**Appendix S1**

**SUPPORTING INFORMATION**

Chlorophyll a – Nutrient and Temperature Relationships, and Predictions for Lakes across perialpine and Balkan Mountain Regions

Oskar Kärcher, Christopher T. Filstrup, Mario Brauns, Orhideja Tasevska, Suzana Patceva, Ariane Walz, Karin Frank, Danijela Markovic

## Fig. S1 Relationship between each exogenous variable and log(Chl-a) for GAMs and the respective two variable sets. The plots display the influence on log(Chl-a) as spline functions of (a) log(TN), (b) log(TP), (c) TN:TP, (d) log(MaxDepth), (d) LSWT (°C) and (e) altitude (m) while all remaining variables are fixed. The area around the influence curve depicted by dotted lines represents the uncertainty, i.e. the 95% confidence interval.



## Fig. S2 Two-dimensional partial dependence plots for combinations of TP and each remaining exogenous variable for the BRT altitude model. log(Chl-a) is depicted as a function of (a) log(TP) and log(MaxDepth), (b) log(TP) and log(TN), (c) log(TP) and altitude (m), and (d) log(TP) and TN:TP, accounting for the averaged effects of the other variables.



## Fig. S3 Three-dimensional partial dependence plot of TP, TN and MaxDepth for the BRT LSWT model. log(Chl-a) is depicted as a function of log(TP), log(TN) and log(MaxDepth) accounting for the averaged effects of the remaining variables. The scale indicates the log10-transformed Chl-a concentration.



## Fig. S4 Three-dimensional partial dependence plot of TP, TN and MaxDepth for the BRT altitude model. log(Chl-a) is depicted as a function of log(TP), log(TN) and log(MaxDepth) accounting for the averaged effects of the remaining variables. The scale indicates the log10-transformed Chl-a concentration.



## Fig. S5 Three-dimensional partial dependence plot of TP, TN and LSWT for the BRT LSWT model. log(Chl-a) is depicted as a function of log(TP), log(TN) and LSWT (°C) accounting for the averaged effects of the remaining variables. The scale indicates the log10-transformed Chl-a concentration.



## Fig. S6 Three-dimensional partial dependence plot of TP, TN and altitude for the BRT altitude model. log(Chl-a) is depicted as a function of log(TP), log(TN) and altitude (m) accounting for the averaged effects of the remaining variables. The scale indicates the log10-transformed Chl-a concentration.



## Table S1 Morphometric characteristics of the studied perialpine and central Balkan mountain lakes.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Lake** | **Lon** | **Lat** | **Altitude  (m)** | **Surface Area (km²)** | **Mean  depth (m)** | **Max depth (m)** | **Volume  (109 m3)** |
| Ammer | 11.13 | 47.98 | 533 | 46.6 | 38 | 81 | 1.75 |
| Biel | 7.17 | 47.08 | 429 | 39.3 | 31 | 74 | 1.18 |
| Brienz | 8.00 | 46.74 | 564 | 29.8 | 173 | 260 | 5.17 |
| Chiem | 12.47 | 47.89 | 518 | 79.9 | 26 | 73 | 2.05 |
| Como | 9.26 | 46.01 | 198 | 145.9 | 154 | 410 | 22.5 |
| Garda | 10.64 | 45.54 | 65 | 369.9 | 136 | 346 | 50.35 |
| Iseo | 10.07 | 45.74 | 186 | 61.8 | 124 | 251 | 8.1 |
| Maggiore | 8.57 | 45.90 | 194 | 213.0 | 177 | 372 | 37.5 |
| Ohrid | 20.73 | 41.04 | 693 | 358.0 | 164 | 288 | 58.64 |
| Prespa | 21.01 | 40.93 | 853 | 274.0 | 16 | 54 | 5.0 |
| Starnberg | 11.31 | 47.90 | 584 | 56.4 | 53 | 128 | 2.99 |
| Thun | 7.71 | 46.70 | 558 | 48.4 | 136 | 217 | 6.5 |
| Zurich | 8.73 | 47.2 | 406 | 67.7 | 52 | 143 | 3.9 |

## Table S2 Adjusted R² of GAMs and PDMs with random/fixed individual effects and with pooled data. We note that for BRTs adjusted R² could not be computed.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **PDM** | | | **GAM** |
| Random individual effects | Fixed  individual effects\* | Pooled |
| LSWT variable set | 0.4323 | -0.0139 | 0.7235 | 0.7823 |
| Altitude variable set | 0.4340 | -0.0172 | 0.7280 | 0.7960 |

\*Note that the usage of fixed individual effects does not incorporate influences of static variables across individuals.

|  |  |  |
| --- | --- | --- |
| **Variables** | **Model** | |
| **LSWT variable set** | **Altitude variable set** |
| constant | 2.2169\*\*\* | 2.0625\*\*\* |
| log(TN) | 0.3932\*\*\* | 0.4018\*\*\* |
| log(TP) | 0.5740\*\*\* | 0.5355\*\* |
| TN:TP | -0.0008 | -0.0009 |
| log(MaxDepth) | -0.2697\*\*\* | -0.2446\*\*\* |
| LSWT | -0.0094 | - |
| Altitude | - | -0.0002\*\*\* |

**Table S3** Computed coefficients of PDMs and the corresponding statistical significance. Note that for both models the pooled PDM was used. The significance (p-value) of the computed coefficients was determined with heteroscedasticity robust standard errors.

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01.

**Table S4** Modelling parameters (tree complexity – *tc*, learning rate – *lr*, bag fraction – *bf*, number of trees – *nt*) of BRT models.

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | | **BRT** | |
| LSWT variable set | Altitude variable set |
| **modelling  parameter** | *tc* | 4 | 4 |
| *lr* | 0.011 | 0.004 |
| *bf* | 0.72 | 0.64 |
| *nt* | 1,850 | 3,300 |

The best combination of the parameters *tc*, *lr*, and *bf*, which generated an optimal tree number of over 1,000 trees, was determined via randomly dividing the dataset 10 times into a calibration (80%) and a testing dataset (20%) and by computing the mean squared error (MSE) for the testing dataset each time. The smallest average MSE based on the testing dataset was used to identify the optimal combination.

**Table S5** Meanrelative variable importance (normalized to 100%) resulting from the validation of BRT models.

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | | **BRT** | |
| LSWT variable set | Altitude variable set |
| **variable** | log(TN) | 9.7% | 8.8% |
| log(TP) | 52.8% | 51.0% |
| TN:TP | 4.4% | 3.8% |
| log(MaxDepth) | 26.9% | 21.8% |
| LSWT | 6.2% | - |
| Altitude | - | 14.6% |

The validation was conducted by random sampling of the data, leading to a calibration (80%) and validation (20%) dataset. After 100 repetitions, the average variable importance from the calibrations was computed.

**Table S6** RMSE (in g/l) of the predictions resulting from the best performing approach, BRTs, for the selected perialpine and central Balkan mountain lakes and annual mean Chl-a levels.

|  |  |  |
| --- | --- | --- |
| **Lake** | **RMSE (g/l)** | |
| **BRT LSWT** | **BRT Altitude** |
| Ammer | 2.08 | 2.19 |
| Biel | 2.12 | 2.33 |
| Brienz | 1.35 | 1.34 |
| Chiem | 2.43 | 2.22 |
| Como | 2.16 | 2.14 |
| Garda | 1.50 | 1.33 |
| Iseo | 2.74 | 1.23 |
| Maggiore | 1.39 | 1.64 |
| Ohrid | 1.18 | 0.80 |
| Starnberg | 0.57 | 0.71 |
| Thun | 0.59 | 0.58 |
| Zurich | 2.95 | 2.93 |
| mean | 1.76 | 1.62 |

To test transferability of the results to the lakes from mountain regions we excluded each of the considered lakes that occurred in the calibration dataset and predicted the log-transformed Chl-a value for the excluded lake. Predictions were then transformed back, leading to RMSEs in g/l. Note that the RMSE could only be calculated if more than one observation was available.