ORIGINAL PAPER

A new diatom training set for the reconstruction of past water pH in the Tatra Mountain lakes

Elwira Sienkiewicz [ID] **· Michał Gąsiorowski · Ladislav Hamerlík · Peter Bitušík · Joanna Stańczak**

Abstract Lakes located in the Polish and Slovak parts of the Tatra Mountains were included in the Tatra diatom database (POL_SLOV training set). The relationship between the diatoms and the water chemistry in the surface sediments of 33 lakes was the basis for the statistical and numerical techniques for quantitative pH reconstruction. The reconstruction of the past water pH was performed using the alpine (AL:PE) and POL_SLOV training sets to compare the reliability of the databases for the Tatra lakes. The results showed that the POL_SLOV training set had better statistical parameters (R^2 higher by 0.16, RMSE and max. bias lower by 0.2 and 0.36, respectively) compared to the AL:PE training set. The better performance of the POL_SLOV training set is particularly visible in the case of Przedni Staw Polski where the curve of the inferred water pH shows an opposite trend for the period from the 1960s to 1990 compared to that based on the AL:PE dataset. The reliability of the inferred pH was confirmed by the comparison with current instrumental measurements.

E. Sienkiewicz (✉) · M. Gąsiorowski · J. Stańczak
Institute of Geological Sciences, Polish Academy of Sciences, Research Centre at Warsaw, St. Twarda 51/55, 00818 Warsaw, Poland
e-mail: esienkie@twarda.pan.pl

L. Hamerlík · P. Bitušík
Department of Biology and Ecology, Faculty of Natural Sciences, Matej Bel University, Tajovského 40, 97401 Banská Bystrica, Slovakia
e-mail: ladislav.hamerlik@umb.sk

P. Bitušík
e-mail: peter.bitusik@umb.sk

L. Hamerlík
Institute of Zoology, Slovak Academy of Sciences, Dubravska cesta 9, 845 06 Bratislava, Slovakia

Introduction

The diatom species composition and water chemistry transfer function have been used to reconstruct past lake-water pH, lake eutrophication and climate changes for many years (Birks et al. 1990; Fritz et al. 1991; Bennion et al. 1996). So far, there are many modern regional datasets (e.g. the SWAP pH training set, the Swedish pH training set) and a few large datasets consisting of several smaller training sets (e.g. the AL:PE pH training set, the Combined pH training set) (Juggins 2001b). They are very suitable tools for the reconstruction of long- and short-term changes in the water environment and are especially useful for studying the periods before the existence of instrumental measurements and monitoring data. These

datasets can also be used to reconstruct various environmental parameters in other lakes not included in the dataset. For example, the AL:PE pH training set was used to reconstruct past water pH, and the Combined TP training set was used to reconstruct the total phosphorus in the lakes located in the Tatra Mountains (Western Carpathians) (Gąsiorowski and Sienkiewicz 2010; Sienkiewicz and Gąsiorowski 2018; Sochuliaková et al. 2018). Previously, the AL:PE dataset was the most adequate to reconstruct water pH in the Tatra lakes, because it consisted of diatom assemblages and water chemistry data from 118 alpine and arctic lakes located in different regions of Europe. The Combined TP dataset was derived from nine datasets (together 347 sites). Both datasets were available in the European Diatom Database (EDDI), and contained lakes with water pH ranging from 4.8 to 8.0 (Cameron et al. 1999) and total phosphorus content between 2 and 1189 µg L^{-1} (Juggins 2001a). The majority of the water environmental parameters and algal community composition of lakes in the Tatra Mts. coincided with the parameters of the lakes and diatom flora included in these datasets. Currently, the website given for the EDDI database is no longer accessible. The biological and chemical data included in this dataset covered a timeframe from 1960 to 1998 and the taxonomic nomenclature has not been updated since then. Thus, we created a modern calibration dataset consisting of surface-sediment diatom assemblages with actual nomenclature and environmental data collected from lakes located in the Polish and Slovak Tatra Mts. (POL_SLOV training set). The surface-sediment samples were collected between 2000 and 2019. This dataset can be also used for reconstruction of water pH for other mountain lakes.

Although mountain lakes are generally located in remote areas, they are also impacted by human activity. The highest contamination level in the Tatra Mountains and in many countries in Europe, Canada and the USA was a result of human-induced acid deposition in the second half of the twentieth century (Sienkiewicz et al. 2006). However, the diatom-inferred reconstruction of water pH (DI-pH) in the Tatra Mts. lakes using the AL:PE training set revealed that during this time period, acidic precipitation had little influence on the water pH and floristic changes in the majority of the analysed lakes (Sienkiewicz and Gąsiorowski 2018).

The first goal of this paper is to examine the relationships between the diatom assemblages and water chemistry of 33 lakes from the Tatra Mts. using quantitative methods, including a modern calibration dataset and diatom-inference models. The second goal is to apply a new dataset to the reconstruction of past water pH changes in selected lakes, i.e. Przedni Staw Polski (PSP), Morskie Oko (MOK) and Toporowy Staw Niżni (TSN) and to compare the results of the DI-pH using the POL_SLOV and AL:PE datasets. To demonstrate the differences between both datasets, we used the weighted averaging (WA) method for the reconstruction of the past lake water pH.

Study sites

A total of 33 lakes located in the Polish (11 lakes) and Slovak (22 lakes) part of the Tatra Mts. (Western Carpathians) were included in the POL_SLOV pH training set (Fig. 1). The relatively small number of lakes included in the dataset is due to the fact, that the majority of water bodies in the Tatra Mts. (especially in the Slovak part) are small and periodic and can be considered rather ponds than lakes (Hamerlík et al. 2014). Some of them freeze to the bottom, which makes it impossible to collect samples in winter. The altitudes of the sampled lakes ranged from 1089 to 2124 m a.s.l.; the lowest and highest values correspond to the locations of Toporowy Staw Niżni (TSN) and Vyšné Terianske pleso (VTP), respectively. TSN is the lowermost lake in the Polish Tatra Mts. while VTP is the water body with the longest period of ice cover (above 200 days per year) in the Slovak Tatras (Šporka et al. 2006). Generally, the ice cover on the Tatra lakes lasts from November to May/July. All studied lakes have glacial origins, and most of them are located in the High Tatra Mountains (Fig. 1). Only three lakes are located outside the High Tatras: Smreczyński Staw (SME) and Prvé Roháčske pleso (ROH) are situated in the Western Tatra Mountains, and TSN is located in the Sucha Woda Valley, which is the border between the High and Western Tatras. The remaining abbreviations of lake names included in the dataset are explained in Figs. 1 and 2. The bedrock of the lakes consists mainly of crystalline rocks (granitoids) with patches of Quaternary glacial sediments. The analysed lakes are permanent water bodies with a relatively wide spectrum of water pH (4.7–7.3). Regarding trophic status, most of the
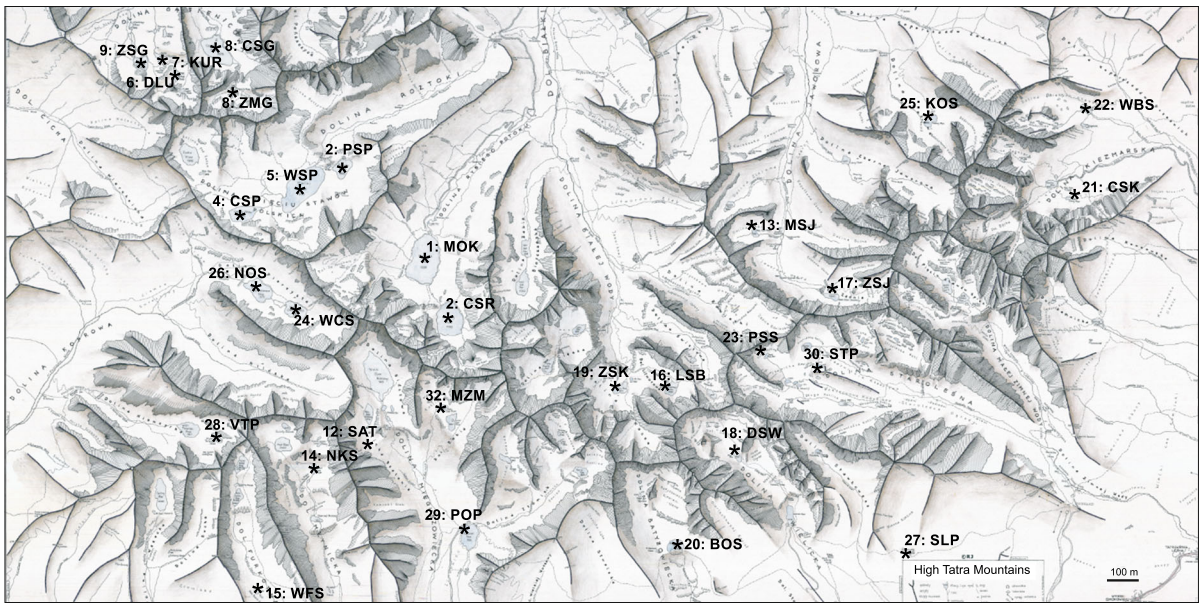
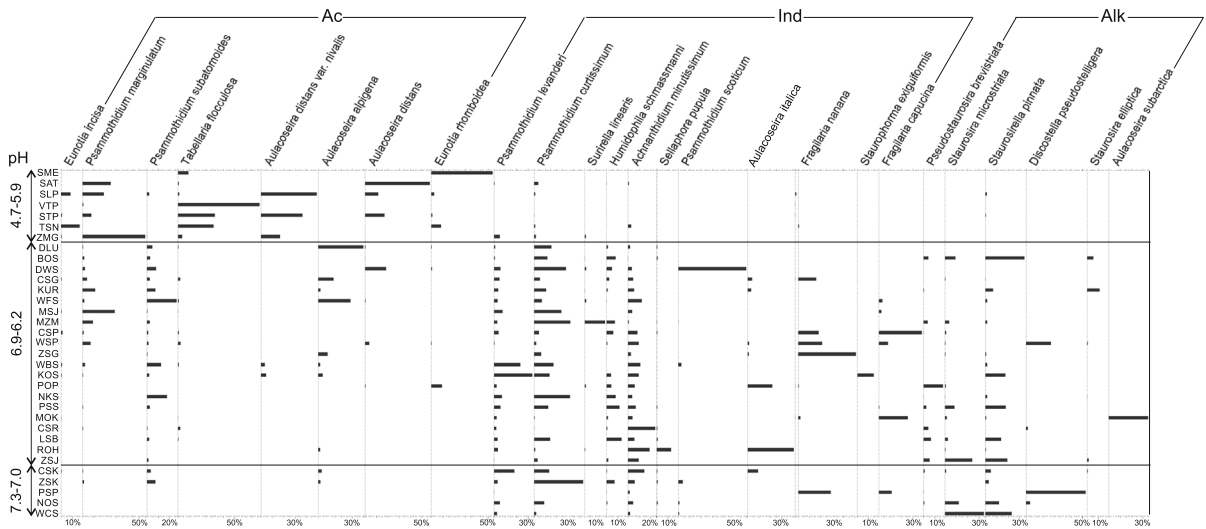**Fig. 1** Locations of the lakes in the High Tatra Mountains that were included in the POL_SLOV dataset

Fig. 2 The percentage abundance of the diatoms in the surface sediments of the 33 lakes in the Tatra Mountains. The diatoms (> 10% in the sample) were ranked according to their pH preferences

studied lakes are oligotrophic, some are oligo-mesotrophic (e.g. PSP) and the minority is dystrophic (e.g. SME). The selected physical and chemical parameters of the surface waters of lakes are listed in Electronic Supplementary Material 1 (ESM 1). The values of the water parameters are based on the median values of all available readings from the lakes for the period 1994–2014 (from 3 to 6 measurements depending on the lake). Details of the measurements are given in Stuchlík et al. (2006) and Kopáček et al. (2015, 2006a, b).

Materials and methods

Sample collection, diatoms and statistical analysis

Surface-sediment samples were collected from the deepest part of the lakes using a Kajak-type gravity corer. The sediments were packed into plastic bags and stored in a refrigerator before analysis. Samples with a volume of 1 cm^3 for diatom analysis were prepared according to the standard methods (Battarbee 1986). An Olympus BX51 light microscope with oil immersion at 1000× magnification was used for the identification of diatom species. At least 300 diatom valves were counted from each sample. The identification of the diatom species was based on Krammer and Lange-Bertalot (1991a, b, 1986, 1988), Lange-Bertalot et al. (2011) and Lange-Bertalot et al. (2017). Division according to pH preferences was based on the Hustedt's classification (Battarbee et al. , 1986).

Detrended canonical analysis (DCA) shown relatively long gradient length (5.5 SD) in diatom data. Hence, canonical correspondence analysis (CCA) was used to identify the relationships between the diatom communities and the environmental variables, such as pH, ANC, dissolved organic carbon (DOC), total phosphorus (TP), total organic nitrogen (TON), Al, Mg, Ca, K, Si and Na. To estimate the species-environment association, two datasets were required: percentage data of the diatom species in each study lake and data on the environmental variables from the same sites. The diatom taxa occurring > 1% in at least one of the studied lakes have been included in the analysis. Next, the statistical significance of the environmental factors as explanatory variables for the diatom community composition was estimated with the Monte Carlo permutation test (999 randomizations) with forward selection of the variables. The CCA analysis and Monte Carlo simulation were performed using CANOCO version 5.0 (ter Braak and Šmilauer 2012).

The reconstruction of the diatom-inferred pH (DI-pH) was conducted using the weighed averaging (WA) method with classical and inverse deshrinking. The taxonomy was harmonized by updating diatom names to current synonyms to reduce inconsistencies between both datasets and improve data quality. Between the AL:PE and POL_SLOV pH training sets, the best results (i.e. the lowest root mean squared error for the training sets; RMSE and the highest coefficient of determination between the inferred and observed values; R^2) were obtained with the weighted average method with inverse deshrinking using the POL_SLOV pH training set (Table 1). For diatom species with maximum relative abundance ≥ 3%, the effective number of occurrences (Hill's N2), weighted average pH optima (WA Opt) and tolerance to water pH (WA Tol) were estimated (ESM 2). Moreover, modern analogue technique (MAT) was applied to evaluate, if fossil assemblages had close analogues in modern data sets. Squared-chord distance was used as dissimilarity measure and bootstrapping was applied as evaluation method. The transfer function was calculated using C2 software version 1.5 (Juggins 2001a).

Results

Dataset characteristics

The lakes included in the POL_SLOV pH training set had a relatively wide range of water pH values (Table 1, Figs. 2, 3), but are not equally distributed along the pH gradient and 60% of sites have modern pH between 6.25 and 7, while pH ranges 4.74–5.5 and 7–7.32 are represented by 15% each. The lowest values of acid neutralizing capacity were calculated for the dystrophic lakes and for the alpine acidic lakes (ANC varied between − 12 and 24), while the highest values were observed in the lakes located on the crystalline rocks with Quaternary glacial sediments containing relatively high concentrations of Ca (the ANC varied between 116 and 147). The lake water TP, TON and Al were high in the forest lakes (approximately 44–59 μg L^{-1}, 21–41 μmol L^{-1}, 4–8 μmol L^{-1}, respectively) and low in the alpine lakes (4–19 μg L^{-1}, 3–17 μmol L^{-1}, 0.1–3 μmol L^{-1}, respectively). Similarly, the amount of DOC was the highest in the dystrophic water bodies, which in our case were forest and meadow lakes (102–661 μmol L^{-1}), and a much lower concentration was observed in the alpine lakes (DOC 10–99 μmol L^{-1}).

Out of the 33 lakes included in the POL_SLOV pH training set, approximately 20% were acidic (pH 4.7–5.9) (Fig. 2). Except for the typical dystrophic lakes situated in the forest zone (below the modern tree line), such as SME, TSN and Slavkovské pleso (SLP) or lakes with a meadow catchment (e.g.

Table 1 The main parameters of the DI-pH reconstructions of both datasets

| Code	RMSE	R2	Ave_Bias	Max_Bias	Jack_R2	Jack_Ave_Bias	Jack_Max_Bias	RMSEP
POL_SLOV								
WA_Inv	0.16055	0.943871	− 1.31881e−015	0.263288	0.862161	− 0.0486116	0.584596	0.274214
WA_Cla	0.165255	0.943871	− 5.65204e−016	0.337876	0.86301	− 0.0499415	0.526339	0.26341
AL:PE								
WA_Inv	0.29788	0.84719	− 1.0507e−007	0.49594	0.78098	0.0010731	0.62429	0.35665
WA_Cla	0.32363	0.84719	2.3438e−007	0.32316	0.78239	0.00068044	0.47118	0.37408
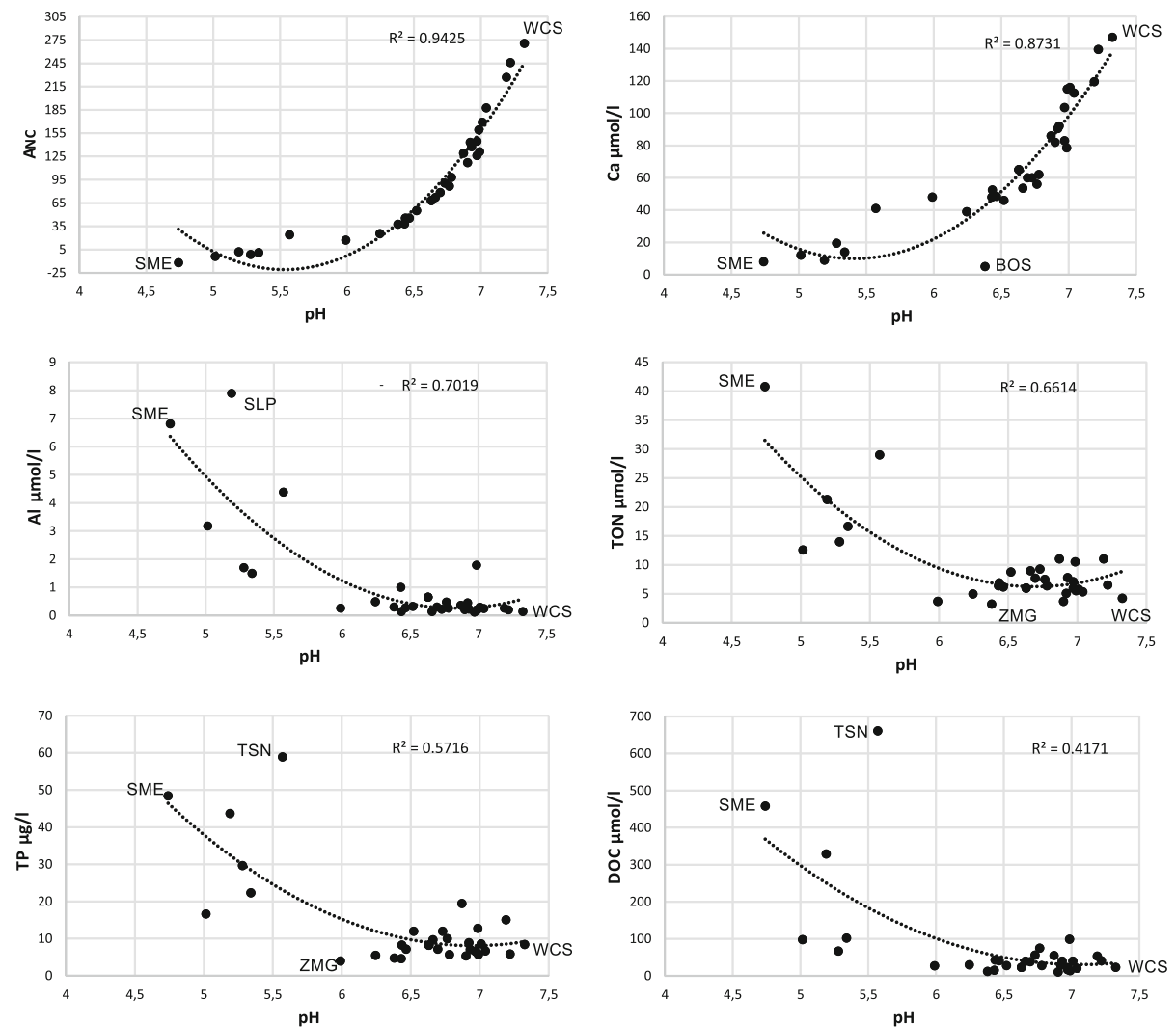
Fig. 3 Correlation of the pH to the selected water chemical variabilities in lakes included in the POL_SLOV dataset. Lakes' abbreviations: *SME* Smreczyński Staw, *TSN* Toporowy Staw Niżni, *WCS* Vyšné Temnosmrečinské pleso, *SLP* Slavkovské pleso, *BOS* Batizovské pleso, *ZMG* Zmarzły Staw Gąsienicowy

Starolesnianske pleso—STP), some clear-water lakes were also included in the group of acidic lakes. These lakes were characterized by some chemical variables similar to dystrophic lakes, i.e. relatively high concentrations of TP and high amounts of DOC, mainly from allochthonous sources. Some examples of such lakes are VTP and Satanie pleso (SAT), with DOC amounting to 67 and 98 μmol L^{-1}, respectively. The concentration of TP in both lakes (approximately 30 and 17 μg L^{-1}, respectively) suggested mesotrophic conditions. All lakes in this group had low acid neutralizing capacity, which was an effect of the low content of carbonates and the other alkali minerals in the bedrock. It results in low Mg and Ca contents in the water (Mg 2.2–20 μmol L^{-1}; Ca 8–48 μmol L^{-1}). The most dominant diatoms in lakes included to this group are acidophilous taxa, such as *Psammothidium marginulatum* (Grun.) Bukhtiyarova & Round, *Tabellaria flocculosa* (Roth) Kützing, *Aulacoseira distans* (Ehr.) Simonsen, *A. distans* var. *nivalis* (W. Smith) E. Y. Haworth and *Eunotia rhomboidea* Hustedt.

More than 63% of the lakes in the dataset had neutral or circumneutral water (pH 6.2–6.9). The higher amounts of Mg and Ca in these lakes compared to lakes belonging to the previous group caused the values of ANC to be higher, varying between 25.5 and 159 μmol L^{-1}. Based on the TP concentrations, except for those in three oligo-mesotrophic water bodies, Vyšné Furkotské pleso (WFS), Popradské pleso (POP) and Prvé Roháčke pleso (ROH), the remaining lakes were oligotrophic (ESM 1). The lowest TP content was usually recorded in lakes with catchments composed of more than 70% rocks, and in lakes with catchment comprising of 30 to 70% meadows and/or rocks. The diatom assemblage composition of the lakes of the group was characterized by high diversity and the dominance of benthic taxa. The most common species were *Psammothidium curtissimum* (J. R. Carter) Aboal, *Achnanthidium minutissimum* (Kütz.) Czarnecki, *Psammothidium levanderi* (Hust.) Bukhtiyarova & Round and *P. subatomoides* (Hust.) Bukhtiyarova & Round. According to their pH preferences, the recorded diatoms were identified as acidophilous (Ac), indifferent (Ind) and alkaliphilous (Alk) taxa (Fig. 2) (van Dam et al. 1994).

Approximately 15% of the studied lakes have water pH above 7.0 but not higher than 7.32. Based on the amounts of TP, only PSP is oligo-mesotrophic (TP 15.1 μg L^{-1}); the remaining lakes from this group are

oligotrophic. The lakes of this group are located in meadow and meadow-rocky catchments (ESM 1), and their ANC is high (169–270.5 μmol L^{-1}). The water bodies in this pH category are characterized by the highest concentrations of Ca (116–147 μmol L^{-1}), indicating that the bedrock is well buffered compared to that in the other studied lakes. The diatom assemblages were dominated by benthic indifferent and alkaliphilous taxa, such as *Psammothidium curtissimum, Pseudostaurosira microstriata* (Marciniak) Flower and *Staurosirella pinnata* (Ehr.) D. M. Williams & Round. However, some planktonic diatoms, e.g. *Fragilaria nanana* Lange-Bertalot and *Discostella pseudostelligera* (Hust.) Houk & Klee, were also found.

The results of the CCA (Fig. 4a, b) showed that six variables explained significantly the variation in the diatom species composition in the study lakes: pH ($p = 0.002$), TON ($p = 0.002$), Mg ($p = 0.002$), TP ($p = 0.04$), ANC ($p = 0.02$) and Al ($p = 0.04$). Other variables' contributions to the model were insignificant ($p > 0.05$) and DOC was deleted from the model due to its high inflation factor (VIF > 20) indicating that it was redundant. Eigenvalues of the first two canonical axes equalled to 0.497 and 0.272, respectively. The total explanatory variables accounted for 37.9%, and pH explained 13.0% of the variance in the dataset. The distribution of the study lakes according to diatom diversity is shown in Fig. 4a. Only 26 species included in the modern training set have effective number of occurrences over 5 (ESM 2), so only these taxa have well defined optima (Hill 1973). The water pH, ANC, Al, TP and TON explain the distribution of the lakes and diatom taxa along axis 1. Their correlation to axis 1 is strong and equals to − 0.88, − 0.60, 0.78, 0.78 and 0.87, respectively. Axis 2 may be a nutrient gradient with weak correlation to magnesium (0.41). Magnesium is an important source responsible for diatom diversity. Arrows corresponding to magnesium and total phosphorus crossing at a right angles and show near-zero correlation. The lakes are split into two groups along axis 1 (Fig. 4a). The first group includes circumneutral lakes with low and moderate TON content (e.g. MOK, PSP and PSS), while the second group contains lakes with low pH and relatively high TON and DOC, such as SME, SAT, VTP and SLP. Generally, lower diatom diversity was observed in lakes with acidic water, but there were some exceptions (e.g. TSN, with pH below 6 and

Fig. 4 Canonical correspondence analysis (CCA) of the POL_SLOV dataset with statistically significant environmental variables. The distribution of all lakes in the dataset according to their diversity. The size of triangles reflects the Hill's N2 index for sites (**a**). The distribution of main diatom taxa in the dataset (**b**). The diameter of circles reflects the Hill's N2 index for main taxa in the dataset. Only 20 taxa with the highest weight value are displayed
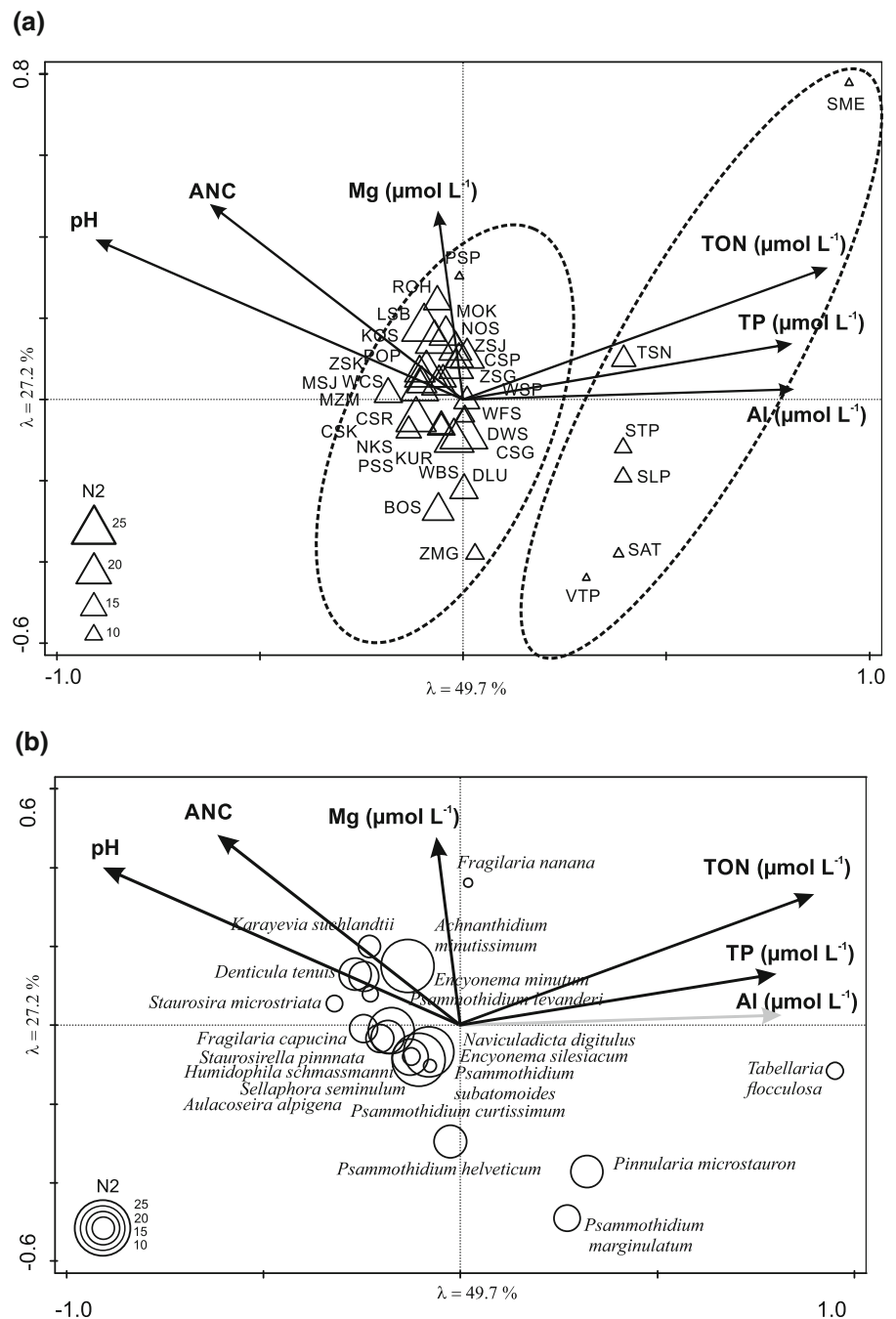
Hill's N2 index 6.63). In contrast, lower diversity (Hill's N2 = 2.77) was found in PSP, despite its slightly alkaline water (ESM 1). The diversity of the diatom flora in these lakes showed that there was only a partial relationship between the pH value and the degree of species diversity.

Vyšne Terianske pleso (VTP), belonging to the group of lakes with pH between 5 and 6, had the lowest diatom diversity (Hill's N2 = 1.89) in the dataset. The highest diversity (Hill's N2 = 17.61) was observed in Nižné Temnosmrečinské pleso (NOS), which belonged to the circumneutral lakes (first group; Fig. 4a). The selected diatom taxa with the largest

weight are shown in Fig. 4b. Most diatom taxa often occurring in neutral/alkaline waters are situated on the top- and bottom-left side of the diagram (along axis 1). Diatoms preferring acidic waters are positioned along axis 2 on the top- and bottom-right side of the diagram.

The diatoms occurring in the POL_SLOV dataset with maximum relative abundances of $\geq 3\%$ in at least one lake, their weighted average pH optima (WA Opt) and tolerances (WA Tol) are shown in ESM 2 and Fig. 5. The lowest pH optima (4.74) were calculated for *Frustulia saxonica* Rabenhorst, *Navicula hoeflerii* Krammer & Lange-Bertalot, *N. subtilissima* Cleve and *Pinnularia divergens* W. Smith. The highest pH optima (7.98) was estimated for *Amphora inariensis* Krammer. The highest and the lowest pH tolerances were determined for *Neidium affine* (Ehr.) Pfitzer and *Cavinula pusio* (Cleve) Lange-Bertalot*,* with tolerance values equal to 1.12 and 0.03 pH units, respectively.

The comparison of pH optima for taxa occurring in $> 20\%$ of the lakes in the AL:PE and POL_SLOV datasets revealed that differences in the values in both training sets ranged between 0 and 0.63 pH units. The highest divergence was observed in the pH optima of *Eunotia rhomboidea*—it equalled 5.52 in the AL:PE dataset (Cameron et al. 1999) and 4.89 in the POL_SLOV dataset.

The best performing diatom-inferred pH transfer function in POL_SLOV dataset was developed using the WA method with classical deshrinking (Table 1). There is a good relationship between estimated and observed pH values ($R^2 = 0.94$). Generally, pH values are underestimated at lower pH (pH < 6), but differences between the predicted and observed values never exceeded 0.5 pH units (Fig. 6). Dissimilarities between the subfossil assemblages and the modern training set were calculated using squared-chord distance as a dissimilarity coefficient. The 10th percentile of dissimilarity range was an approximate threshold value to indicate 'good analogue' (Horton and Edwards 2005). This threshold value for POL_SLOV dataset was 0.795, which means, that subfossil assemblages with higher minimal distance to the closest modern assemblage have no good analogue in the training set.

Reconstruction and comparison of the diatom-inferred pH in sediment cores using the AL:PE pH training set and the POL_SLOV pH training set

So far, the reconstructions of water pH for nine lakes from the Polish part of the Tatra Mountains has been published (Sienkiewicz and Gąsiorowski 2018). The reconstructions were completed using the AL:PE database as a training set. In the current study, we selected three sediment cores collected from MOK, TSN and PSP, to test the new transfer function based on the POL_SLOV dataset. The diatom stratigraphy, geochemical and isotopic data of MOK, TSN and PSP were presented by Gąsiorowski and Sienkiewicz (2010, 2013) and Sienkiewicz and Gąsiorowski (2014).

The DI-pH reconstruction of the three mountain lakes is shown in Fig. 7. The DI-pH$_{AL:PE}$ in MOK ranged between 6.49 and 6.93, while that based on the POL_SLOV training set varied between 6.73 and 6.97. The DI-pH$_{AL:PE}$ and DI-pH$_{POL\_SLOV}$ from the surface sediments (the topmost 1 cm of the sediment column) equalled to 6.89 and 6.95, respectively. The pH model based on the POL_SLOV dataset gave slightly higher values in comparison to those of the model based on the AL:PE dataset and was in better agreement with values measured from the water (pH = 6.93). The occurrence of fossil taxa in the AL:PE dataset varied between 86.7 and 98.8%, while in the POL_SLOV dataset it equalled to 94.4–100%.

In the case of TSN, both curves had the same trend and pH range between 5.48 and 6.25 for the AL:PE dataset and between 5.41 and 6.30 for the POL_SLOV dataset. The DI-pH$_{AL:PE}$ and DI-pH$_{POL\_SLOV}$ values for the recent sediments were 5.50 and 5.46, respectively. The current water pH measured in TSN was 5.57. Both pH curves almost overlapped, and similarly to MOK, the reconstruction based on the POL_SLOV dataset gave slightly higher values, especially before the beginning of the twentieth century. Fossil diatoms occurred in the AL:PE dataset from 62.2 to 90.1%. However, in the POL_SLOV training set occurrence of fossil taxa ranged between 79.6 and 96.3% (ESM 3).

The largest differences between the DI-pH values from the different datasets were observed in the PSP lake. In contrast to the results of the previous lakes, the DI-pH$_{AL:PE}$ showed higher values until the 1960s compared to that based on the POL_SLOV dataset. From the 1960s to 1990, the DI-pH curves indicate opposite trends. The values of DI-pH$_{AL:PE}$ had a wider range (6.65–7.28) than those based on the POL_SLOV dataset (6.69–7.07). The reconstructed water pH for the youngest sediments was 6.69 and 7.07 using the AL:PE and POL_SLOV datasets, respectively, while

Fig. 5 The pH optima and tolerances of different species inferred by weighted averaging
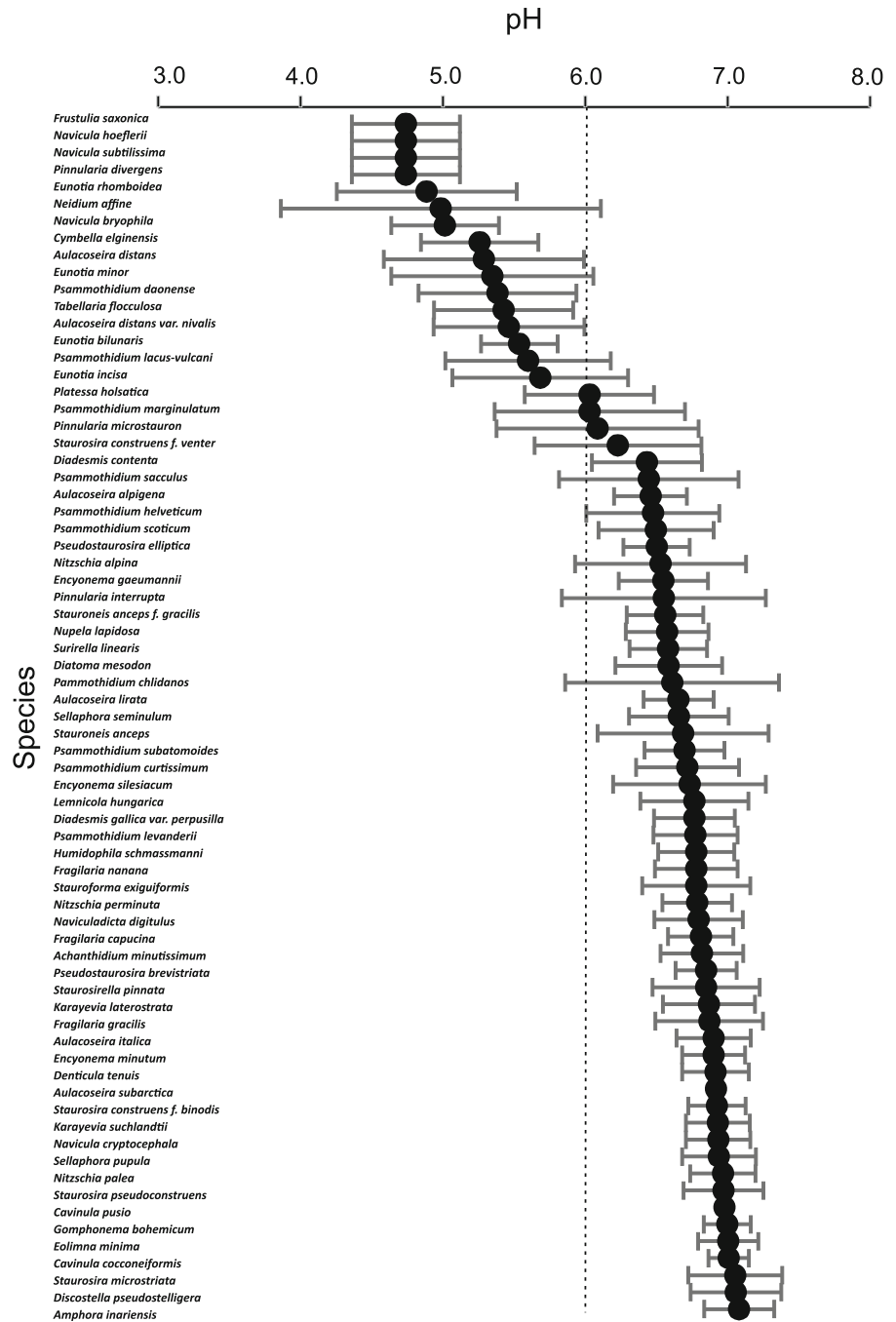
the measured water pH was 7.19. The presence of subfossil diatoms in the AL:PE dataset varied from 72.9 to 97.6%, while in the POL_SLOV dataset was between 93.5 and 100% (ESM 3). These differences may be caused by the fact, that approximately 20% of the diatoms occurring in the POL_SLOV dataset did not appear in the AL:PE dataset. However, among them only five taxa occurred in relative abundance over 5% in the lake and the frequency of other absent diatoms was very low (from single valves to a few percentages). Among the species absent in the AL:PE dataset, *Pseudostaurosira microstriata* had the highest
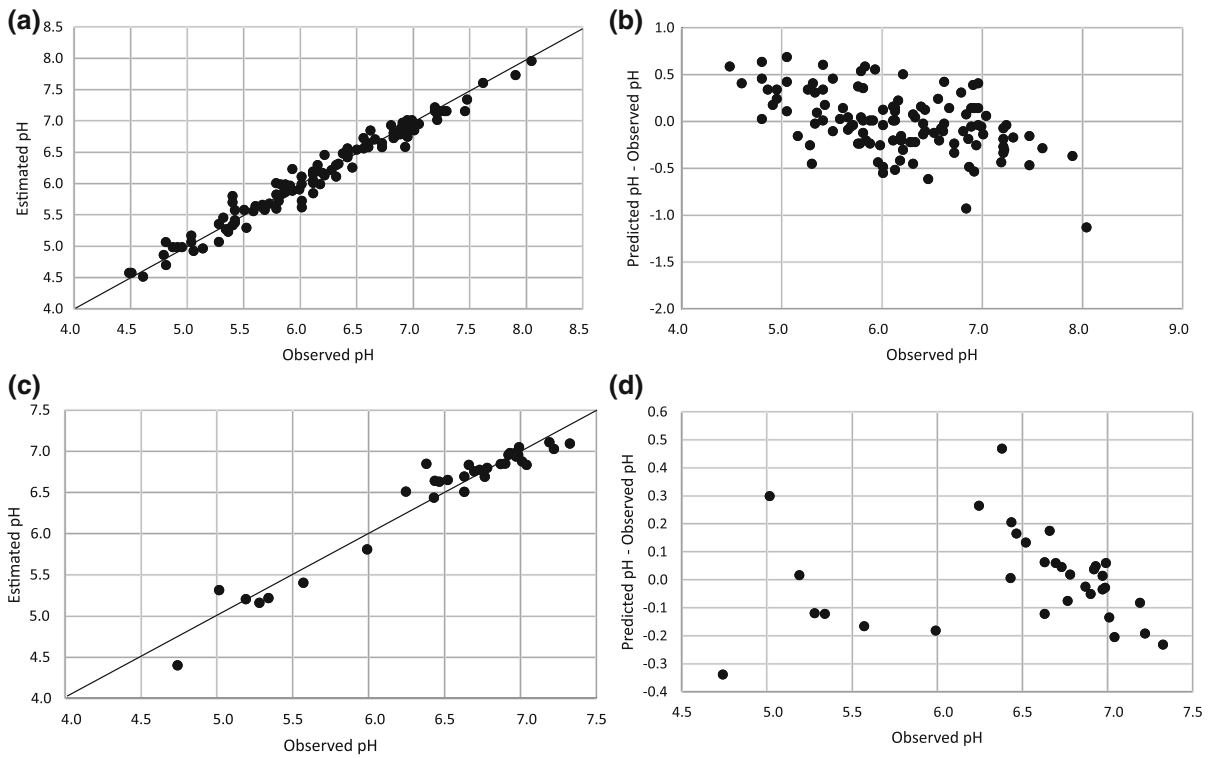
Fig. 6 Plots of observed versus diatom estimated pH for AL:PE (**a**) and POL_SLOV datasets (**c**) and plots of (predicted pH-observed pH) against observed pH for the AL:PE (**b**) and POL_SLOV datasets (**d**)
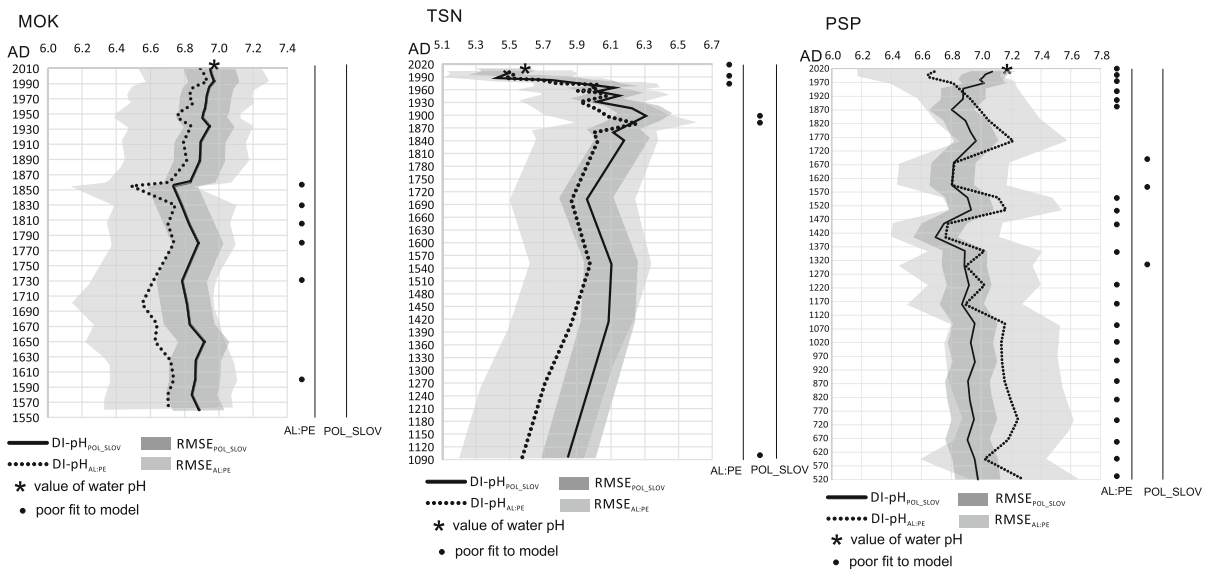

Fig. 7 Comparison of the diatom-inferred pH values in the three lakes in the Tatra Mountains based on the POL_SLOV and AL:PE datasets

frequency (33%) in the POL_SLOV training set followed by *Cymbella elginensis* Krammer (16%). The lack of these two taxa in the AL:PE dataset for the reconstruction of DI-pH in MOK had little influence because of their low frequency (*P. microstriata* < 2.5% and *C. elginensis* < 0.3%). Neither of the aforementioned taxa occurred in TSN, and in PSP the frequency of *C. elginensis* was low (0.6%), while it was high in *P. microstriata* (∼ 12.5%). So, in order not to exclude a relatively abundant taxon, *P. microstriata* was harmonized together with *P. brevistriata* using the AL:PE dataset in the DI-pH reconstruction. In the POL_SLOV set seven diatom taxa occurred with a frequency > 5% (*Discostella pseudostelligera, Staaurosirella pinnata, Aulacoseira subarctica, A. distans, A. distans* var. *nivalis, Achnanthidium minutissimum* and *Psammothidium marginulatum*) which are reported as a complex of different taxa in the AL:PE dataset (i.e. amalgamated).

Discussion

A new dataset was created based on the diatom community and water chemistry of the Tatra Mts. lakes. In total, 220 diatom species from 35 genera were identified from the surface sediments of the 33 lakes. Many of them are common and widespread taxa characteristic of alpine lakes (Koinig et al. 1998; Marchetto et al. 2004). This number of lakes was already large enough to calculate statistically significant results for simple systems dominated by single strong environmental gradient (Juggins and Birks 2012), which in this case is water pH, but datasets containing higher numbers of lakes are generally characterized by higher diversity of biotic and abiotic features. However, datasets including lakes located in only one region, e.g. from the Tatra Mts., usually have better parameters of the reconstruction method, i.e. lower root mean square error for the training set (RMSE) and better squared correlation between the inferred and observed values (R^2). A relatively small area causes the taxa in the modern training set to be to a high extent the same as those in the subfossil dataset, as result of geographical limitation. In a small region, such as the Tatra Mts., the lakes are usually located on a similar bedrock, and the development of the diatom communities is related to the physical environment in which they live. This means that the modern and subfossil diatom assemblages are better harmonized than those in the training sets created from lakes scattered along a huge region or merged from smaller datasets from different areas. On the other hand, a large number of lakes from different regions, but with relatively unified characteristics, ensures a high diversity of species and of the physical and chemical parameters of lakes. A problematic feature of the reconstruction may be that some of the subfossil taxa do not occur in the modern diatom calibration dataset and have a relatively high frequency in the fossil record. These taxa cause the "no modern analogue" problem, and the reconstruction is less reliable (Bigler et al. 2002). However, using datasets created from generally small areas sometimes has the negative aspect often connected with an insufficient number of lakes.

Six lakes belonging to the POL_SLOV dataset showed extreme values of the water chemical parameters (Fig. 3). WCS had the highest pH values and the highest ANC and Ca concentrations, while the lowest pH was noted in the dystrophic SME. The values of ANC > 200 indicated that three lakes, i.e. PSP, NOS and Vyšné Temnosmrečinské pleso (WCS), were located on bedrock insensitive to acidification (U.S. EPA 1988, 2003). The least productive lake in the dataset was VTP, together with the dystrophic lakes, such as SME, TSN and SLP, which had high concentrations of TP, TON and DOC originating from allochthonous sources, but were characterized by low primary productivity. The humic substances present in dystrophic water bodies cause the formation of stable complexes with nutrients that are not available for algae, resulting in low productivity in these lakes. The highest correlation among chemical parameters ($R^2 = 0.94$) occurred between water pH and ANC. Without doubt, one of the weaknesses of POL_SLOV dataset is the uneven distribution of sites along the pH gradient resulting in less reliable reconstruction of pH at ranges < 6.25 and > 7.00.

The canonical correspondence analysis of the dataset showed that pH, Mg and TON were the most important of the tested parameters ($p = 0.002$), explaining the diversity and composition of the diatom communities in the samples (Fig. 3). Parameters such as TP ($p = 0.004$), ANC ($p = 0.02$) and Al ($p = 0.04$) were also responsible for diatom distribution, but to a lesser extent. Generally, two groups of lakes can be distinguished, namely, 1) lakes with low water pH and

low diversity (Hill's N2 < 5), the majority of which were dystrophic lakes; and 2) lakes with water pH varying from slightly acidic to slightly alkaline with higher diversity (N2 > 5), which were mostly oligotrophic water bodies. The first group of lakes was dominated by *P. marginulatum, A. distans, A. distans* var. *nivalis, E. rhomboidea*. In the second group, there were mainly indifferent and alkaliphilous taxa, such as *P. curtissimum, A. minutissimum, S. pinnata* and *D. pseudostelligera*. The type of catchment and the ratio of the catchment area to the lake area also affected the chemical variables of the lake water and, consequently, the diatom diversity. The lakes with forested catchments have high amounts of total phosphorus, dissolved organic carbon and total organic nitrogen. Their concentrations mainly depend on the amount of vegetation and organic matter in the catchment (Kopáček et al. 2000). Higher content of phosphorus usually indicates more fertile conditions, but not in dystrophic lakes. The highest concentrations of TP were noted in dystrophic acid lakes (ESM 1). In this type of lakes, nutrients with humic substances form the stable complexes that are not available for algae and despite the high concentration of TP lakes have low productivity. Generally, together with the decreasing amount of vegetation in the catchment, i.e. from the forest areas through the meadows to the rocky areas, also the content of organic matter and nutrients decreases in water bodies. An indirect factor also limiting diatom diversity is the increasing altitude (especially > 1700 m a.s.l.), which causes longer ice-cover periods and shorter growing seasons. The decrease of vegetation cover, greater rainfall, less evapotranspiration and thicker mineral soils along the altitude often results in lower acid neutralizing capacity. At higher altitudes, spring melt typically starts later, causing postponed weathering of the base cations delivered to water bodies i.e. the potential increase of ANC occurs later in comparison with those at lower altitudes. An example of a lake with low diversity is VTP (N2 = 1.89), which is situated at 2124 m a.s.l. In the surface sediments of that lake only 18 diatom species were identified. More than 70% of diatoms occurring there belong to acidophilous *Tabellaria flocculosa*.

Reconstruction of the past water pH performed from the sediments of MOK, TSN and PSP showed that in the first and second lake, the DI-pH$_{POL\_SLOV}$ gave slightly higher values than the DI-pH$_{AL:PE}$, but both curves had the same trends. MOK is the only one lake with a natural population of brown trout in the Polish part of the Tatra Mts. The fish introduction in 1881 coincides with the increase of *Aulacoseira subarctica* (Sienkiewicz and Gąsiorowski 2014). The significant decrease of DI-pH in TSN was in 1960–1990. It was connected with the highest concentrations of sulphate and nitrogen compounds in the Tatra area (Kopáček et al. 2001). During this time, the DI-pH dropped by 0.5 pH units. As confirmed by isotope analysis, this was probably due to the deposition of products resulting from combustion of fossil fuels (Gąsiorowski and Sienkiewicz 2013). The period of recovery after 1990 was marked by a slight increase of pH (Fig. 7). However, the values of DI-pH are still very low because the process of dystrophication is still in progress. In the case of PSP, the general trend of the pH curves was similar, with the exception of the period from the 1960s to 1990 (Fig. 6). During this period, PSP was dominated by planktonic *D. pseudostelligera* and *F. nanana*. The changes in diatom community as well as the disappearance of large *Daphnia* were synchronised with the fish introduction in the second half of the 1960s (Sienkiewicz and Gąsiorowski 2018). DI-pH$_{AL:PE}$ indicated a clear decreasing trend between the second half of the 1700s and 1990. During the last few decades the DI-pH$_{AL:PE}$ indicates an increase, while, the DI-pH$_{POL\_SLOV}$ shows permanent increasing trend for the second half of the 1800s until now. These differences can be the effect of various frequencies of certain diatoms in the modern training sets and fossil samples. Even though, the sum of taxa percentages in the calibration dataset represents even 100% in some samples of the total subfossil count, subfossil species are often recorded at different relative abundances than observed in the modern dataset what causes poor modern analogues (ESM 3). In the AL:PE dataset, the maximum frequency of *D. pseudostelligera* and *F. nanana* amounted to 3.7 and 3.4%, respectively, while in the surface sediments of PSP, these diatoms occurred at 52 and above 28%. In the POL_SLOV dataset, the maximum frequency of both diatom species was approximately 50%. A better fit between modern and fossil diatoms was observed in the Tatra dataset in comparison with the AL:PE training set. This trend was especially visible in the reconstruction of the diatom-inferred pH in PSP for the youngest sediments, which was confirmed by

instrumental measurements (Fig. 7). Although in two out of three lakes the reconstructions of pH have similar trends, an example of PSP shows that using the AL:PE training set may give false image of changes in water pH. The differences in pH values in the youngest sediment were not large (approximately 0.1 pH units), but trend of changes was completely different during the last 150 years. Similar situations could perhaps occur in other previously studied lakes, thus, the POL_SLOV dataset can be used to estimate DI-pH in lakes mentioned earlier and eventually to revise these reconstructions and also use it in other, not yet analysed Tatra lakes, and perhaps in lakes in the Western Carpathians.

Conclusions

The POL_SLOV dataset consists of the diatom-water chemistry transfer function from the 33 lakes located in the Polish and Slovak Tatra Mountains and contains 220 diatom taxa. The reconstruction of the past water pH in three Tatra lakes was performed using the AL:PE and POL_SLOV training sets to estimate which dataset was more reliable. The results of the water pH reconstruction using the weighted average method (WA) based on both datasets were relatively similar in MOK and TSN. However, the application of the Tatra diatom dataset gave a significant reduction in prediction error and maximum bias. The coefficient of determination (R^2) between the inferred and observed values was higher for the POL_SLOV training set and equalled 0.94. Both the observed and inferred pH values and the sum of the fossil diatoms in the modern dataset were better with the use of the Tatra diatom dataset. The poor quantitative fit among the modern and fossil taxa could not give reliable results for the pH reconstruction, as in the case of the values for PSP for the period from the 1960s to 1990. There the curve of the DI- $pH_{AL:PE}$ shows the opposite trend in relation to that in the curve of the DI-pH_{POL_SLOV}. The current measurements of water pH indicate that the POL_-SLOV dataset was more reliable for the pH reconstruction of lakes located in the Tatra Mountains. The new dataset allows revision of the past pH reconstruction for the Tatra lakes previously published and the estimation of pH changes for other lakes not yet explored.

References

Battarbee RW (1986) Diatom analysis. In: Berglund BE (ed) Handbook of holocene palaeoecology and palaeohydrology. Wiley, Hoboken, pp 527–570

Battarbee RW, Smol JP, Meriläinen J (1986) Diatoms as indicators of pH: an historical review. In: Smol JP, Battarbee RB, Davis RB, Meriläinen J (eds) Diatoms and lake acidity. The Netherlands, pp 5–16

Bennion H, Juggins S, Anderson N (1996) Predicting epilimnetic phosphorus concentrations using an improved diatom-based transfer function, and its application to lake eutrophication management. Environ Sci Techol 30:2004–2007

Bigler C, Larocque I, Peglar SM, Birks HJB, Hall RI (2002) Quantitative multiproxy assessment of long-term patterns of Holocene environmental change from a small lake near Abisko, northern Sweden. The Holocene 12:481–496

Birks HJB, Line JM, Juggins S, Stevenson AC, ter Braak CJF (1990) Diatoms and pH reconstruction. Philos T R Soc Lond Ser B 327:263–278

ter Braak CJF, Šmilauer P (2012) Canoco reference manual and user's guide: software of ordination (version 5.0). Microcomputer power (Ithaca, NY, USA)

Cameron NG, Birks HJB, Jones VJ, Berge F, Catalan J, Flower RJ, Garcia J, Kawecka B, Koinig KA, Marchetto A, Sánchez-Castillo P, Schmidt R, Šiško M, Solovieva N, Šte-

fková E, Toro M (1999) Surface-sediments and epilithic diatom pH calibration sets for remote European mountain lakes (AL:PE Project) and their comparison with the Surface Water Acidification Programme (SWAP) calibration set. J Paleolimnol 22:291–317

Fritz SC, Juggins S, Battarbee RW, Engstrom DR (1991) Reconstruction of past changes in salinity and climate using a diatom-based transfer function. Nature 352:706–708

Gąsiorowski M, Sienkiewicz E (2010) 20th century acidification and warming as recorded in two alpine lakes in the Tatra Mountains (South Poland, Europe). Sci Total Environ 408:1091–1101

Gąsiorowski M, Sienkiewicz E (2013) The sources of carbon and nitrogen in mountain lakes and the role of human activity in their modification and determined by tracking stable isotope composition. Water Air Soil Poll 224:1498

Hamerlík L, Svitok M, Novikmec M, Očadlík M, Bitušík P (2014) Local, among-site, and regional diversity patterns of benthic macroinvertebrates in high altitude waterbodies: ponds differ from lakes? Hydrobiologia 723:41–52

Hill MO (1973) Diversity and evenness: a unifying notation and its consequences. Ecology 54:427–432

Horton BP, Edwards RJ (2005) The application of local and regional transfer functions to reconstruct former sea levels, North Norfolk, England. The Holocene 15:216–228

Juggins S (2001a) C2—software for ecological and palaeoecological data analysis and visualization user guide version 1.5., 73

Juggins S (2001b) The European diatom database, User Guide, version (1.0)

Juggins S, Birks HJB (2012) Quantitative environmental reconstructions from biological data. In: Birks H, Lotter A, Juggins S, Smol J (eds) Tracking environmental change using lake sediments. Development in paleoenvironmental research, vol 5. Springer, Berlin, pp 431–494

Kawecka B, Galas J (2004) Diversity of epilithic diatoms in high mountain lakes under the stress of acidification (Tatra Mts., Poland). Ann Limnol Int J Lim 39(3):239–253

Koinig KA, Schmidt R, Sommaruga-Wögrath S, Tessadri R, Psenner R (1998) Climate change as the primary cause for pH shifts in a high alpine lake. Water Air Soil Poll 104:167–180

Kopáček J, Stuchlík E, Straškrabová V, Pšenáková P (2000) Factors governing nutrient status of mountain lakes in the Tatra Mountains. Freshwater Biol 43:369–383

Kopáček J, Veselý J, Stuchlík E (2001) Sulphur and nitrogen fluxes and budgets in the bohemian Forest and Tatra Mountains during the industrial revolution (1850–2000). Hydrol Earth Syst Sc 5:391–405

Kopáček J, Stuchlík E, Hardekopf D (2006) Chemical composition of the Tatra Mountain lakes: recovery from acidification. Biologia 61(Supplement 18):S21–S33

Kopáček J, Borovec J, Hejzlar J, Kotorová I, Stuchlík E, Veselý J (2006) Chemical composition of modern and pre-acidification sediments in the Tatra Mountain lakes. Biologia 61(Supplement 18):S65–S76

Kopáček J, Bičárová S, Hejzlar J, Hynštová M, Kaňa J, Mitošinkova M, Porcal P, Stuchlík E, Turek J (2015) Catchment biochemistry modifies long-term effects of acidic deposition on chemistry of mountain lakes. Biogeochemistry 125:315–335

Krammer K, Lange-Bertalot H (1986) Süßwasserflora von Mitteleuropa. Bacillariophyceae. I. Teil: Naviculaceae. Gustav Fisher Verlag, Stuttgart

Krammer K, Lange-Bertalot H (1988) Süßwasserflora von Mitteleuropa. Bacillariophyceae. 2. Teil: Bacillariaceae, Epithemiaceae, Surirellaceae. Gustav Fisher Verlag, Stuttgart

Krammer K, Lange-Bertalot H (1991) Süßwasserflora von Mitteleuropa. Bacillariophyceae. 3. Teil: Centrales, Fragilariaceae. Gustav Fisher Verlag, Stuttgart, Eunotiaceae

Krammer K, Lange-Bertalo H (1991) Süßwasserflora von Mitteleuropa. Bacillariophyceae. 4. Teil: Achnanthaceae, Kritische Ergäzungenzu Navicula (Lineolatae) und Gomphonema Gesamtliteraturverzeichnis Teil 1–4. Gustav Fisher Verlag, Stuttgart

Lange-Bertalot H, Bąk M, Witkowski A, Tagliaventi N (2011) Eunotia and some related genera. In: Lange-Bertalot H (ed) Diatoms of Europe 6. A.R.G. Gantner Verlag K.G, Ruggell

Lange-Bertalot H, Hofmann G, Werum M, Cantonati M (2017) Freshwater benthic diatoms of Central Europe: over 800 common species used in ecological assessment. Koeltz Botanical Books, Oberreifenberg

Marchetto A, Mosello R, Rogora M, Manca M, Boggero A, Morabito G, Musazzi S, Tartari GA, Nocentini AM, Pugnetto A, Bettinetti R, Panzani P, Armiraglio M, Cammarano P, Lami A (2004) The chemical and biological response of two remote mountain lakes in the Southern Central Alps (Italy) to twenty years of changing physical and chemical climate. J Limnol 63(1):77–89

Marciniak B (1986) Late quaternary diatoms in the sediments of Przedni Staw Lake (Polish Tatra Mountains). Hydrobiologia 143:255–265

Sienkiewicz E, Gąsiorowski M, Hercman H (2006) Is acid rain impacting Sudetic lakes? Sci Tot Environ 369:139–149

Sienkiewicz E, Gąsiorowski M (2014) Changes in the trophic status of three mountain lakes - natural or anthropogenic process? Pol J Environ Stud 23(3):875–892

Sienkiewicz E, Gąsiorowski M (2018) Limited acid deposition inferred from diatoms during the 20th century—a case study from lakes in the Tatra Mountains. J Environ Sci 64:92–106

Sochuliaková L, Sienkiewicz E, Hamerlík L, Svitok M, Fidlerová D, Bitušík P (2018) Reconstructing the trophic history of an alpine lake (High Tatra Mts.) using subfossil diatoms: disentangling the effects of climate and human influence. Water Air Soil Poll 229–298

Stuchlík E, Kopáček J, Fott J, Hořická Z (2006) Chemical composition of the Tatra lakes: response to acidification. Biologia 61(Supplement 18):S11–S20

Šporka F, Livingstone DM, Stuchlík E, Turek J, Galas J (2006) Water temperatures and ice cover in lakes in the Tatra Mountains. Biologia 61(Supplement 18):S77–S90

U.S. EPA (1988) Chemical characteristics of streams in the mid-Atlantic and southeastern United States. Volume I: population descriptions and physico-chemical relationships. EPA/600/3-88/021A. Washington, DC

U.S. EPA (2003) Response of surface water chemistry to the Clean Air Act Amendments of 1990. EPA/620/R-03/001. Research Triangle Park, NC

van Dam H, Mertens A, Sinkeldam J (1994) A coded checklist and ecological indicator values of freshwater diatoms from the Netherlands. Neth J Aquatic Ecol 28:117–133