smith. And the TP content below S52 (including S52) was 0.05–0.32 mg/L, in the diatom samples corresponding to this TP range, *Aulacoseira granulata*, *Aulacoseira ambigua* (Grunow) Simonsen, *Cyclotella meneghiniana*, *Diploneis elliptica* (Kützing) Cleve, *Ulnaria acus* (Kützing) M. Aboal, *Staurosirella pinnata* (Ehrenberg) D. M. Williams & Round, *Nitzschia palea*, and *Surirella peisonis* Pantocsek dominated. With the decrease of TP content, the diatom assemblage showed a change from relatively simple to relatively complex composition of the dominant species. *Aulacoseira granulata*, *Aulacoseira ambigua*, and *Cyclotella meneghiniana* became the dominant species in the study area since they were present in most sampling points, accounted for a large percentage and were often found in eutrophic water, meanwhile, they could tolerate a certain amount of salinity, which is consistent with the salinization of water environment in the study area.

Table 2 Result of preliminary CCA for 15 environmental variables

CCA axes	1	2
Eigenvalues	0.329	0.241
Species-environment correlations	0.923	0.884
Cumulative percentage variance of species data	6.3	10.9
of species-environment relation	17.1	29.7

3.3 CCA result

The DCA result of the data on the diatoms collected from the 71 sampling points showed that the cumulative variance value of the species data explained by the first two axes was 16.9%, and the maximum value of the gradient length in the first two axes was 5.426, which was greater than 3. This indicates that the diatom species had a distinct unimodal response relationship with respect to the first two axes; thus, CCA is suitable for exploring the relationship between diatom community structure and environment.

The preliminary CCA constrained to 15 environmental variables was run, the analytical result show that the cumulative percentage variance of species data obtained from the first two axes was 10.9%, and axis 1 (R=0.923) and axis 2 (R=0.884) show a significant species-environment correlation (Table 2). The first two axes reflected 29.7% of the cumulative percentage variance of species-environment relation and explained 36.63% of total eigenvalues of the data on diatoms.

The CCA result also reveal that the environmental variables with a variance inflation factor (VIF) greater than 20 were conductivity, Na^+, HCO_3^-, salinity (TDS)

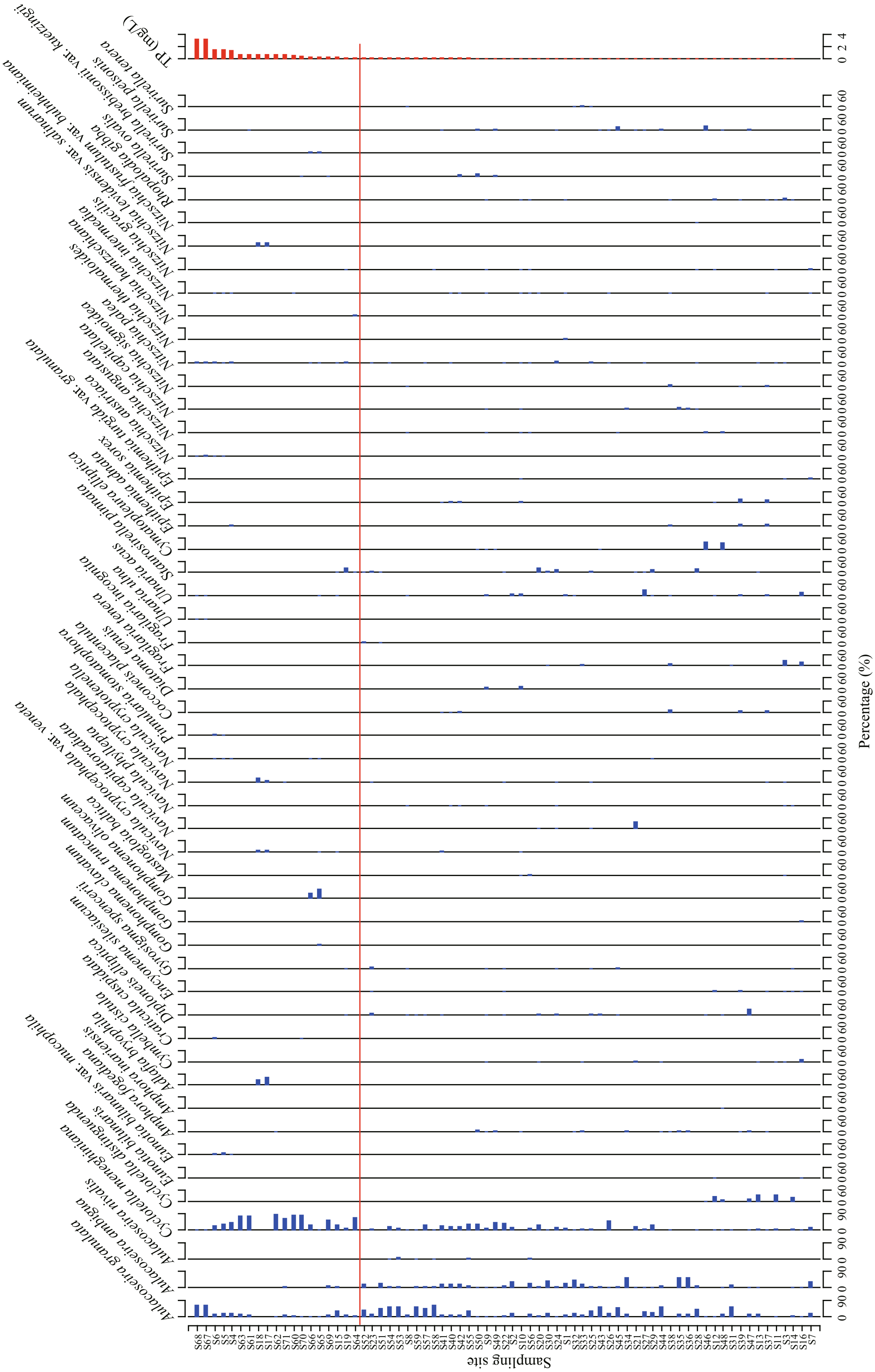

Fig.2 Distribution of dominant species of planktonic diatoms in the lakes on the Songnen Plain

Table 3 Analytical results of marginal effect of 10 environmental variables

Environment variable	$\lambda 1$	$\lambda 2$	Variance interpretation	P value
TP	0.274	0.499	5.2	0.002
SO_4^{2-}	0.241	0.499	4.6	0.002
Cl^-	0.205	0.501	3.9	0.002
K^+	0.204	0.498	3.9	0.002
CO_3^{2-}	0.190	0.498	3.6	0.002
Mg^{2+}	0.168	0.500	3.2	0.002
pH	0.163	0.500	3.1	0.002
Ca^{2+}	0.152	0.497	2.9	0.004
WT	0.142	0.492	2.7	0.002
TN	0.100	0.498	1.9	0.086

and total alkalinity (TA), indicating that the typical correlation coefficients between the five environmental variables and the axis were unstable and they were not suitable for further explanation (Birks et al., 1990). After deletion of the abovementioned variables, the remaining 10 environmental variables were used for the subsequent marginal effect test.

A series of cCCA constrained to a single environmental variable at a time was performed, i.e. the marginal effect analysis, the results showed that TP was the most important environmental variable, explaining 5.2% of the information on distribution of diatom species (Table 3). As for the eigenvalue ratio ($\lambda 1/\lambda 2$) of the 10 environmental variables, the eigenvalue ratio of TP was greater than 0.5, the eigenvalue ratio of SO_4^{2-} was approximately equal to 0.5, and the others were all less than 0.5. In addition, among the ten environmental variables, TN did not pass the significance test ($P>0.05$), while TP, SO_4^{2-}, Cl^-, K^+, CO_3^{2-}, Mg^{2+}, pH, Ca^{2+} and WT values were able to explain separately diatom data, with 95% confidence level ($P<0.05$). The further CCA analysis result of these 9 environmental variables significantly explained 25.31% of total eigenvalue, the cumulative percentage variance of species data obtained from the first two axes was 9.7%, the cumulative variance reflected 38.2% of the species-environment relation, an improvement over 29.7% for 15 environmental variables.

In the CCA ordination biplot (Fig.3) constrained to the 9 environmental variables, the TP had a good correlation ($R=0.824$) with the first axis; SO_4^{2-}, Cl^-, CO_3^{2-} and pH were correlated with the first axis ($R=0.651$, 0.487, 0.409 and 0.325 respectively) as well as with the second axis ($R=0.476$, 0.397, 0.046 and 0.155 respectively); Ca^{2+} and K^+ were mainly correlated with the negative half axis of the second axis ($R=-0.313$ and -0.519 respectively); WT and Mg^{2+} had poor correlation with the above two axes, whose impacts on the data could be ignored. Fig.3 also better reflected the relationship between the diatom species and the environmental variables (through the arrangement of the projection points of these species on environmental indexes). In the forward direction of TP with the maximum correlation with the first axis, some distributed species, such as *Nitzschia austriaca* Hustedt, *Pinnularia microstauron* var. *brebissonii* (Kützing) Mayer, *Craticula cuspidata* (Kützing) D. G. Mann, *Gomphonema parvulum* (Kützing) Kützing, *Nitzschia levidensis* var. *salinarum* Grunow and *Pseudostaurosira brevistriata* (Grunow) D. M. Williams & Round, were all nutrient-tolerant species. And in the forward direction of pH, some basophilic diatom species were distributed, such as *Nitzschia levidensis* var. *salinarum*, *Navicula pygmaea* Kützing, *Cyclotella meneghiniana* and *Aulacoseira granulata*, which was consistent with the characteristics of the alkaline water quality in the study area. The arrows for the three environmental variables of SO_4^{2-}, Cl^-, CO_3^{2-} were in the same quadrant, some species, such as *Gomphonema clavatum* Ehrenberg, *Surirella brebissonii* var. *kuetzingii* Krammer & Lange-Bertalot, *Pinnularia subcapitata* W. Gregory, *Navicula cryptocephala* var. *veneta* (Kützing) Rabenhorst, and *Hippodonta capitata* (Ehrenberg) Lange-Bertalot, Metzeltin & Witkowski, were distributed along the forward direction of these three variables, showing that they had a positive correlation with the three environmental variables. In the lower part in the figure with higher K^+ and Ca^{2+} values, the relative abundance of the distributed species, such as *Surirella ovalis* Brébisson, *Cocconeis placentula* Ehrenberg, *Epithemia adnata* (Kützing) Brébisson, *Fragilaria incognita* Reichardt, *Pinnularia maior* (Kützing) Cleve and *Caloneis amphisbaena* (Bory de Saint Vincent) Cleve, increased with the increase of the concentration of K^+ and Ca^{2+}. On the other hand, the species distributed along the reverse direction of these environmental variables in the figure showed a negative correlation with these environmental variables respectively.

It could be concluded that TP was the most significant and important one out of all selected environmental variables according to the marginal effect results and the relationship between each environmental variable and the first axis in CCA

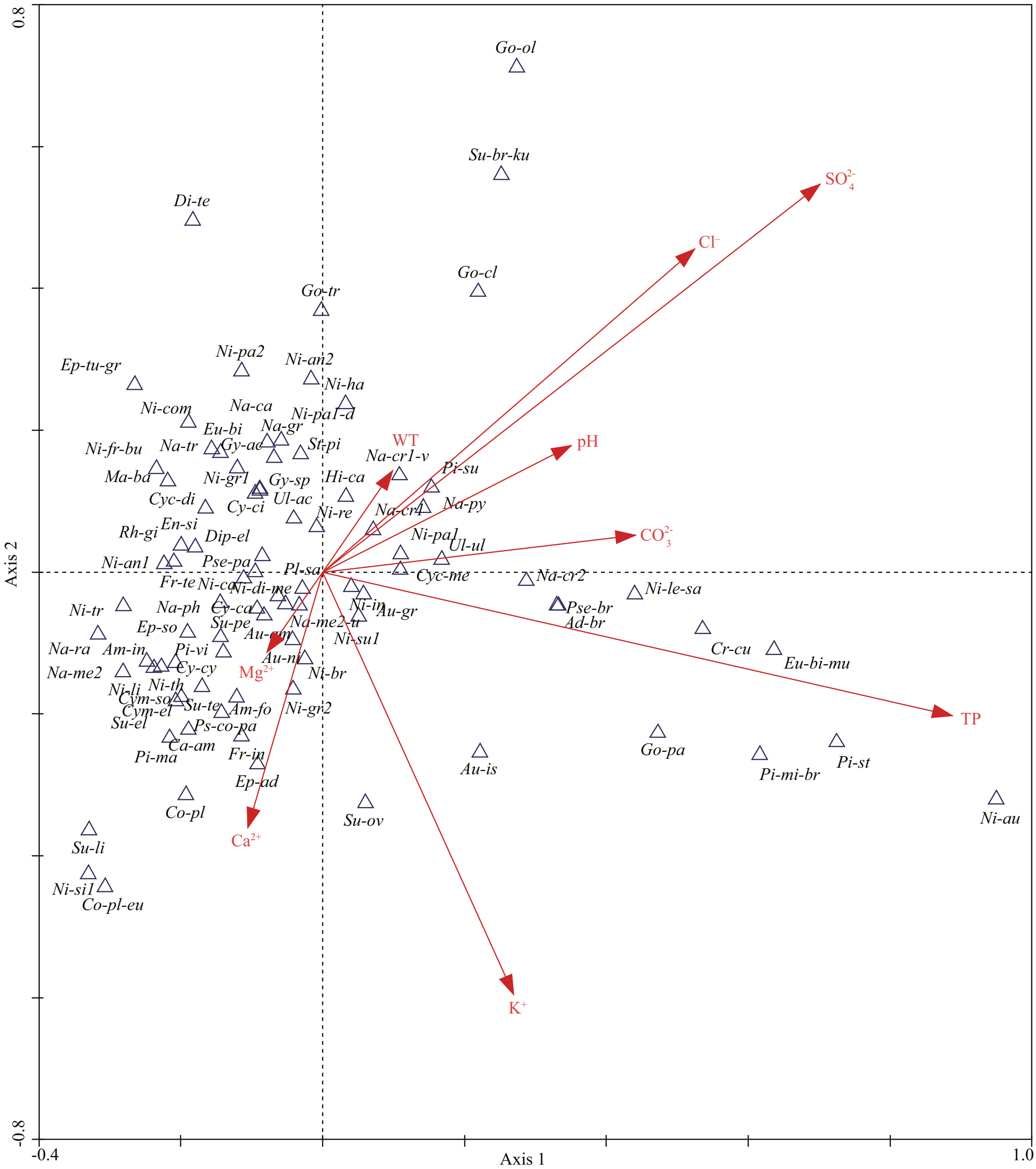

Fig.3 CCA ordination biplot of 9 environmental variables and planktonic diatom species

analysis, thus it had been selected as the target factor of the transfer function.

DCCA constrained only to TP was run, the effective gradient length of the first axis was analyzed and tested to be 2.576 for the data collected from 71 sampling points, which is greater than the critical value of the unimodal response (2.0 SD), indicating a unimodal response pattern between the diatom distribution and TP; thus, the diatom-total phosphorus transfer function might be created with the nonlinear weighted average model.

3.4 Establishment of diatom-total phosphorus transfer function

The statistical test results of the data collected from 71 sampling points modeled by various WA and WA-PLS methods are shown in Table 4 and five components were included into the results of the WA-PLS transfer function model. When modeling with these 71 sampling points, WA-PLS component 5 showed the highest correlation coefficient (R^2_Jack), the smallest maximum bias and smaller root-mean-

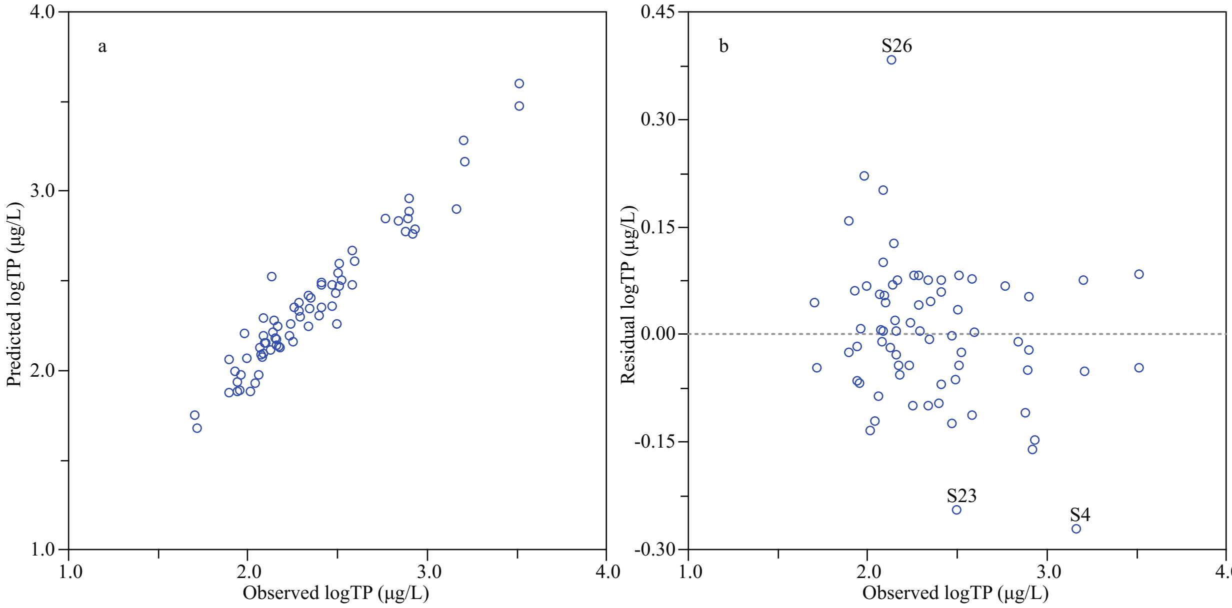

Fig.4 Observed versus predicted annual mean TP concentrations (log μg/L) (a), and observed versus residual annual mean TP concentrations (log μg/L) (b), based on WA-PLS (component 5) model of 71 sites

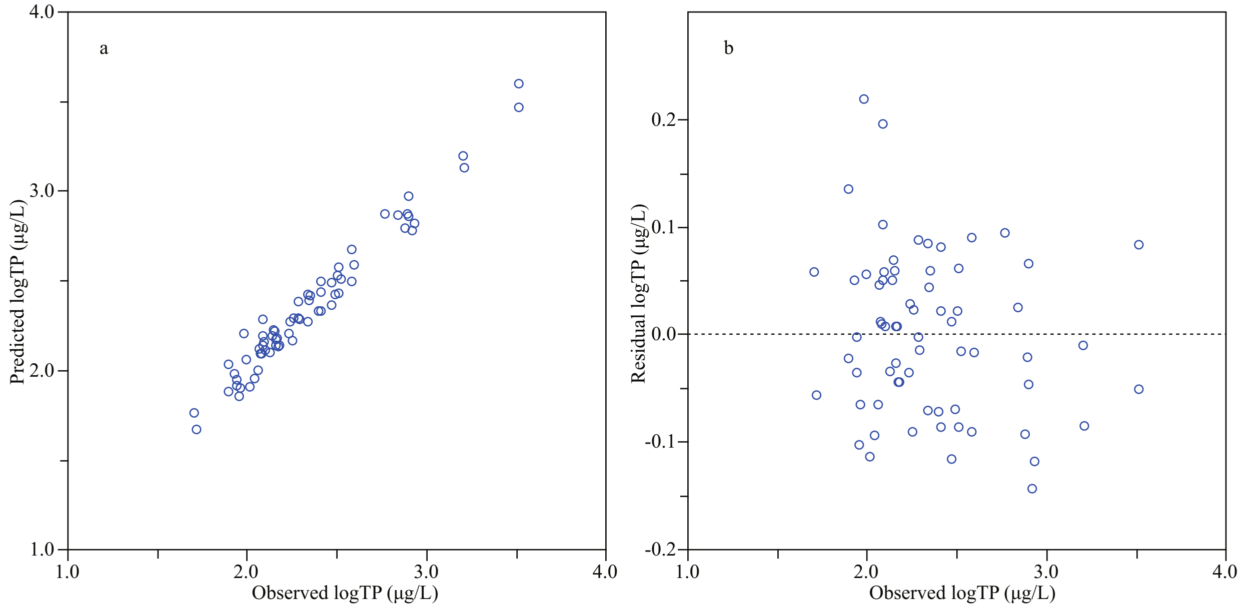

Fig.5 Observed versus predicted annual mean TP concentrations (log μg/L) (a), and observed versus residual annual mean TP concentrations (log μg/L) (b), based on WA-PLS (component 5) model of 68 sites

square error prediction (RMSEP). Compared with other regional TP models in the world, it had become the best modeling method due to its stronger derivation ability. However, it could be seen from its residual distribution map (Fig.4) that the three sampling points S4, S23, and S26 had higher derivation residual values, which were considered to be outliers and deleted. After deletion of samples from S4, S23, and S26, the performance ability of the transfer function was improved: the RMSEP was reduced to 0.169, and the correlation coefficient between the observed value and the predicted value (R^2_Jack) was increased to 0.823; the fitted curve of the observed value and the predicted value was much closer to the 1:1 diagonal (Fig.5). Furthermore, another important index for determining the function's derivation ability—the Max. Bias was increased to 0.094, but it was not very evident. In summary, the WA-PLS (component 5) model after residual component extraction was the best model.

Table 4 Comparison of test results of diatom-total phosphorus multi-transfer function models for 71 sampling points on the Songnen Plain

Model	RMSEP	R^2_Jack	Max. Bias
Inverse WA	0.237	0.640	0.786
Classical WA	0.259	0.646	0.634
Inverse WA_{tol}	0.257	0.581	0.398
Classical WA_{tol}	0.271	0.586	0.248
WA-PLS (Component 1)	0.237	0.6406	0.785
WA-PLS (Component 2)	0.207	0.726	0.440
WA-PLS (Component 3)	0.200	0.748	0.189
WA-PLS (Component 4)	0.199	0.758	0.124
WA-PLS (Component 5)	0.202	0.759	0.083

4 DISCUSSION

Diatom-total phosphorus transfer functions are helpful in estimating the epilimnetic TP concentrations in order to reconstruct the former trophic statuses of lakes in the study area (Witak et al., 2017). Quantitative reconstructions are useful because they can simplify the complex, multivariate relationships between biota and their physical and chemical environments. However, this methodology is not without problems (Juggins, 2013; Juggins et al., 2013), the most important of these result from attempts to reconstruct non-causal environmental variables and from the effects of secondary variables (Juggins, 2013). Therefore, palaeoecologists should give greater attention to what could and could not be reconstructed.

The quantitative study of diatoms has regional characteristics. The environmental gradients affecting the distribution of diatoms are different in different lakes. In arid and semi-arid areas, climate change affects the salinity and water level of lake water, and the distribution of organisms is significantly affected by the salinity or atmospheric effective humidity (Fritz et al., 1991). In the area affected by acid rain, organisms indicate the change of pH value (Cameron et al., 1999). And in the area of economic development, the problem of lake eutrophication is relatively serious, and organisms are more sensitive to the change of nutrients (Liu et al., 2018). Therefore, for a research area, it is necessary to determine what environmental component plays the leading role. Due to the acceleration of population growth and economic development, the industrial and agricultural production pollution and human activities have aggravated water environmental pollution on the Songnen Plain wetland in recent years. In this study, CCA analysis of the relationship between planktonic diatoms and the environment also revealed that TP was the most significant factor affecting the distribution of planktonic diatoms in this area, for economic development areas, it is not surprising that most of the diatom information is explained by nutritional variables. TN is another important index to evaluate the nutritional status of lakes; however, we found that the effect of TN was not significant (P>0.05), which might be due to that TN depends on other variables and cannot be independently used to explain diatom data. In addition, salinization was mentioned as a pressure on these lakes. Salinity could be represented by conductivity or TDS, both of them had been removed because they were collinear with other variables in the preliminary CCA. Nevertheless, the presence of a large number of saline-tolerant species, such as *Aulacoseira granulata*, *Aulacoseira ambigua*, *Cyclotella meneghiniana*, *Pseudostaurosira brevistriata*, *Cocconeis placentula*, *Gyrosigma acuminatum* (Kützing) Rabenhorst, *Navicula cryptocephala* Kützing, *Navicula pygmaea*, *Nitzschia hantzschiana* Rabenhorst, and *Nitzschia linearis* W. Smith, suggested that salinization was one of the characteristics of water quality in the study area.

The modern dataset used in this paper was developed based on modern phytoplankton samples rather than on the planktonic component of surface sediment diatom assemblages. Although the diatoms in the Chagan Lake sedimentary core were found largely similar to those in the modern dataset through the preliminary experiment, the use of the approach to reconstruct TP from fossil assemblages in sediment cores remains questionable as the sample type used to generate a modern dataset (modern phytoplankton samples) was not the same as that for which the transfer function will be applied (i.e. sedimentary assemblages). Usually, modern datasets are based on samples taken from the surface sediments in the centers or at the deepest points of the lakes (Adler and Hübener, 2007). A single sample is considered to represent a good approximation of the composition of a lake's diatom species, including planktonic and benthic taxa. However, the combination of benthic and planktonic species in one dataset might be problematic, because benthic and planktonic species do not depend on the same environmental variables (Hansson, 1992; Blumenshine et al., 1997). Philibert and Prairie (2002) found that the inclusion of both benthic and planktonic species resulted in the transfer function with the most predictive power. In the future, relevant data of benthic diatoms in this area will be added to the background dataset. Also, the distribution of sampling points in the modern dataset was limited and the water samples were collected only once at each sampling point from June to August in 2008. In this case, we need to recognize the limitations of a single spot water chemistry value for capturing variables that experience high intra-annual variability, so it is necessary to expand the background data of other years in future research to improve the predictive ability of the model.

The results show that TP was the most important and significant environmental variable to explain the distribution of diatoms, and it was the best variable for establishing the transfer function. In the CCA constrained to the nine environmental variables (Fig.3), the distribution characteristics of diatom species in the direction of TP gradient change reflected the tolerance of species to TP, the farther the TP index forward end, the higher the tolerance of the distributed species to TP. Because of the good correlation between TP and the first axis, it could be considered in general that the species distributed on the positive half axis of the first axis had relatively high tolerance to TP and preferred a nutritious water environment, and the number of these species was close to 1/3 of the total number of diatom species. However, most of the species were distributed in the negative half axis of the first axis, which indicated that the optimum TP content for the survival of these species was relatively low. Combined with the assemblage distribution of dominant species (Fig.2), it could also be found that the number of diatom species was less and the diversity was relatively low at the sampling points with higher TP content (above S52). The competition among different algae, especially the development of blue and green algae in eutrophication water and their demand for phosphorus in a water body, were the main reasons for the decrease of the number and the diversity of planktonic diatoms (Bennion, 1994; Hall and Smol, 1999).

In addition, the TP content of 71 sampling points in the study area was 51–3 304 μg/L. According to the environmental quality standard for surface water issued by the People's Republic of China, no water quality in the study area below Class III (≤0.05 mg/L) was considered as a mesotrophic to eutrophic type (Xiao et al., 2011). Therefore, when reconstructing the TP contents in the historical period with the model, some diatom species that prefer low-nutrition in the sedimentary core might be absent in the modern diatom dataset, which affected the quantitative reconstruction of TP based on the diatom transfer function and consequently led to the overestimation of TP concentration. The root cause of this overestimation phenomenon was the limited length of the TP gradient of the dataset and the lack of sampling points at low TP concentration (<51 μg/L). To solve this problem, the number of sampling points should be increased, especially the number of sampling points at the low value end of the TP gradient.

No matter in DCA analysis, preliminary CCA analysis or further CCA analysis, the cumulative variance value of the species data explained by the first two axes was very low: 16.9%, 10.9%, and 9.7%, respectively. The above results are similar to those on the diatom in the lakes of the southern England and Chinese Tibetan Plateau found by Bennion (1994) and Yang et al. (2001), which might be that the diatom data contains many zero values (Jongman et al., 1995), namely, a species is unlikely to appear when the water environment is above or below its tolerance range. Ter Braak (1986) pointed out that the appearance of this low variance percentage did not affect its interpretation of species data.

In terms of the establishment of transfer function, the four models established by the WA method showed large prediction errors (RMSEP) and small correlation coefficients (R^2_Jack). Of them, the results of the Inverse WA regression model are basically the same as those of the WA-PLS component 1. The second component was generated by optimizing and recalculating the residual information amount correlation in the first component of WA-PLS. Thus, it could be deduced that each component was generated by correcting and improving the previous component. To a certain extent, WA-PLS was an extension and development of WA (Battarbee et al., 2002; Zhuang et al., 2014). It was apparent from the results of the transfer functions (Table 4) that the transfer functions established by WA-PLS had lower prediction errors (RMSEP) and higher correlation coefficients (R^2_Jack) than the models established by the WA method. Therefore, the diatom-total phosphorus transfer functions established by WA-PLS had significantly more powerful derivation ability than the regression models established by WA method in the present study.

By comparing the TP transfer models established in other parts of the world (Table 5), the transfer function obtained in the present study had a smaller

Table 5 Comparison of test results of regional total phosphorus transfer function

Study area	Number of sampling points	Total phosphorus range (μg/L)	RMSEP	R^2_Jack	Reference
Michigan, USA	41	1–15	0.41 ln (μg/L)	0.73	Fritz et al., 1993
Alpine and pre-alpine	86	2–226	0.346 lg (μg/L)	0.57	Wunsam and Schmidt, 1995
Southeastern Finland	78	3–125	0.193 (μg)	0.73	Miettinen, 2003
Middle and lower reaches of the Changjiang (Yangtze) River, China	43	30–548	0.12 lg (μg/L)	0.82	Dong et al., 2006
Northern Poland	46	20–33	0.22 lg (μg/L)	0.42	Witak et al., 2017
Central-eastern Finland	51	3–120	0.191 lg (μg/L)	0.72	Tammelin et al., 2017
Songnen Plain in China	68	51–3304	0.168 lg (μg/L)	0.82	This article

prediction error (RMSEP) and higher regression correlation coefficient (R^2_Jack) between the observed values and the predicted values, so the accuracy of the model established in this paper was satisfactory. The strong derivation ability of the model that might be benefited from the TP data used to establish the model in this study had a wide gradient (51–3 304 μg/L). In general, the longer the environmental gradient, the closer the response pattern of the species to the unimodal mode, and the more reliable the calculation results (Potapova et al., 2004). In addition, the analysis results of CCA indicated that TP was the unique major environmental parameter that could be used to establish the transfer function among the pre-selected 15 environmental variables. Only TP was an important and significant environmental variable affecting the distribution of diatom species in the present study. Differently, both pH and TP values had become the main environmental gradients due to the acidity problem in many lakes in Europe, leading to the diatom species response to both the change in TP and the pH value (Enache and Prairie, 2002; Dong et al., 2006). For this reason, a certain error occurred during the calculation of that model, which caused a reduction in its derivation ability.

5 CONCLUSION

In summary, there was a certain regional specificity of the lakes on the Songnen Plain, planktonic diatoms from 71 sampling points in the study area displayed some responses to the measured environment gradients, and TP gradient was the most important in controlling the distribution of planktonic diatoms. The strengths and contribution of this study can be summarized as follows: (i) it is the first time that the relationship between planktonic diatoms and environment variables has been explored at the regional level of the Songnen Plain, (ii) it is also the first attempt to establish a diatom-environment (TP) transfer function for the region, and (iii) provision of a test of the potential for developing inference models based on modern phytoplankton samples rather than surface sediment diatom assemblages.

Perspectives to this work would be to quantify the role of the TP perturbations that affected the diatom communities in the lakes changing their compositions. The planktonic diatom-TP transfer function developed in the current research can be applied readily and appropriately to the reconstruction of lake water quality, the determination of reference conditions, and in ecological assessment for the lakes on the Songnen Plain.

6 DATA AVAILABILITY STATEMENT

The authors declare that the data supporting the findings of this study are available within the article.

7 ACKNOWLEDGMENT

We would like to express our gratitude to the State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences for their strong support for the collection of specimens used in this paper. And we also thank Bin XUE, the researcher from Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences and Professor Haifeng XIAO from Harbin Normal University for their provision of hydrochemical environmental variables in the study.

References

Adler S, Hübener T. 2007. Spatial variability of diatom assemblages in surface lake sediments and its implications for transfer functions. *Journal of Paleolimnology*, **37**(4): 573-590.

Andrén E, Telford R J, Jonsson P. 2017. Reconstructing the history of eutrophication and quantifying total nitrogen reference conditions in Bothnian Sea coastal waters. *Estuarine*, *Coastal and Shelf Science*, **198**: 320-328.

Battarbee R W, Grytnes J A, Thompson R, Appleby P G, Catalan J, Korhola A, Birks H J B, Heegaard E, Lami A. 2002. Comparing palaeolimnological and instrumental evidence of climate change for remote mountain lakes over the last 200 years. *Journal of Paleolimnology*, **28**(1): 161-179.

Battarbee R W. 1999. The importance of palaeolimnology to lake restoration. *Hydrobiologia*, **395-396**: 149-159.

Bennion H. 1994. A Diatom-Phosphorus transfer function for shallow, eutrophic ponds in southeast England. *Hydrobiologia*, **275-276**(1): 391-410.

Bigler C, Hall R I. 2002. Diatoms as indicators of climatic and limnological change in Swedish Lapland: a 100-lake calibration set and its validation for paleoecological reconstructions. *Journal of Paleolimnology*, **27**(1): 97-115.

Birks H J B, Line J M, Juggins S, Stevenson A C, Ter Braak C J F. 1990. Diatoms and pH reconstruction. *Philosophical transactions of the Royal Society B*: *Biological Sciences*, **327**(1240): 263-278.

Birks H J B. 1995. Quantitative palaeoenvironmental reconstructions. *In*: Maddy D, Brew J S eds. Statistical Modelling of Quaternary Science Data. Cambridge University Press, Cambridge, UK. p.161-236.

Blumenshine S C, Vadeboncoeur Y, Lodge D M, Cottingham K L, Knight S E. 1997. Benthic-pelagic links: responses of benthos to water-column nutrient enrichment. *Journal of the North American Benthological Society*, **16**(3): 466-479.

Cameron N G, Birks H J B, Jones V J, Berges F, Catalan J, Flower R J, Garcia J, Kawecka B, Koinig K A, Marchetto A, Sánchez-Castillo P, Schmidt R, Šiško M, Solovieva N, Štefková E, Toro M. 1999. Surface-sediment and epilithic diatom pH calibration sets for remote European mountain lakes (AL: PE Project) and their comparison with the Surface Waters Acidification Programme (SWAP) calibration set. *Journal of Paleolimnology*, **22**(3): 291-317.

Chen G J, Dalton C, Leira M, Taylor D. 2008. Diatom-based total phosphorus (TP) and pH transfer functions for the Irish Ecoregion. *Journal of Paleolimnology*, **40**(1): 143-163.

Craig A. 2018. An investigation into the ecological effects a medieval settlement had on the White Loch of Myrton, Scotland using diatom analysis. *Meliora*: *International Journal of Student Sustainability Research*, **1**(2), http://doi.org/10.22493/Meliora.1.2.0009.

Dai W N, Tan J W, Li Y, et al. 2018. Research progress on water environment treatment and protection of typical wetland in Songnen plain. *Protection Forest Science and Technology*, (7): 55-57. (in Chinese with English abstract)

Denicola D M, De Eyto E, Wemaere A, Irvine K. 2004. Using epilithic algal communities to assess trophic status in Irish lakes. *Journal of Phycology*, **40**(3): 481-495.

Dong X H, Yang X D, Wang R, Pan H X. 2006. A diatom-total phosphorus transfer function for lakes in the middle and lower reaches of Yangtze River. *Journal of Lake Science*, **18**(1): 1-12. (in Chinese with English abstract)

Duthie H C, Yang J R, Edwards T W D, et al. 1996. Hamilton Harbour, Ontario: 8300 years of limnological and environmental change inferred from microfossil and isotopic analyses. *Journal of Paleolimnology*, **15**(1): 79-97.

Enache M, Prairie Y T. 2002. WA-PLS diatom-based pH, TP and DOC inference models from 42 lakes in the Abitibi clay belt area (Québec, Canada). *Journal of Paleolimnology*, **27**(2): 151-171.

Finkelstein S A, Bunbury J, Gajewski K, Wolfe A P, Adams J K, Devlin J E. 2014. Evaluating diatom-derived Holocene pH reconstructions for Arctic lakes using an expanded 171-lake training set. *Journal of Quaternary Science*, **29**(3): 249-260.

Fritz S C, Juggins S, Battarbee R W, Engstrom D R. 1991. Reconstruction of past changes in salinity and climate using a diatom-based transfer function. *Nature*, **352**(6337): 706-708.

Fritz S C, Kingston J C, Engstrom D R. 1993. Quantitative trophic reconstruction from sedimentary diatom assemblages: a cautionary tale. *Freshwater Biology*, **30**(1): 1-23.

Hall R I, Smol J P. 1999. Diatoms as indicators of lake eutrophication. *In*: Stoermer E F, Smol J P eds. The Diatoms: Applications for the Environmental and Earth Sciences. Cambridge University Press, Cambridge, UK. p.128-168.

Hansson L A. 1992. Factors regulating periphytic algal biomass. *Limnology and Oceanography*, **37**(2): 322-328.

Hu H J, Wei Y X. 2006. Systematics, Taxonomy and Ecology: the Freshwater Algae of China. Science Press, Beijing, China. 1023p. (in Chinese)

Huang C, Gu X H, Hu W Y. 2000. Culturing of *Scenedesmus obliquus* in saline-alkali waters and its biological purification. *Journal of Lake Sciences*, **12**(1): 63-67. (in Chinese with English abstract)

Huang Y Y, Wang X T, Liu W, Huang S F, Lei Y D. 2013. Evaluation of a benthic diatom-conductivity transfer function model from Dongjiang basin. *Ecological Science*, **32**(5): 564-570. (in Chinese with English abstract)

Jones V J, Juggins S. 1995. The construction of a diatom-based chlorophyll *a* transfer function and its application at three lakes on Signy Island (maritime Antarctic) subject to differing degrees of nutrient enrichment. *Freshwater Biology*, **34**(3): 433-445.

Jongman R H G, ter Braak C J F, van Tongeren O F R. 1995. Data Analysis in Community and Landscape Ecology. Cambridge University Press, Cambridge, UK. p.91-164.

Juggins S, Anderson N J, Hobbs J M R, Heathcote A J. 2013. Reconstructing epilimnetic total phosphorus using diatoms: statistical and ecological constraints. *Journal of Paleolimnology*, **49**(3): 373-390.

Juggins S. 2003. C2 User Guide: Software for Ecological and

Palaeoecological Data Analysis and Visualisation. University of Newcastle, Newcastle, UK. p.7-70.

Juggins S. 2013. Quantitative reconstructions in palaeolimnology: new paradigm or sick science? *Quaternary Science Reviews*, **64**: 20-32.

Kociolek J P, Balasubramanian K, Blanco S, Coste M, Ector L, Liu Y, Kulikovskiy M, Lundholm N, Ludwig T, Potapova M, Rimet F, Sabbe K, Sala S, Sar E, Taylor J, Van De Vijver B, Wetzel C E, Williams D M, Witkowski A, Witkowski J. 2019. DiatomBase. http://www.diatombase.org. Accessed on 2019-10-13.

Krammer K, Lange-Bertalo H. 1997a. Bacillariophyceae. 1. Teil: Naviculaceae. *In*: Ettl H, Gerloff J, Heynig H, Mollenhauer D eds. Süsswasserflora von Mitteleuropa, Band 2/1. Spektrum Akademischer Verlag, Heidelberg, Berlin. p.1-876.

Krammer K, Lange-Bertalo H. 1997b. Bacillariophyceae. 2. Teil: Bacillariaceae, Epithemiaceae, Surirellaceae. *In*: Ettl H, Gerloff J, Heynig H, Mollenhauer D eds. Süsswasserflora von Mitteleuropa, Band 2/2. Spektrum Akademischer Verlag, Heidelberg, Berlin. p.1-611.

Krammer K, Lange-Bertalo H. 2004a. Bacillariophyceae. 3. Teil: Centrales, Fragilariaceae, Eunotiaceae. *In*: Ettl H, Gerloff J, Heynig H, Mollenhauer D eds. Süsswasserflora von Mitteleuropa, Band 2/3. Spektrum Akademischer Verlag, Heidelberg, Berlin. p.1-599.

Krammer K, Lange-Bertalo H. 2004b. Bacillariophyceae. 4. Teil: Achanathaceae, Kritische Ergänzungen zu Achnanthes s. l., Navicula s. str., Gomphonema. Gesamtliteraturverzeichnis Teil 1-4. *In*: Ettl H, Gerloff J, Heynig H, Mollenhauer D eds. Süsswasserflora von Mitteleuropa, Band 2/4. Spektrum Akademischer Verlag, Heidelberg, Berlin. p.1-468.

Krawczyk D W, Witkowski A, Moros M, Lloyd J M, Høyer J L, Miettinen A, Miettinen A. 2017. Quantitative reconstruction of Holocene sea ice and sea surface temperature off West Greenland from the first regional diatom data set. *Paleoceanography and Paleoclimatology*, **32**(1): 18-40.

Liu B, McLean C E, Long D T, Steinman A D, Stevenson R J. 2018. Eutrophication and recovery of a Lake inferred from sedimentary diatoms originating from different habitats. *Science of the Total Environment*, **628-629**: 1 352-1 361.

Liu S S, Chen S Y, Yao M, Zhang Q H. 2012. Lakes sedimentary diatom application in environmental study. *Journal of Linyi University*, **34**(6): 78-83. (in Chinese with English abstract)

Luostarinen T, Miettinen A, Husum K. 2017. Variability of sea surface temperature and sea ice extent in Kongsfjorden, Svalbard during the last 2000 years. *In*: Proceedings of the 19th EGU General Assembly. EGU, Vienna, Austria. 1007p.

Ma L S. 2018. Study on the Distribution of Diatoms and Their Correlation with Environmental Factors in Sediments of Wetlands in the Sanjiang Plain. Chinese Academy of Sciences, China, Beijing. 101p. (in Chinese with English abstract)

Marchetto A, Musazzi S. 2001. Comparison between sedimentary and living diatoms in Lago Maggiore (N. Italy): implications of using transfer function. *Journal of Limnology*, **60**(1): 19-26.

Miettinen J O. 2003. A diatom-total phosphorus transfer function for freshwater lakes in southeastern Finland, including cross-validation with independent test lakes. *Boreal Environment Research*, **8**: 215-228.

Philibert A, Prairie Y T. 2002. Is the introduction of benthic species necessary for open-water chemical reconstruction in diatom-based transfer functions? *Canadian Journal of Fisheries and Aquatic Sciences*, **59**(6): 938-951.

Pokras E M, Molfino B. 1986. Oceanographic control of diatom abundances and species distributions in surface sediments of the tropical and southeast Atlantic. *Marine Micropaleontology*, **10**(1-3): 165-188.

Potapova M G, Charles D F, Ponader K C, Winter D M. 2004. Quantifying species indicator values for trophic diatom indices: a comparison of approaches. *Hydrobiologia*, **517**(1-3): 25-41.

Potapova M G, Minerovic A D, Veselá J, Smith C R. 2019. Diatom New Taxon File at the Academy of Natural Sciences (DNTF-ANS), Philadelphia. http://symbiont.ansp.org/dntf/. Accessed on 2019-10-13.

Qi Y Z, Li J Y. 2004. Flora Algarum Sinicarum Aquae Dulcis. Science Press, Beijing, China. p.1-161. (in Chinese)

Reavie E D, Heathcote A J, Chraïbi V L S. 2014. Laurentian Great Lakes phytoplankton and their water quality characteristics, including a diatom-based model for paleoreconstruction of phosphorus. *PLoS One*, **9**(8): e104705.

State Environmental Protection Administration of China. 2002. Monitoring and Analysis Method of Water and Waste Water. China Environmental Science Press, Beijing, China. 836p. (in Chinese)

Tammelin M, Kauppila T, Viitasalo M. 2017. Factors controlling recent diatom assemblages across a steep local nutrient gradient in central-eastern Finland. *Hydrobiologia*, **799**(1): 309-325.

Ter Braak C J F, Juggins S. 1993. Weighted averaging partial least squares regression (WA-PLS): An improved method for reconstructing environmental variables from species assemblages. *Hydrobiologia*, **269-270**(1): 485-502.

Ter Braak C J F, Prentice I C. 1988. A theory of gradient analysis. *Advances in Ecological Research*, **18**: 221-317.

Ter Braak C J F, Šmilauer P. 1998. Canoco Reference Manual and User's Guide to Canoco for Windows: Software for Canonical Community Ordination (Version 4). Centre for Biometry Wageningen, Wageningen, the Netherlands. 351p.

Ter Braak C J F, Verdonschot P F M. 1995. Canonical correspondence analysis and related multivariate methods in aquatic ecology. *Aquatic Sciences*, **57**(3): 255-289.

Ter Braak C J F. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology*, **67**(5): 1 167-1 179.

Ter Braak C J F. 1987. The analysis of vegetation-environment relationships by canonical correspondence analysis. *Vegetatio*, **69**(1-3): 69-77.

Ter Braak C J F. 1988. Canoco-A Fortran Program for Canonical Community Ordination by Partial Detrended Canonical Correspondence Analysis, Principal Components Analysis and Redundancy Analysis (Version 2. 1). TNO Institute of Applied Computer Science, Wageningen. 95p.

Wang S M, Dou H S. 1998. China Lakes Chronicle. Science Press, Beijing, China. 580p. (in Chinese)

Wei D, Yan H, Shan S X, Yan B X. 2001. Hydrochemical characteristics of salt marsh wetlands in western Songnen Plain. *Journal of Geographical Sciences*, **11**(2): 217-223.

Wetzel R G. 2001. Limnology: Lake and River Ecosystems. 3rd edn. Academic Press, San Diego, USA. 1006p.

Winter J G, Duthie H C. 2000. Epilithic diatoms as indicators of stream total N and total P concentration. *Journal of the North American Benthological Society*, **19**(1): 32-49.

Witak M, Hernández-Almeida I, Grosjean M, Tylmann W. 2017. Diatom-based reconstruction of trophic status changes recorded in varved sediments of Lake Zabinskie (northeastern Poland), AD 1888-2010. *Oceanological and Hydrobiological Studies*, **46**(1): 1-17.

Wu J C, Zhang J F. 2009. Water Environmental Chemistry. China Water Resources and Hydropower Press, Beijing, China. 259p. (in Chinese)

Wunsam S, Schmidt R. 1995. A diatom-phosphorus transfer function for alpine and pre-alpine lakes. *Memorie dell Istituto Italiano di Idrobiologia*, **53**: 85-99.

Xiao H F, Xue B, Yao S C et al. 2011. Water quality of lakes evolution in Songnen Plain. *Wetland Science*, **9**(2): 120-124. (in Chinese with English abstract)

Xu B. 2018. Establishment of Diatom-Salinity Transfer Function and Reconstruction of Paleo-salinity in the Minjiang Estuary and its Adjacent Sea Area. Anhui University of Science & Technology, Huainan, China. 84p. (in Chinese with English abstract)

Yang X D, Wang S M, Xia W L, Li W C. 2001. Application of CCA for study on modern lake diatoms and environment in the Tibetan Plateau. *Science in China Series D*: *Earth Sciences*, **44**(S1): 343-350.

Yang X D. 2004. Diatom Transfer Functions and Quantitative Reconstructions of Environment: Case Studies of Lakes in Qinghai-Xizang (Tibetan) Plateau and the Middle and Lower Reaches of Yangtze River. Chinese Academy of Sciences, China. p.1-87. (in Chinese with English abstract)

Yao S C, Xue B, Lv X G, Xiao H F. 2010. The hydrochemical characteristic of lakes in Songnen Plain. *Wetland Science*, **8**(2): 169-175. (in Chinese with English abstract)

Zhang Z S, Huang X F. 1991. Research Methods of Freshwater Plankton. Science Press, Beijing, China. 414p. (in Chinese)

Zhou H T, Na X D, Zang S Y. 2016. Dynamic change of red-crowned crane habitat suitability in the west Songnen Plain during the past 30 years. *Chinese Journal of Ecology*, **35**(4): 1 009-1 018. (in Chinese with English abstract)

Zhuang C C, Zhan Q, Wang Z H. 2014. An attempt on construction of diatom-based sea-level transfer functions in modern tidal flat at Yangtze River estuary. *Journal of Palaeogeography*, **16**(4): 557-568. (in Chinese with English abstract)