data

Data Descriptor

Urals lakes diatom dataset for inferring of total phosphorus and electric conductivity of surface water

Anna Maslennikova [1*]

[1] Federal State Budgetary Institution of Science South Urals Research Center of Mineralogy and Geoecology, Urals Branch, Russian Academy of Sciences; adenophora@inbox.ru

* Correspondence: adenophora@inbox.ru

Abstract: This paper presents the 90-lakes dataset from the Urals. The dataset contains information on total phosphorus (TP), electric conductivity (EC) of water and diatom assemblages (393 taxa abundance in %) of modern sediments (115 samples). Lakes were sampled for sediments and water in the period between 2014 and 2022. The performance of transfer functions was evaluated by a determination coefficient of predicted and observed values (r^2_{boot}), mean square error of prediction ($RMSEP_{boot}$) and maximum bias (Max $Bias_{boot}$) as estimated by bootstrapping. The best electric conductivity inference model (r^2_{boot} = 0.8, $RMSEP_{boot}$ = 0.21 log10 µS cm^{-1}, Max $Bias_{boot}$ = 0.22 log10 µS cm^{-1}) was developed by weighted averaging (WA) with classical deshrinking. The best performing model for TP reconstructions was weighted averaging partial least squares (WA-PLS) Component 2 (r^2_{boot} = 0.64, $RMSEP_{boot}$ = 0.19 log10 µg l^{-1}, Max $Bias_{boot}$ = 0.4 µg l^{-1}). The obtained transfer functions could be applied in paleolimnological studies for reconstructions of TP varied between 5.6 µg l^{-1} and 188 µg l^{-1} and EC varied between 61 and 4500 µS cm^{-1}. The dataset can also be applied to a TP and EC optimum calculation in the diatom species ecology research.

Dataset: https://doi.org/10.1594/PANGAEA.963466

Dataset License: CC-BY

Keywords: diatom assemblage; transfer function; calibration dataset; total phosphorus; electric conductivity; lake water; surface sediments; quantitative reconstruction; Urals

1. Summary (required)

Initially, the research objective was to obtain a calibration dataset for reconstructing electrical conductivity of lake waters and to study variations in effective moisture in the Lateglacial and the Holocene of the Urals. The first variant of the dataset included 72 samples. The data on a total phosphorus content in summer water sampled one time were available only for 39 samples [1]. The resulting transfer function was used to reconstruct electric conductivity variations in the Holocene for four lakes in the Southern and Middle Urals [2–4]. An additional material was selected to study the eutrophication of Urals lakes. The eutrophication problem, as well as for other lakes in the world, is very relevant for the Urals lakes. Quantitative paleolimnological reconstructions are needed to study phosphorus trends in lake water. Application of the transfer function based on the combined TP diatom dataset from the European Diatom Database Initiative (EDDI) [5–10] to the phosphorus reconstruction of the Southern Urals lakes gives overestimated values or unsatisfying result due to the absence of modern analogue samples. The first results obtained with the help of the regional TP transfer function based on 73 lakes calibration dataset (without publication of the TP-calibration dataset) demonstrated only a slight overestimation of diatom inferred TP values [11]. However, $RMSEP_{boot}$ was relatively high (0.3), and studied lakes fossil diatom assemblages, especially for oligotrophic and oligo-mesotrophic lakes, often had poor modern analogues. Therefore, a new goal of our research was to obtain a calibration dataset to reconstruct TP not only in mesotrophic, eutrophic, hypereutrophic, but also in oligotrophic and oligo-mesotrophic lakes. The results obtained were expected to be used to create and publish a calibration dataset of 90 lakes to reconstruct total phosphorus in

Citation: To be added by editorial staff during production.

Received: date
Revised: date
Accepted: date
Published: date

the Urals lakes. In addition, the EC transfer function was expected to be improved during the work.

Currently, the dataset includes 90 lakes located on the eastern macroslope of the Southern and Middle Urals (Russia) (Fig. 1). Median values of total phosphorus concentrations of surface water varied between 5.6 and 313 μg l^{-1}. According to Kratzer and Brezonik [12], 48 lakes are eutrophic, 31 are mesotrophic, 9 are hypereutrophic, and 2 are oligotrophic. The best performing model for TP reconstructions was WA-PLS Component 2 (r^2_{boot} = 0.6, $RMSEP_{boot}$ = 0.21 log10 μg l^{-1}, Max $Bias_{boot}$ = 0.74 μg l^{-1}). For the calculation purposes, diatom abundances (%) were logarithmically transformed, and 236 taxa presented in the dataset more than three times were included. Observed and estimated TP values were revealed to be characterized by a substantial underestimation in two hypereutrophic lakes (Shchelkunskoe and Achakul). The lakes were close and shallow with depth variations between 1 and 2.5 m for Achakul, 1 and 2 m for Shchelkunskoe. The lakes were distinguished by high interannual and interseasonal TP variations. After deleting these lakes from the dataset, the transfer function statistics became much better: r^2_{boot}= 0.64, $RMSEP_{boot}$ =0.19 μg l^{-1}; Max $Bias_{boot}$ = 0.4 μg l^{-1}. For the remaining lakes, median TP ranged between 5.6 μg l^{-1} and 188 μg l^{-1}. Inclusion of all 393 taxa into the model increased a maximum bootstrapped bias from 0.74 μg l^{-1} to 0.79 μg l^{-1} in the complete 90 lake dataset and from 0.4 μg l^{-1} to 0.45 μg l^{-1} in the dataset without Achakul and Shchelkunskoe lakes.

According to the water salinity classification based on electrical conductivity ranges [13], 51 lakes in the dataset were freshwater (<500 μSm cm^{-1}), 31 lakes were slightly brackish (500–2000 μSm cm^{-1}), six lakes were moderately brackish (2000–5000 μSm cm^{-1}), and two were subsaline (15000-45000 μSm cm^{-1}). The WA with classical deshrinking, using 236 taxa, was the best performing model for the EC transfer function: it combined a relatively high bootstrap r^2 value (0.83), low maximum bias (0.33), and mean square error of prediction (0.22 μSm cm^{-1}). The comparison between measured and estimated EC values made it possible to reveal a substantial discrepancy for two subsaline lakes (Kurochkino and Polovinnoe). After excluding these lakes from the dataset, the transfer function statistics was as follows: r^2_{boot} = 0.8, Max $Bias_{boot}$ = 0.22 μSm cm^{-1}, $RMSEP_{boot}$ =0.21 μSm cm^{-1}. Median EC range for the remaining lakes was between 61 and 4500 μSm cm^{-1}. The maximum bias increased slightly when using all 393 taxa: from 0.33 μSm cm^{-1} to 0.38 μSm cm^{-1} in the complete 90 lake dataset and from 0.22 μSm cm^{-1} to 0.24 μSm cm^{-1} in the dataset without subsaline lakes.

Application of the transfer functions to TP and EC reconstructions based on the oligotrophic Lake Yaktykul (Bannoe) sediment core showed the presence of good modern analogues for all samples and a good agreement between measured and diatom inferred parameters.

The resulting database can be used in quantitative reconstructions of electric conductivity and total phosphorus content not only for the Urals lakes, but also for lakes with similar ranges of hydrochemical parameters and for lakes in other regions after merging with the regional datasets. Quantitative TP and EC reconstructions will improve understanding of lake ecosystem shifts drivers. Merging the Urals lakes dataset with other regional and global datasets makes it possible to obtain more robust TP and EC optimums for diatom species which contribute to the diatom ecology research.

2. Data Description (required)

Data contains information on 90 Urals lakes (Figure 1).

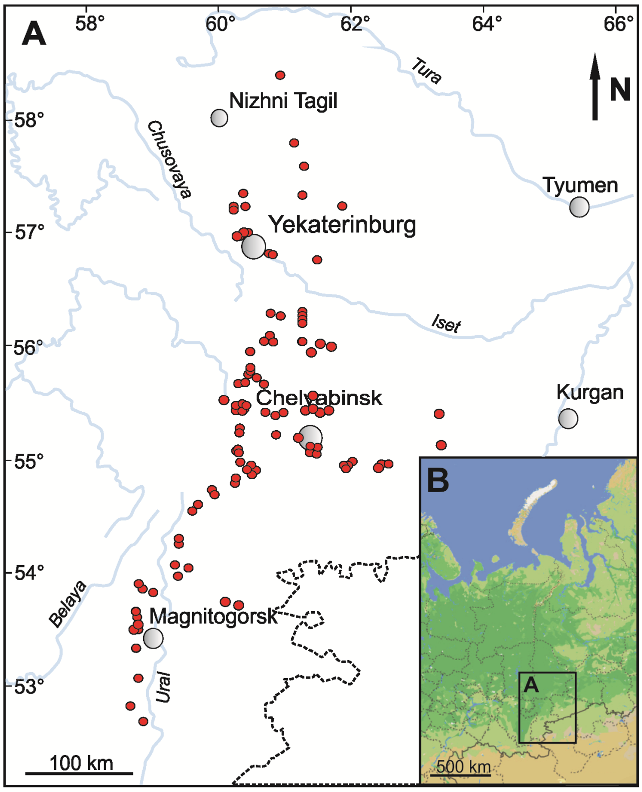

Figure 1. Location of the study area in Russia (B), Urals lakes calibration dataset sampling sites (red circles) (A).

The main data of the dataset are parameters of water (total phosphorus in µg l^{-1} and electric conductivity in µS cm^{-1}) (Table 1) and diatom abundances (relative abundance, %) (Table 2) for 393 taxa in 90 lakes (115 samples). These data are represented as tab-delimited text in repository PANGAEA. Additional metadata with diatom names, authorities, codes is represented in txt format (Table 2).

Table 1. Table of parameters (Ural_lakes_TP_EC.tab, https://doi.org/10.1594/PANGAEA.963463)

Column name	Description
Event	Code of each lake consisting of its name and field number (as required by repository PANGAEA)
Latitude	Lake coordinates
Longitude	Lake coordinates
ID	Field number of lake consisting of a code (Pov, Povs, Povj, BK, Turk, etc.) of the expedition carried out by the Laboratory of Mineralogy of Technogenesis and Geoecology (FSBIS South Urals Research Center of Mineralogy and Geoecology, UB RAS) and a number of a lake or a sediment core. For example: Pov100 means the lake №100 studied in the surface sediment sampling expeditions from 2014 to 2019. BK1-1 means a surface sample of sediment core №1 obtained in the expeditions from 2020 to 2022 carried out to study Lake Bannoe (Yaktykul).
Lake	Full name of a lake
Depth sampling	Water depth (in meters) of a point where surface sediment was sampled.
TP log	Decimal logarithm of a total phosphorus (TP) median

	value. Median value was calculated based on TP measured in spring, summer, and autumn water of a lake. In most cases, TP measured within several years before or after sediment sampling was also included in the calculation.
EC log	Decimal logarithm of an electric conductivity (EC) median value. Median value was calculated based on EC measured in spring, summer, and autumn water of a lake. In most cases, EC measured within several years before or after sediment sampling was also included in the calculation.
TP [μg/l] (Median values)	Median value of total phosphorus concentrations in μg l^{-1}
TP [μg/l] (Median values)	Average value of total phosphorus concentrations in μg l^{-1}
TP std dev [±]	Standard deviation of total phosphorus in μg l^{-1}
Cond electr [μS/cm] (Median values)	Median value of electric conductivity in μSm cm^{-1}
Cond electr [μS/cm] (Average values)	Average value of electric conductivity in μSm cm^{-1}
EC std dev [±]	Standard deviation of electric conductivity in μSm cm^{-1}

Table 2. Table of diatom abundances (Ural_lakes_diatoms.tab, https://doi.org/10.1594/PANGAEA.963463)

Column name	Description
Event	Code of each lake consisting of its name and field number (as required by repository PANGAEA)
Latitude	Lake coordinates
Longitude	Lake coordinates
ID	Field number of lake consisting of a code (Pov, Povs, Povj, BK, Turk, etc.) of the expedition carried out by the Laboratory of Mineralogy of Technogenesis and Geoecology (FSBIS South Urals Research Center of Mineralogy and Geoecology, UB RAS) and a number of a lake or a sediment core. For example: Pov100 means the lake №100 studied in the surface sediment sampling expeditions from 2014 to 2019. BK1-1 means a surface sample of sediment core №1 obtained in the expeditions from 2020 to 2022 carried out to study Lake Bannoe (Yaktykul).
Lake	Full name of a lake
Name[1]	Name of a diatom taxon verified using the Algaebase [14]
Authority[1]	Authority of a diatom taxon verified using the Algaebase [14]
Code[1]	A diatom taxon code in the Urals lake dataset
A. brevipes -.....Diatoms indet [%]	Columns which contain information on abundances (in % of total sum) of 393 diatom species.

[1] The column from the table of additional metadata (file.txt) of diatoms dataset which is located between 'Keyword(s)' and 'Funding' and that is shown in the middle of the website: https://doi.pangaea.de/10.1594/PANGAEA.963464

3. Methods (required)

A total of 90 lakes in the Southern and Middle Urals were sampled in the 2014 and 2016 expeditions and in the period from 2018 to 2022. Lakes were sampled in the point with average or maximum depth (depending on a lake surface area). Water samples were collected by hand below the surface (0.3–0.5 m deep). Surface sediment samples (from uppermost 1 cm up to 3 cm in lakes with watered sediments) were collected using a small (0.5 m) gravity corer. Lake depth was determined with a hand-held acoustic depth meter. Electrical conductivity was measured using a hand-held Hanna HI9333000. Bottles were filled and kept in a cool, dark place and transported to the laboratory within 1–2 days, or in the case of several distant lakes, within 4 days.

Total phosphorus was measured in unfiltered lake water by acid persulfate digestion and spectrophotometry. For the diatom analysis purpose, lake sediment samples were treated with nitric and perchloric acids to remove organic matter. Slides were mounted using high refractive mountants: Elyashev's mountant [15] and Naphrax mountant. Diatom species on the slides were observed, identified, and counted with a Mikmed 6 var. 7 microscope using bright-field oil immersion optics at 1000× magnification. Taxonomic identification measurements were carried out with a ToupView 3.7. software, and photomicrographs were obtained with a ToupCAM UCMOS14000KPA digital camera. When necessary, taxa were identified with the help of electron microscopy (TE SCAN VEGA 3). All diatom nomenclature was verified using the Algaebase [14].

The diatom species relative abundance data were log-transformed for numerical analysis. The diatom-based inference models were developed by simple weighted averaging and weighted averaging partial least squares methods [16,17]. The performance of all transfer functions was evaluated with the use of a mean square error of prediction, as estimated by bootstrapping and associated statistics such as maximum bias, and a determination coefficient between predicted and observed values [18]. Transfer functions of electrical conductivity and total phosphorus were developed using a C2 software [19]. Water analysis and diatom identification with the following numerical analysis were carried out in the South Urals Research Center of Mineralogy and Geoecology, Urals Branch, Russian Academy of Sciences (Miass, Russia).

4. User Notes (optional)

For quantitative reconstructions and for estimation of EC and TP optima of diatom species, the tables with environmental variables (Ural_lakes_TP_EC.tab) and diatom species counts (Ural_lakes_diatoms.tab) should be transferred to Exel and then should be imported into the C2 software or similar program according to the C2 manual [19]. To reconstruct TP and EC, the user should also import a table with fossil diatom assemblages.

5. Patents

Funding: This research was funded 1) by Russian Science Foundation, grant number 21-17-00071, https://rscf.ru/project/21-17-00071/(the data on TP transfer function and partly on EC transfer function (samples of 2021 and 2022 expeditions)). 2) by the State Contract of South Urals Research Center of Mineralogy and Geoecology, UB RAS (the approved EC transfer function). The APC was funded by Russian Science Foundation (50%) and South Urals Research Center of Mineralogy and Geoecology, UB RAS (50%).

Institutional Review Board Statement: No ethical guidelines or contents were relevant for the research reported in this article.

Informed Consent Statement: Not applicable

Data Availability Statement: All data are available in the website: https://doi.pangaea.de/10.1594/PANGAEA.963466

Acknowledgments: The author thanks Vasily Gulakov for help in field work (South Urals Research Center of Mineralogy and Geoecology, UB RAS, Russia), Lyubov Lapshina and Lyudmila Udachina (South Urals Federal Research Center of Mineralogy and Geoecology of the Urals Branch of the Russian Academy of Sciences) for generous help in lakes water analysis.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Maslennikova, A.V. Southern and Middle Urals Lakes Diatom Abundance and Hydrochemistry 2019.
2. Maslennikova, A.V. Development and Application of an Electrical Conductivity Transfer Function, Using Diatoms from Lakes in the Urals, Russia. *J Paleolimnol* **2020**, *63*, 129–146, doi:10.1007/s10933-019-00106-z.
3. Maslennikova, A.V.; Udachin, V.N.; Anfilogov, V.N. Holocene Environments of Anomalous Uranium Concentrations in Sediments of Syrytkul Lake (Southern Urals). *Dokl. Earth Sc.* **2020**, *492*, 323–326, doi:10.1134/S1028334X2005013X.
4. Maslennikova, A.V. Holocene Environments in the Middle Urals: Palaeolimnological Proxies from the Lake Tavatui (Russia). *Quaternary International* **2022**, *622*, 51–64, doi:10.1016/j.quaint.2022.02.033.
5. Lotter, A.F. Subfossil and Modern Diatom Plankton and the Paleolimnology of Rotsee (Switzerland) since 1850. *Aquatic Science* **1989**, *51*, 338–350, doi:10.1007/BF00877176.
6. Bennion, H. A Diatom-Phosphorus Transfer Function for Shallow, Eutrophic Ponds in Southeast England. *Hydrobiologia* **1994**, *275*, 391–410, doi:10.1007/BF00026729.
7. Wunsam, S.; Schmidt, R. A Diatom-Phosphorus Transfer Function for Alpine and Prealpine Lakes. *Mem. Ist. Ital. Idrobiol.* **1995**, *53*, 85–99.
8. Bennion, H.; Juggins, S.; John, A. Predicting Epilimnetic Phosphorus Concentrations Using an Improved Diatom-Based Transfer Function and Its Application to Lake Eutrophication Management. *Environmental Science and Technology* **1996**, doi:10.1021/es9508030.

9. Bennion, H.; Duigan, C.A.; Haworth, E.Y.; Allott, T.E.H.; Anderson, N.J.; Juggins, S.; Monteith, D.T. The Anglesey Lakes, Wales, UK — Changes in Trophic Status of Three Standing Waters as Inferred from Diatom Transfer Functions and Their Implications for Conservation. *Aquatic Conservation: Marine and Freshwater Ecosystems* **1996**, *6*, 81–92, doi:10.1002/(SICI)1099-0755(199606)6:2<81::AID-AQC174>3.0.CO;2-G.
10. Rioual, P. Diatom Assemblages and Water Chemistry of Lakes in the French Massif Central: A Methodology for Reconstruction of Past Limnological and Climate Fluctuations during the Eemian Period. Doctoral, UCL (University College London), 2000.
11. Maslennikova, A.; Udachin, V.; Deryagin, V.; Artemyev, D.; Filippova, K.; Gulakov, V.; Udachin, N.; Aminov, P. Sediment Records of Lake Eutrophication and Oligotrophication under the Influence of Human Activity and Climate Warming in the Urals Metallurgical Region (Russia). *Hydrobiologia* **2023**, *850*, 1669–1698, doi:10.1007/s10750-023-05177-8.
12. Kratzer, C.R.; Brezonik, P.L. A Carlson-Type Trophic State Index for Nitrogen in Florida Lakes1. *JAWRA Journal of the American Water Resources Association* **1981**, *17*, 713–715, doi:10.1111/j.1752-1688.1981.tb01282.x.
13. Stewart, R.E.; Kantrud, H.A. *Classification of Natural Ponds and Lakes in the Glaciated Prairie Region*; U.S. Fish and Wildlife Service, Bureau of Sport Fisheries and Wildlife, 1971;
14. Guiry, M.D.; Guiry, G.M. AlgaeBase. World-Wide Electronic Publication, National University of Ireland, Galway 2023.
15. Proshkina-Lavrenko, A.I. *Diatomovye vodorosli SSSR. Iskopaemye i sovremennye [Diatoms of the USSR. Fossil and modern]*; Nauka Publ.: St-Petersburg, 1974;
16. ter Braak, C.J.F.; Juggins, S. Weighted Averaging Partial Least Squares Regression (WA-PLS): An Improved Method for Reconstructing Environmental Variables from Species Assemblages. *Hydrobiologia* **1993**, *269*, 485–502, doi:10.1007/BF00028046.
17. Birks, H.J.B.; Birks, H.J.B. D.G. Frey and E.S. Deevey Review 1: Numerical Tools in Palaeolimnology – Progress, Potentialities, and Problems. *Journal of Paleolimnology* **1998**, *20*, 307–332, doi:10.1023/A:1008038808690.
18. Birks, H.J.B.; Simpson, G.L. 'Diatoms and pH Reconstruction' (1990) Revisited. *J Paleolimnol* **2013**, *49*, 363–371, doi:10.1007/s10933-013-9697-7.
19. Juggins, S. C2 (Version 1.5). Software for Ecological and Palaeoecological Data Analysis and Visualisation 2007.