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

Chlorophyll a relationships with nutrients and temperature, and predictions for lakes across perialpine and Balkan mountain regions

Suggested citation referring to the original publication:
Inland Waters 10 (2020) 1, pp. 29 - 41
DOI: https://doi.org/10.1080/20442041.2019.1689768
ISSN: 2044-2041, 2044-205X

Journal article | Version of record

INLAND WATERS
2020, VOL. 10, NO. 1, 29–41
https://doi.org/10.1080/20442041.2019.1689768

Chlorophyll *a* relationships with nutrients and temperature, and predictions for lakes across perialpine and Balkan mountain regions

Oskar Kärcher [a,b], Christopher T. Filstrup [c], Mario Brauns[d], Orhideja Tasevska[e], Suzana Patceva[f], Niels Hellwig [a,b], Ariane Walz[a], Karin Frank[d,g,h] and Danijela Markovic[b]

[a]Institute of Earth and Environmental Sciences, University of Potsdam, Potsdam, Germany; [b]Faculty of Business Management and Social Sciences, Osnabrück University of Applied Sciences, Osnabrück, Germany; [c]Large Lakes Observatory & Minnesota Sea Grant, University of Minnesota, Duluth, MN, USA; [d]Department of Ecological Modelling, UFZ – Helmholtz Centre for Environmental Research, Leipzig, Germany; [e]Department of Zooplankton, Hydrobiological Institute, Ohrid, Macedonia; [f]Department of Phytoplankton, Hydrobiological Institute, Ohrid, Macedonia; [g]Institute for Environmental System Research, Osnabrück University, Osnabrück, Germany; [h]iDiv – German Centre for Biodiversity Research Halle-Jena-Leipzig, Leipzig, Germany

ABSTRACT
Model-derived relationships between chlorophyll *a* (Chl-*a*) and nutrients and temperature have fundamental implications for understanding complex interactions among water quality measures used for lake classification, yet accuracy comparisons of different approaches are scarce. Here, we (1) compared Chl-*a* model performances across linear and nonlinear statistical approaches; (2) evaluated single and combined effects of nutrients, depth, and temperature as lake surface water temperature (LSWT) or altitude on Chl-*a*; and (3) investigated the reliability of the best water quality model across 13 lakes from perialpine and central Balkan mountain regions. Chl-*a* was modelled using in situ water quality data from 157 European lakes; elevation data and LSWT in situ data were complemented by remote sensing measurements. Nonlinear approaches performed better, implying complex relationships between Chl-*a* and the explanatory variables. Boosted regression trees, as the best performing approach, accommodated interactions among predictor variables. Chl-*a*–nutrient relationships were characterized by sigmoidal curves, with total phosphorus having the largest explanatory power for our study region. In comparison with LSWT, utilization of altitude, the often-used temperature surrogate, led to different influence directions but similar predictive performances. These results support utilizing altitude in models for Chl-*a* predictions. Compared to Chl-*a* observations, Chl-*a* predictions of the best performing approach for mountain lakes (oligotrophic–eutrophic) led to minor differences in trophic state categorizations. Our findings suggest that both models with LSWT and altitude are appropriate for water quality predictions of lakes in mountain regions and emphasize the importance of incorporating interactions among variables when facing lake management challenges.

ARTICLE HISTORY
Received 21 September 2018
Accepted 2 November 2019

KEYWORDS
chlorophyll *a*; nutrients; Ohrid-Prespa region; perialpine lakes; water temperature

Introduction

Despite water quality improvements in many European lakes, eutrophication is still the most serious threat to lake water quality (Poikane et al. 2014). Eutrophication and water quality status are often evaluated by using phytoplankton biomass as one indicator. Specifically, chlorophyll *a* (Chl-*a*) is used as a proxy of primary producer biomass (cf. Kasprzak et al. 2008) and for defining ecologically relevant lake water quality targets (Poikane et al. 2014). Understanding Chl-*a*–total nitrogen (TN) and Chl-*a*–total phosphorus (TP) relationships is critical to understanding lake ecosystem health and management (Rapport et al. 1998). Studies on Chl-*a*–nutrient relationships have a long history (Sakamoto 1966, Dillon and Rigler 1974) and emphasize a need for accurate models. Most empirical studies have shown TP to be a better predictor of Chl-*a* than TN, supporting the view that phosphorus more frequently limits the production of phytoplankton biomass in lakes (cf. Abell et al. 2012), but TN may colimit phytoplankton biomass under certain conditions (Elser et al. 2007, Bracken et al. 2015). Also, McCauley et al. (1989) found a significant interaction term between TN and TP when predicting Chl-*a*, in which TN has a large influence on Chl-*a* at high TP. Filstrup and Downing (2017) found a similar importance of TN on predicting Chl-*a* depending on

CONTACT Oskar Kärcher o.kaercher@hs-osnabrueck.de Faculty of Business Management and Social Sciences, Osnabrück University of Applied Sciences, Caprivistr. 30A, Osnabrück 49076, Germany
Supplemental data for this article can be accessed https://doi.org/10.1080/20442041.2019.1689768.

TP concentration, emphasizing modelling approaches that can incorporate interactions among predictor variables.

Most previous studies have documented relationships between Chl-*a* and TN or TP (Phillips et al. 2008, Abell et al. 2012) with different linear approaches on log-transformed axes ranging from ordinary least squares (Bachmann et al. 2012) to panel data models (Magumba et al. 2014). Linear methods can be limited in describing complex interactions between exogenous and endogenous variables because only one influence direction for each variable is possible. Nonlinear methods, however, can account for varying responses along environmental gradients. Few studies have considered nonlinear modelling approaches. For example, McCauley et al. (1989) noted sigmoidal Chl-*a*–TP relationships by using nonlinear regression, a finding subsequently supported by Filstrup et al. (2014). Furthermore, Hollister et al. (2016) modelled the trophic state with random forest (Breiman 2001), and Lu et al. (2016) developed an artificial neural network model (e.g., McCulloch and Pitts 1943) to predict Chl-*a* for a lake in the United States. Although the tendency to model these relationships by using nonlinear approaches is growing, many questions remain regarding model form and the accuracy of predictions, especially across mountain regions.

Temperature can strongly influence Chl-*a* concentrations observed at given nutrient concentrations (e.g., Kraemer et al. 2017); however, in some cases water quality–temperature relationships are inferred from lake surface water temperature (LSWT) surrogates. Magumba et al. (2014), for example, used altitude to control for the effect of temperature on the Chl-*a* concentration. Comparisons between Chl-*a*–altitude and Chl-*a*–LSWT relationships are insufficiently explored. Thus, the Chl-*a*–temperature versus Chl-*a*–altitude relationships need to be further investigated. The latter is particularly important for mountain lakes and lakes in the foothills, which are regarded as highly susceptible to environmental perturbations such as warming (Huber et al. 2005, Battarbee et al. 2009, Markovic et al. 2017). Preserving and improving the ecosystem services provided by lakes requires information on the level of ongoing changes as well as scenarios to estimate possible future developments.

Here, we investigated Chl-*a* concentrations of European lakes across a broad trophic gradient (from ultra-oligotrophic to hypertrophic) using linear and nonlinear statistical modelling approaches: panel data models (PDMs), generalized additive models (GAMs), and boosted regression trees (BRTs). Different statistical approaches ensure the consistency of resulting parameterized relationships. In addition to in situ measurements of the nutrients (TN, TP) and lake morphometric parameters (maximum depth), Chl-*a* models included LSWT in situ measurements supplemented by remote sensing-based LSWT data (MacCallum and Merchant 2013, Riffler et al. 2015). Our goal was to identify the statistical approach and variable set yielding the best model performance. For the best performing statistical method, water quality drivers were identified by calculating the variable importance. Comparisons of predicted and observed Chl-*a* concentrations were used to test the prediction ability of the best performing statistical method for perialpine and central Balkan mountain lakes. In summary, the objectives of this study were to (1) model Chl-*a* concentration in lakes (with the focus on performance comparisons across multiple statistical models and on the identification of water quality drivers), (2) investigate the Chl-*a* nutrient relationships and Chl-*a* temperature relationships inferred from altitude and LSWT using multivariate modelling, and (3) test the best performing model for predicting the trophic state of lakes in European mountain regions.

Methods

Water quality data

In situ observations of physical and chemical water quality variables were mainly obtained from the Waterbase-Lakes database provided by the European Environment Agency (EEA; http://www.eea.europa.eu/data-and-maps/data/waterbase-lakes-10, accessed 26 April 2017). Waterbase contains timely, reliable, and policy-relevant data collected from European Economic Area member countries through the Water Information System for Europe (WISE) data collection process managed by the EEA (see http://dd.eionet.europa.eu/datasets/3163 for more details). Because the number of observations for the aggregation period “summer” was low in the Waterbase-Lakes dataset, we focused on the annual mean values between 1989 and 2012 for Chl-*a*, TP, TN, Secchi depth, and LSWT. Annual mean values resulting from an in situ aggregation length <10 months were excluded from the analyses because of their high uncertainty. For lakes with multiple monitoring stations, only stations with the longest series of observations of the parameters of interest were included. Lake maximum depth (MaxDepth) and surface elevation (altitude) were considered as static parameters. Information on altitude gathered from Google Earth was added if lake observations were missing altitude data. For Lake Ohrid, we additionally used water quality data obtained from the Hydrobiological Institute in Ohrid.

Because the LSWT of many lakes was not monitored or irregularly monitored, remote sensing data were used to complement and extend traditional lake sampling methods, facilitating the understanding of recent trends and the current state of lake ecosystems (Peterson and Parker 1998, Turner et al. 2003, McPhearson and Wallace 2008). Specifically, in situ observations of LSWT were complemented by the satellite-based LSWT data derived from the Riffler et al. (2015) dataset and from the ARC-Lake (ATSR Reprocessing for Climate Lake Surface Temperature) dataset (http://www.laketemp.net/home/, accessed April 2017). The data are mainly based on Advanced Along Track Scanning Radiometers (AATSRs) and the Advanced Very High Resolution Radiometer (AVHRR) and were most abundant from 1989 to 2013.

The dataset used to calibrate the Chl-*a* models included 721 sets of annual mean values of the parameters Chl-*a*, TP, TN, and LSWT (Table 1). The dataset (n = 721) covered 157 lakes with various trophic states. Consequently, lakes with low Chl-*a* (min = 0.4 µg/L), TN (min = 50 µg/L), and TP (min = 2 µg/L) concentrations as well as lakes with high Chl-*a* concentrations (>500 µg/L) were included. The highest amount of annual mean data was available inter alia for Lake Vesterborg (24 yr). The lakes in the dataset were distributed across Europe, ranging from southern to northern (approximately 37°N to 60°N) and western to eastern (approximately 9°W to 29°E) Europe (Fig. 1a).

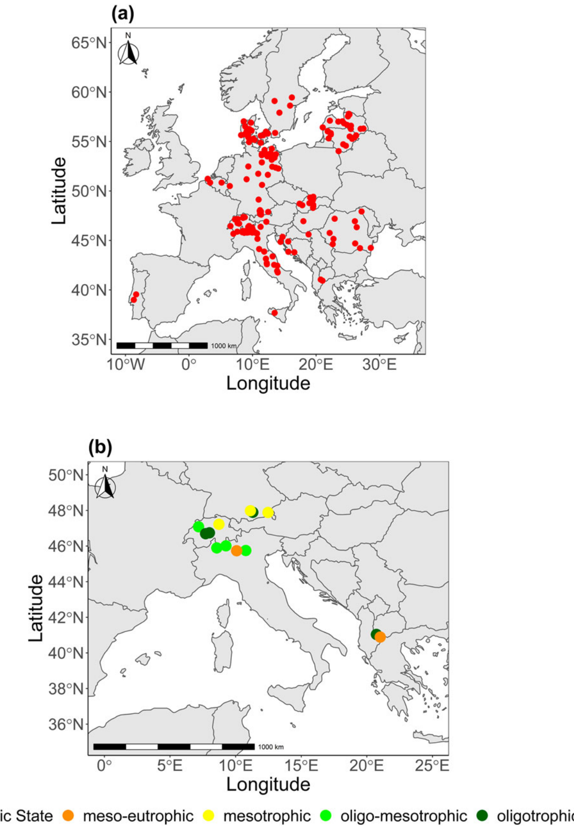

Figure 1. Locations of (a) the 157 European lakes used for modelling water quality and (b) the 13 perialpine and central Balkan mountain lakes including trophic states. For the 13 selected lakes in (b) the trophic state assessment was based on combinations of annually averaged values of Chl-*a*, TP, and Secchi depth from 2005 to 2008.

Chlorophyll a modelling

Chl-*a* was modelled using 2 different variable sets, each including TN, TP, TN:TP (by weight), and MaxDepth, with all variables except TN:TP $\log_{10}$-transformed prior to analysis (hereafter log). To identify variations resulting from the utilization of LSWT and LSWT surrogates, we extended the basic variable set with either LSWT or altitude.

The variable selection for modelling Chl-*a* was based on previous studies finding that TP and TN influence Chl-*a* concentration, either in isolation or together (Gunkel and Casallas 2002, Abell et al. 2012). Furthermore, TN:TP was included for the identification of limiting nutrients (Bachmann et al. 2003) and its influence on Chl-*a* nutrient relationships (Prairie et al. 1989). The utilization of MaxDepth accounts for the residence time of water (Kalff 2002, Londe et al. 2016) and the vertically mixed proportion of the water column, which can influence Chl-*a* nutrient relationships (e.g., Weithoff et al. 2000). Increasing temperature has been demonstrated to positively affect Chl-*a* concentration, yet temperature surrogates are generally used (e.g., Carvalho et al. 2009). Although water transparency influences a wide range of biogeochemical processes, Secchi depth was not included as a predictor because of the high influence of Chl-*a* concentration on water transparency.

Relationships between Chl-*a* and the 2 investigated variable sets were parameterized with 3 different modelling approaches: PDMs, GAMs, and BRTs. The statistical approaches are available in the R packages *plm* (PDM; Croissant and Millo 2008), *mgcv* (GAM; Wood 2011), and *gbm* (BRT; Ridgeway 2017). PDMs account for

Table 1. Summary statistics of Chl-*a*, the explanatory variables, and geographical characteristics of the 157 European lakes in the modelling dataset.

Variable	Unit	Minimum	Mean	Maximum	Standard deviation
Chl-*a*	µg/L	0.4	26.5	551.7	48.13
TN	mg/L	0.05	1.72	8.17	1.37
TP	mg/L	0.002	0.08	0.59	0.09
TN:TP (by weight)		0.59	42.11	316.67	43.66
MaxDepth	m	1	37.8	410	70.50
LSWT	°C	4.4	11.0	22.7	2.59
Altitude	m	0	145.5	2432	275.53
Lon	°E	−8.7	11.6	28.6	4.27
Lat	°N	37.7	53.1	59.5	4.43

possible linear relationships. Three PDM types were used: (1) fixed individual effects, (2) random individual effects, and (3) pooled. Therefore, we compared the performance of these PDM types for the 2 variable sets and used the best performing PDM for the respective variable set for further analyses. For the final PDMs, the standard errors of the coefficients were computed with a robust covariance matrix estimator to address potential heteroscedasticity (Long and Ervin 2000, Greene 2012). In contrast to PDMs, GAMs can account for nonlinear relationships between the explanatory and dependent variables and consequently do not assume a linear influence of each explanatory variable on Chl-*a*. The third approach (BRT) takes into account the possibility of relationships among the exogenous variables. BRTs are nonparametric and able to express nonlinear relationships and interactions between predictors based on the structure of decision trees (Elith et al. 2008). The single trees of the BRT model are determined successively by fitting the residuals at each step. For this procedure, 4 parameters in the BRT model are defined manually: the learning rate (*lr*) defines the contribution of each tree to the whole model, tree complexity (*tc*) indicates the final number of nodes, bag fraction (*bf*) introduces randomness into the model, and (*nt*) represents the number of trees. We determined *nt* for different combinations of *lr*, *tc*, and *bf* with the R function *gbm.step* of the R package *dismo* (Hijmans et al. 2016). Following Elith et al. (2008), combinations leading to fewer than the recommended optimum of at least 1000 trees were excluded from further examination. Subsequently, the 4 parameters yielding the smallest average testing mean squared error (MSE), determined by randomly dividing the dataset 10 times into a calibration (80%) and a testing dataset (20%), were set as optimal model values. The calibrated BRT model can be viewed as a sum of *nt* trees, each multiplied by *lr*, and including variable interactions if $tc > 1$, whereas *bf* represents the random fraction of the data used to propose the next tree.

Determination of the best performing approach was based on indices that compare predicted and observed Chl-*a* values; that is, the fraction of variation explained (R^2) and root mean squared error (RMSE). Standard model selection criteria, such as the Akaike information criterion (AIC), are unsuitable for nonparametric models. Here, the performance evaluation was conducted via bootstrapping, with random data splitting into calibration (80%) and validation (20%) datasets, repeated 100 times. Based on the bootstrapping data samples, the mean R^2 and RMSE from the calibration and the validation dataset were assessed. The model resulting in the highest validation mean R^2 and lowest mean RMSE is hereafter referred to as the "best performing approach."

For all models, we identified the influence direction (partial dependence curves) of each explanatory variable, whereas the variable importance and the multidimensional partial dependence plots were computed only for the best performing approach. The influence direction (partial dependence) for nonlinear models is depicted as a function of an individual explanatory variable displaying the response variable while all remaining variables are either fixed or kept at their mean value. The variable importance for BRTs is based on the number of selections for splitting, weighted by the squared improvement to the model as a result of each split, and averaged over all trees (Friedman and Meulman 2003). Uncertainty estimation additionally included testing the variable importance accuracy by computing the average variable importance from 100 repetitions with the calibration dataset for the best performing approach. The importance of each explanatory variable for the best performing approach was used to identify Chl-*a* concentration drivers. After identifying the variables with the highest importance, we additionally investigated multidimensional partial dependence plots by using the *gbm.plot* function integrated in the R-package *gbm* (see also Friedman and Meulman 2003, Friedman and Popescu 2008, Lampa et al. 2014).

Predicting chlorophyll a *concentration in perialpine and central Balkan mountain lakes*

Our study included major large lakes located in or near the European Alps (cf. Riffler et al. 2015) and 2 major freshwater biodiversity hotspots, Balkan lakes Ohrid and Prespa (Fig. 1b), each with significantly different morphometric and trophic characteristics. The surface areas varied from 29.8 km^2 (Lake Brienz) to 369.9 km^2 (Lake Garda), with maximum depths ranging from 73 m (Lake Chiem) to >400 m (Lake Como; Supplemental Table S1). The trophic status spanned from oligotrophic to meso-eutrophic (Table 2). Specifically, 4 lakes were classified as oligotrophic (Brienz, Ohrid, Starnberg, Thun), 4 as oligo-mesotrophic (Biel, Como, Garda, and Maggiore), 3 as mesotrophic (Ammer, Chiem, and Zurich), and 2 as meso-eutrophic (Iseo and Prespa). Lake selection was guided by the availability of in situ water quality observations and satellite-based lake surface water temperature data.

To evaluate the prediction ability for perialpine and central Balkan mountain lakes, we predicted Chl-*a* and the trophic state lake-by-lake for the 13 lakes in the modelling dataset with the best performing approach. For the Chl-*a* prediction, each respective lake was excluded beforehand in the model calibration. Subsequently, the Chl-*a* prediction was assessed by comparing predicted

Table 2. Organisation for Economic Co-operation and Development (OECD) lake classification. The classification of the trophic status was conducted following OECD fixed boundary recommendations (Premazzi and Chiaudani 1992). Chlorophyll *a* (Chl-*a*), total phosphorus (TP), and Secchi depth values represent the minimum and maximum of the annual averages based on observations collected between 2005 and 2008. Missing required water quality parameters are represented by —.

Lake	Chl-*a* (μg/L)	TP (mg/L)	Secchi depth (m)	Trophic state
Ammer	3.22–4.29	0.008–0.008	4.0–4.0	mesotrophic
Biel	1.00–1.95	0.015–0.017	—	oligo-mesotrophic
Brienz	0.55–0.93	0.004–0.009	—	oligotrophic
Chiem	4.42–4.60	0.008–0.011	3.7–4.8	mesotrophic
Como	1.60–7.20	0.005–0.026	4.2–12.2	oligo-mesotrophic
Garda	1.69–3.75	0.019–0.030	6.2–12.5	oligo-mesotrophic
Iseo	2.73–6.35	0.051–0.099	4.4–5.4	meso-eutrophic
Maggiore	0.83–4.67	0.003–0.014	4.5–10.6	oligo-mesotrophic
Ohrid	0.53	0.008	—	oligotrophic
Prespa	4.87–7.97	0.044–0.060	—	meso-eutrophic
Starnberg	1.83–1.91	0.006–0.006	5.9–7.0	oligotrophic
Thun	1.15–1.65	0.003–0.006	—	oligotrophic
Zurich	4.68–6.33	0.034–0.035	—	mesotrophic

and observed trophic states for 2005–2008 and used to calculate the RMSE.

Results

Chlorophyll a modelling

Log(Chl-*a*) concentration was positively related to log(TP) ($R^2 = 0.64$, $p < 0.001$) and log(TN) ($R^2 = 0.40$, $p < 0.001$) concentration (Fig. 2a–b). Values of log(TN) and log(TP) were positively correlated ($R^2 = 0.43$, $p < 0.001$; Fig. 2c), but LSWT was not significantly correlated with log(Chl-*a*) (Fig. 2d).

Among the studied PDMs (fixed individual effects, random individual effects, and pooled), adjusted R^2 was highest for the pooled modelling data (Supplemental Table S2). Therefore, we continued our analysis using the pooled data without including lake specific individual effects. For both variable sets, the coefficients, and thus influence directions, of the pooled data model showed significant positive influences of TN and TP on Chl-*a*, whereas MaxDepth had a significant negative influence (Supplemental Table S3). For individual variable sets, altitude had a significant negative influence on Chl-*a*, whereas LSWT had no significant effect (Supplemental Table S3).

In general, all models had moderate to good mean R^2 validation (Table 3). Compared to the model using LSWT, the use of altitude as an LSWT surrogate resulted in marginal improvement of validation mean R^2 and RMSE by 0.02 and 0.01 maximum, respectively. Furthermore, with a mean validation R^2 of 0.82–0.84 and a mean validation RMSE of 0.23–0.24 for the LSWT and altitude models, BRTs emerged as the best performing approach (corresponding BRT model parameter specifications in Supplemental Table S4). The lowest performance was attributed to PDMs, suggesting that using nonlinear models and models that allow interactions among exogenous variables improve the prediction modelling accuracy of Chl-*a*. The greatest differences between the summary statistics for the calibration and the validation samples were found in BRTs.

Overall, the partial dependence curves (i.e., influence directions) for the 2 best performing approaches, BRTs (Fig. 3) and GAMs (Supplemental Fig. S1), indicated similar relationships between Chl-*a* and the explanatory variables. However, the uncertainty of the GAM-based estimates increased near extreme predictor values. The relationship between TN and Chl-*a* in BRT models was nonlinear and showed a slight increase of Chl-*a* with increasing TN concentration until a peak was reached at log(TN) = 0.4 (TN = 2.5 mg/L; Fig. 3). Chl-*a* and TP displayed a sigmoidal relationship with an acceleration of the positive slope at log(TP) = −1.9 (TP = 0.013 mg/L) and a deceleration at around log(TP) = −1.3 (TP = 0.05 mg/L; Fig. 3). The partial dependence curve describing the influence direction for TN:TP showed a slight decrease with increasing TN:TP. For TN:TP ratios above ∼170 (by weight), no particular influence direction was observed. However, GAMs identified a strictly negative relationship for TN:TP with increasing uncertainty for greater TN:TP ratios (Supplemental Fig. S1). For MaxDepth, both GAM and BRT suggested a positive influence until log(MaxDepth) = 0.6 (MaxDepth ∼4 m) and a negative influence afterward (Fig. 3, Supplemental Fig. S1). Note that lakes with a maximum depth > 100 m (log(MaxDepth) > 2) were rare in the modelling dataset, implying influence directions with higher uncertainty (Supplemental Fig. S1). Similarly, both GAMs and BRTs identified a nonlinear interaction between LSWT and Chl-*a* (Fig. 3, Supplemental Fig. S1). The BRT models showed an alternating influence direction for LSWT with a slight negative tendency followed by an increase from 16 °C until 20 °C (Fig. 3). The BRT-based Chl-*a* altitude

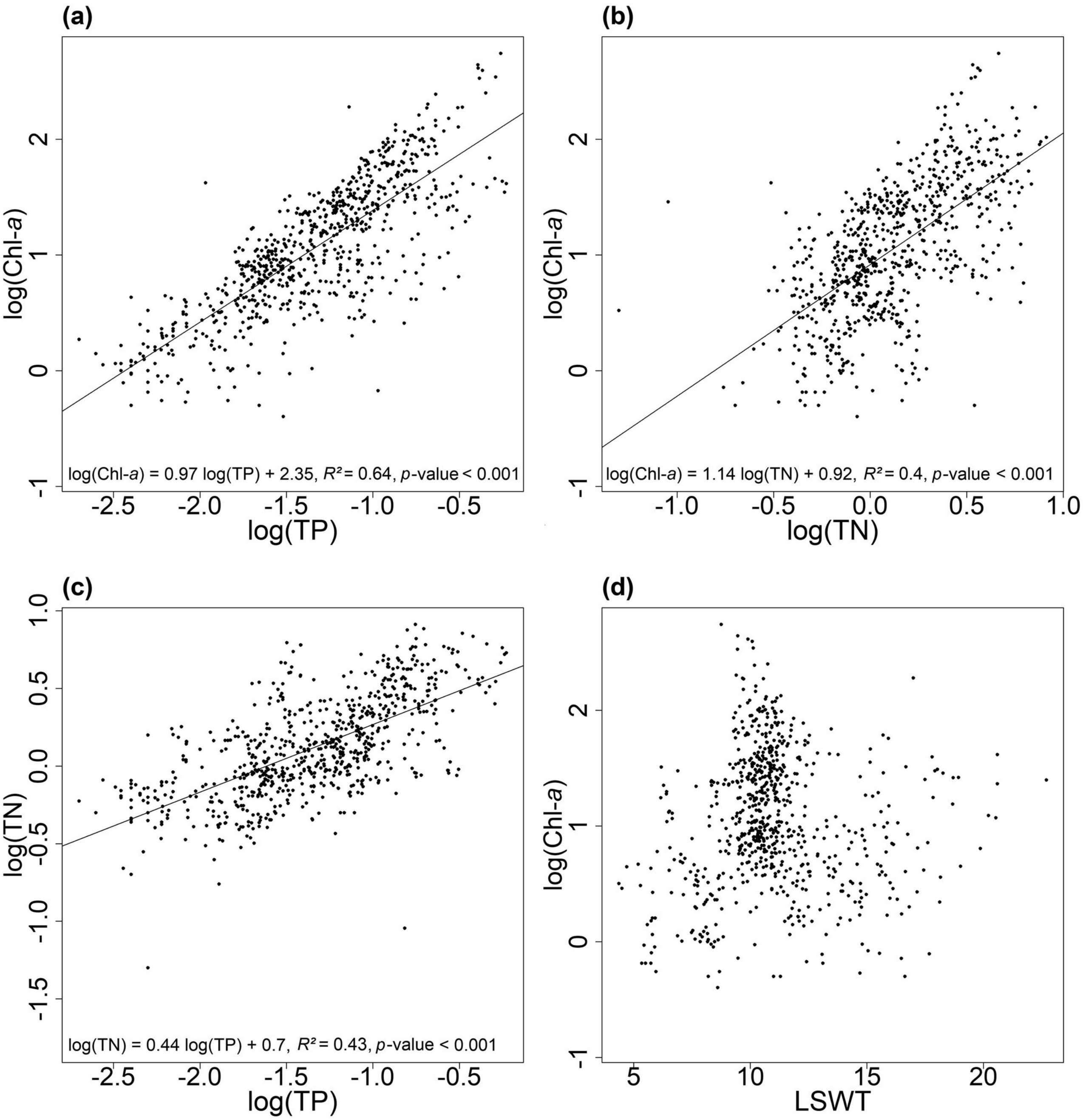

Figure 2. Bivariate scatterplots of (a) log(Chl-*a*) and log(TP), (b) log(Chl-*a*) and log(TN), (c) log(TN) and log(TP), (d) log(Chl-*a*) and LSWT (°C) and the corresponding R^2 and *p*-value for significant relationships only.

relationship showed a decrease in Chl-*a* concentration with increasing altitude (Fig. 3) while GAM showed more variation in the Chl-*a*–altitude relationship, accompanied by increasing uncertainty for lakes with an altitude above ~700 m (Supplemental Fig. S1).

Because validation and calibration of the best performing BRT models confirmed nearly identical results for variable importance (Table 4, Supplemental Table S5), we considered the following results from the calibration with the whole dataset (Table 4). TP seemed to have the strongest influence (>50%) on Chl-*a* concentration in European lakes for the BRT LSWT and altitude models (Table 4). Lake characteristic Max-Depth (26.8% and 22.3% for the LSWT and altitude model, respectively) was identified as the second most important variable, followed by TN (9.2%) for the LSWT and altitude (15.2%) for the altitude model. TN (7.8%) had an even lower variable importance in the altitude model than the LSWT model, but LSWT had a calculated variable importance of 5.8%. Both BRT models assigned TN:TP the smallest variable importance (3–4%).

Table 3. Summary statistics of the predictive accuracy of predictions of PDMs, GAMs, and BRTs including the mean R^2 and mean RMSE (for log-transformed values) for the calibration (80%) and validation dataset (20%). The validation was conducted by bootstrapping 100 times. Predictive accuracy was determined for the calibration (80%) and validation dataset (20%). Afterward, averages of the R^2 and RMSE were calculated.

		Calibration		Validation	
Method	Model	Mean R^2	Mean RMSE	Mean R^2	Mean RMSE
PDM	LSWT variable set	0.72	0.29	0.72	0.30
	Altitude variable set	0.73	0.30	0.72	0.30
GAM	LSWT variable set	0.79	0.26	0.76	0.28
	Altitude variable set	0.81	0.25	0.77	0.28
BRT	LSWT variable set	0.92	0.15	0.82	0.24
	Altitude variable set	0.92	0.17	0.84	0.23

Two-dimensional (2-D) partial dependence plots were considered for combinations of the variable with the highest variable importance (TP) and the remaining predictor variables. For the same variable combinations, the BRT LSWT and altitude model yielded similar 2-D partial dependence plots (Fig. 4, Supplemental Fig. S2).

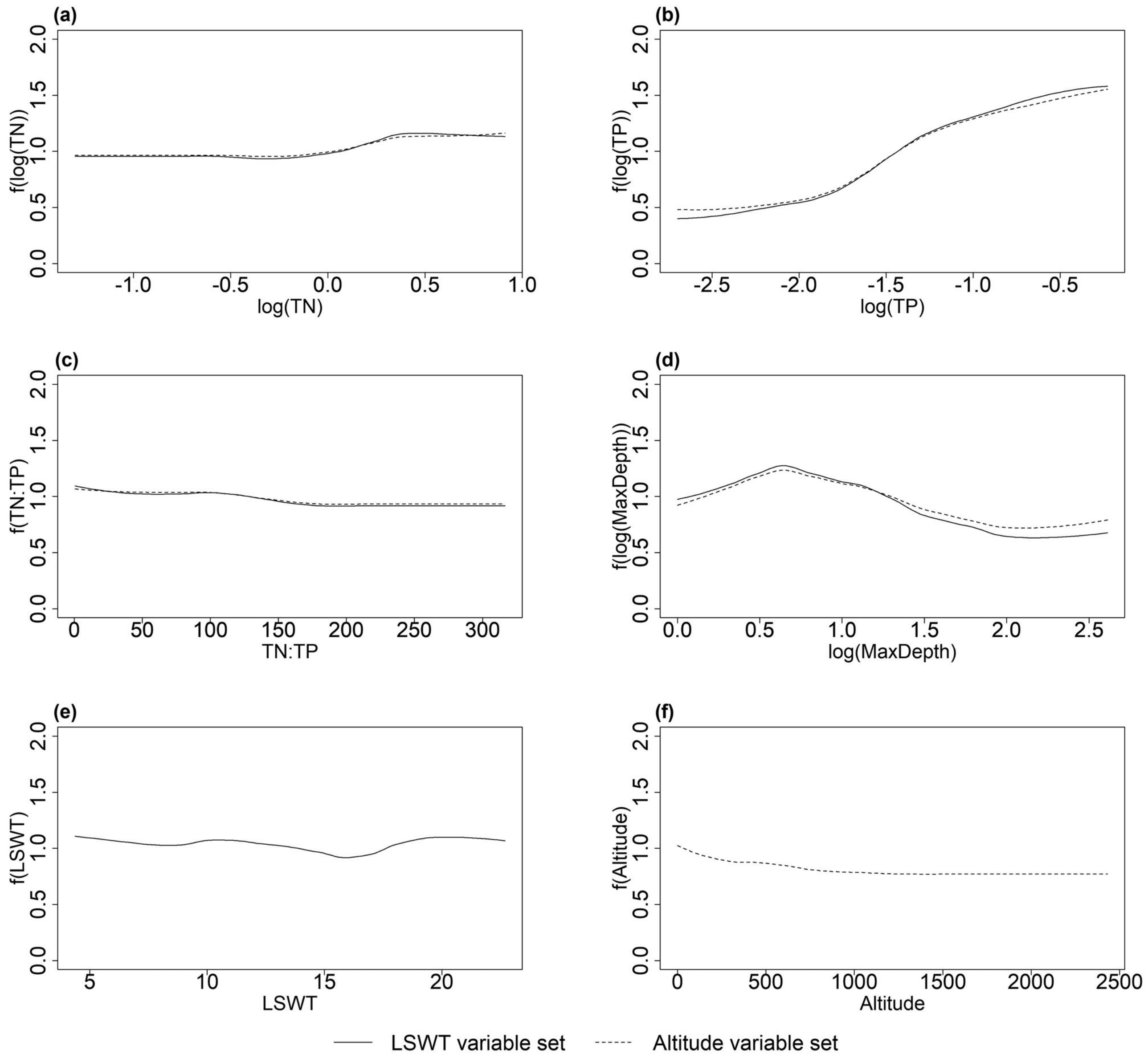

Figure 3. Smoothed partial dependence curves of each exogenous variable for the 2 BRT models. Log(Chl-*a*) is depicted as a function of (a) log(TN), (b) log(TP), (c) TN:TP, (d) log(MaxDepth), (e) LSWT (°C), and (f) altitude (m) while all remaining variables are kept at their mean values.

Table 4. Relative variable importance (as % of total) resulting from the calibration of BRT models with the whole dataset.

		BRT	
Model		LSWT variable set	Altitude variable set
Variable	log(TN)	9.2%	7.8%
	log(TP)	53.9%	51.4%
	TN:TP	4.3%	3.3%
	log(MaxDepth)	26.8%	22.3%
	LSWT	5.8%	—
	Altitude	—	15.2%

The combination of the 2 variables with the highest calculated variable importance (TP and MaxDepth) suggested a high Chl-*a* concentration for lakes with higher TP concentrations (>0.1 mg/L) and lower maximum depth (2–4 m; Fig. 4a, Supplemental Fig. S2a). With increasing maximum depth, the maximum Chl-*a* concentration decreased for a given TP concentration. The sigmoidal structure of the dependence curve and the magnitude of change along the nutrient gradient is

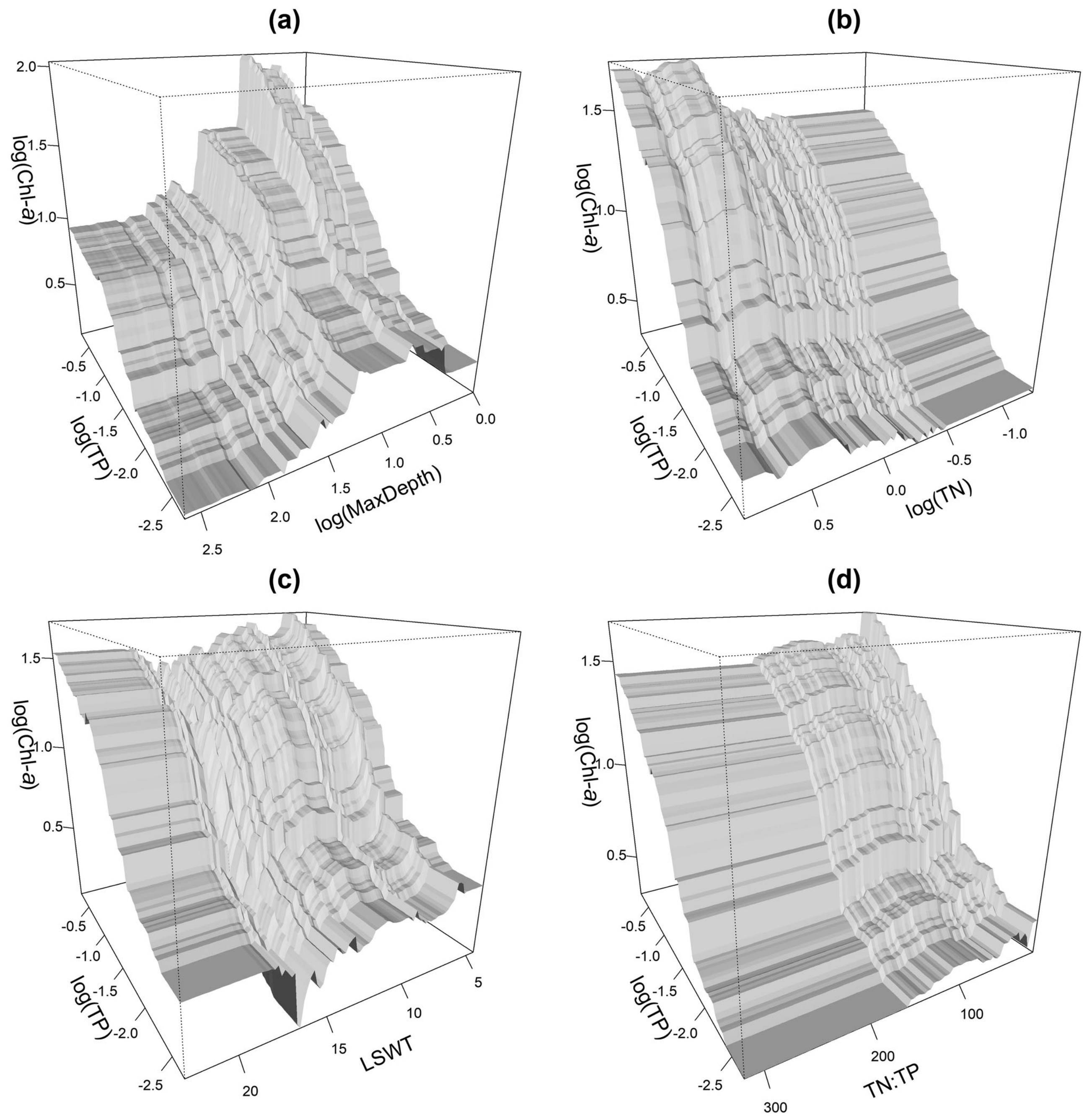

Figure 4. Two-dimensional partial dependence plots for combinations of TP and each remaining exogenous variable for the BRT LSWT model. Depicted is log(Chl-*a*) as a function of (a) log(TP) and log(MaxDepth), (b) log(TP) and log(TN), (c) log(TP) and LSWT (°C), and (d) log(TP) and TN:TP, accounting for the averaged effects of the other variables.

similar regardless of depth; however, the Chl-*a* concentration shifted to higher values at low lake depths (Fig. 4a). Furthermore, increases in nutrient loadings were accompanied by increases in Chl-*a* (Fig. 4b, Supplemental Fig. S2b), displaying high Chl-*a* concentration resulting from the combination of high TN and TP concentrations, with rising TN having a slightly stronger effect if TP was already high. However, a high TN and low TP concentration was connected to a notably lower Chl-*a* concentration than vice versa. Variable combinations with variables characterized by a lower computed variable importance (i.e., TP and LSWT, TP and altitude, or TP and TN:TP) seemed to have no clear, discernible trend (Fig. 4c–d, Supplemental Fig. S2c–d). For TP and LSWT and TP and TN:TP, the 2-D partial dependence plots were mainly driven by the sigmoidal structure manifested by TP, with no strong influence of the second variable on the Chl-*a* concentration. For TP and altitude, however, lakes with high TP concentration at lower elevations seemed to have a positive effect on Chl-*a* (Supplemental Fig. S2c).

Three-dimensional partial dependence plots visualizing 3-way variable interactions of TN, TP, and MaxDepth in the BRT LSWT and altitude model showed high Chl-*a* concentrations for lakes with low maximum depths combined with high TN and TP concentrations (Supplemental Fig. S3, S4). As the maximum depth increased, the maximum Chl-*a* concentration decreased. The maximum Chl-*a* concentration (~log(Chl-*a*) = 1.1, Chl-*a* ~ 13 µg/L) in lakes deeper than log(MaxDepth) = 1.3, MaxDepth ~ 20 m, was only reached with a combination of high TN and TP concentrations (i.e., log(TN) > 0.3, TN ~ 2 mg/L and log(TP) > −1.0, TP ~ 0.1 mg/L, respectively). In regard to the relationships between the temperature variables and nutrients, the 3-way interactions among TN, TP, and LSWT as well as TN, TP, and altitude did not vary along the respective gradients of the temperature variables (Supplemental Fig. S5, S6).

Predicting chlorophyll a concentration in perialpine and central Balkan mountain lakes

For lakes in European mountain regions, the best performing approach, BRTs, was used to predict Chl-*a* concentration and classify the resultant predictions into trophic states. Specifically, the prediction accuracy of the BRTs was evaluated by calculating the RMSE (µg/L) and comparing predicted and observed trophic states for 13 selected lakes. Differences in the average RMSE between the LSWT and altitude models were marginal, with values of 1.76 and 1.62 µg/L, respectively (Supplemental Table S6). Maximum RMSE was observed for Lake Iseo with the LSWT variable set (2.74 µg/L). In addition, for Lake Iseo, the RMSE difference between the 2 variable sets was largest (2.74 µg/L − 1.23 µg/L = 1.51 µg/L); however, for the remaining lakes the difference was <0.38 µg/L. Overall, RMSE for Lakes Starnberg and Thun was lowest (0.57–0.71 µg/L). Note that RMSE for Lake Prespa could not be calculated because more than one observation/prediction for the calculation is needed.

In comparison with the respective annual mean Chl-*a* values, varying overestimations and underestimations of Chl-*a* concentrations were present across the lakes (Table 5) but only led to minor or no variations in the trophic state assessment for the period 2005–2008, including TP and Secchi depth for the classification. For 10 of the 13 lakes, the predicted trophic state agreed with the observed water quality status. Only for Lake Biel was the observed trophic state underestimated by the prediction (oligo-mesotrophic vs. mesotrophic), whereas it was overestimated for Lakes Ammer and Chiem (mesotrophic vs. oligo-mesotrophic; Table 5).

Discussion

Chlorophyll a modelling

BRTs resulted in the overall best performance in modelling water quality for both datasets. Therefore, further investigations of Chl-*a* nutrient and temperature relationships and predictions for lakes from mountain regions were based on the results of the gradient boosted model. Because the tree complexity was >1 in both the BRT LSWT and altitude models, variable interactions were fitted in our water quality models (Elith et al. 2008). Thus, we were able to examine multidimensional partial dependence plots, which reflect the different interactions of 2 or more predictor variables (Friedman and Meulman 2003) and allow study of the detailed nature of interaction effects (Friedman and Popescu 2008).

In general, the computed variable importance (see Elith et al. 2008) as an indicator of the explanatory power can be considered a determinant of the influence extent displayed by the partial dependence plots (i.e., greater changes in Chl-*a* are expected if changes of more important variables are present). As such, for the most important variable, TP (~53%), we observed a sigmoidal Chl-*a*–TP relationship, indicating increasing Chl-*a* concentrations for increasing TP until an upper Chl-*a* maximum was reached (Filstrup et al. 2014). Multidimensional partial dependence plots emphasized the computed variable importance and dominance of TP in the variable interactions by showing greater changes in Chl-*a* concentration for increases in TP compared to the other variables considered. However, all

Table 5. Chl-*a* observations and predictions (min-max) of the BRT LSWT and altitude models, and the corresponding trophic state categorisations using the Chl-*a* observations and predictions for 2005–2008. Note that for lakes with only one prediction value, only one complete variable set necessary for a model prediction was available. The classification of the trophic status was conducted following OECD fixed boundary recommendations (Premazzi and Chiaudani 1992).

	Observation		Prediction		
Lake	Chl-*a* (μg/L)	Trophic state	BRT LSWT (μg/L)	BRT Altitude (μg/L)	Trophic state
Ammer	3.22–4.29	mesotrophic	1.39	1.27	oligo-mesotrophic
Biel	1.00–1.95	oligo-mesotrophic	3.38–4.00	3.44–3.91	mesotrophic
Brienz	0.55–0.93	oligotrophic	1.18–1.92	1.37–1.96	oligotrophic
Chiem	4.42–4.60	mesotrophic	1.54	1.90	oligo-mesotrophic
Como	1.60–7.20	oligo-mesotrophic	1.54–3.53	1.45–3.17	oligo-mesotrophic
Garda	1.69–3.75	oligo-mesotrophic	1.8–4.05	2.34–3.37	oligo-mesotrophic
Iseo	2.73–6.35	meso-eutrophic	6.59–8.86	4.92–6.97	meso-eutrophic
Maggiore	0.83–4.67	oligo-mesotrophic	1.14–2.09	1.05–1.70	oligo-mesotrophic
Ohrid	0.53	oligotrophic	2.15	1.27	oligotrophic
Prespa	4.87–7.97	meso-eutrophic	3.34	2.51	meso-eutrophic
Starnberg	1.83–1.91	oligotrophic	1.07	1.01	oligotrophic
Thun	1.15–1.65	oligotrophic	1.09–1.32	1.08–1.35	oligotrophic
Zurich	4.68–6.33	mesotrophic	2.36–3.34	2.70–2.97	mesotrophic

interaction analyses showed higher Chl-*a* concentrations when TN and TP concentration were high. The influence of TN (~8%) could be characterized by smaller expected changes in Chl-*a* for varying TN concentrations in the analyzed lakes, but still having a positive influence on phytoplankton biomass with a sigmoidal partial dependence curve and a peak of TN ≈ 2.5 mg/L. This threshold is similar to the threshold identified in Filstrup and Downing (2017), whereas after the threshold (TN > 2.5 mg/L) in our study, the relationship was flat and not negative. This differing relationship for TN above the mentioned threshold is likely because the data presented here do not include high TN values (>10 mg/L) present in the study of Filstrup and Downing (2017). The influence of TN, however, can vary with respect to TP concentrations. The 2-D interaction analysis with TP showed increases in Chl-*a* when TN concentration increased in lakes with already high phosphorus concentrations. Accordingly, Filstrup and Downing (2017) detected a markedly stronger effect of TN in TP-rich and especially hypereutrophic lakes than in mesotrophic or eutrophic lakes, implying different Chl-*a*–TN relationships in nutrient-rich lakes. In addition, for mesotrophic and eutrophic lakes, a similar sigmoidal relationship was found (Filstrup and Downing 2017). Thus, our findings support previous studies demonstrating the nonlinear response of Chl-*a* to nutrients (McCauley et al. 1989, Filstrup et al. 2014) and suggest that Chl-*a* may display greater responses to phosphorus reductions in the European lakes considered in this study.

Guildford and Hecky (2000) noted that the TN:TP ratio constitutes an important indicator of nutrient limitation, with phytoplankton growth becoming increasingly phosphorus limited at higher ratios. Here, the influence extent and directions of the least important variable TN:TP (~4%) implied only marginal negative effects on Chl-*a* for increasing TN:TP ratios and thus increasing phosphorus limitation. TN:TP is known to have little effect when the system is strongly phosphorus limited but becomes more important under more balanced conditions (e.g., Cardinale et al. 2009). Multidimensional interaction analyses also revealed no further discernible effects on water quality. Our data mainly consisted of phosphorus-limited lakes (high TN:TP). Low (<10) TN:TP ratios were only present in 62 of 721 observations. Prairie et al. (1989) found that TP explained more variance in Chl-*a* than TN with increasing TN:TP, a finding that matches the detected influence direction of TN:TP and high relative importance of TP in our data. The higher predictive ability of TP, however, does not imply lower correlation of TN with Chl-*a* for high TN:TP compared to TP (Prairie et al. 1989). In addition, according to the BRT models, MaxDepth up to ~4 m had a positive influence on Chl-*a*. This positive influence direction was also supported in the 2-D and 3-D interaction analysis but might be biased by the high number of hypereutrophic lakes (Chl-*a* > 25 μg/L; Premazzi and Chiaudani 1992) with low depths in the dataset (118 of 196 observations with MaxDepth ≤ 4 m had Chl-*a* > 25 μg/L). A decreasing trend of the partial dependence curve for MaxDepth > 4 m indicated a negative effect on Chl-*a* for increasing maximum depth, which can also be explained by lower nutrient loadings per volume and the well-established positive relationship between algal biomass and light availability (Sakamoto 1966, Scheffer 1998). The computed relative variable importance for MaxDepth (~25%) emphasizes the strong influence of lake morphometry on Chl-*a* concentration and indicates interactions among explanatory variables.

The BRT model with a negative linear partial dependence curve for the temperature surrogate altitude (e.g., Carvalho et al. 2009) confirmed a previously found positive Chl-*a*–temperature relationship, expressed as the inverse relationship between Chl-*a* and altitude. By contrast, the Chl-*a*–temperature relationship found by the BRT LSWT model in general did not confirm this positive linear influence of temperature on Chl-*a*. The Chl-*a*–LSWT relationship was characterized by slightly alternating influence directions, with only small impacts on Chl-*a* concentration. The nonlinear relationship may be explained by the reduced effect of water temperature when nutrient levels are low (Elliott et al. 2006) or different responses to warming in phytoplankton-rich lakes, where warming tends to increase Chl-*a*, and in phytoplankton-poor lakes, where warming leads to decreasing Chl-*a* (Kraemer et al. 2017). However, further analyses of variable interactions with nutrients showed no significant increases or decreases along the temperature gradient. Thus, combinations of LSWT with other predictors, especially nutrients, and their influence on phytoplankton communities should be further investigated. We note that the importance of altitude (~15%) and LSWT (~6%) in our study implied relatively small influences of temperature on Chl-*a*.

Interaction effects on water quality must be evaluated because previous studies have found different relationships among predictors for different lake settings (McCauley et al. 1989, Elliott et al. 2006, Filstrup and Downing 2017, Kraemer et al. 2017). Our results suggest that nonlinear models incorporating the underlying relationships among explanatory variables are more qualified than linear approaches. As such, the nonlinear gradient BRT method identified interactions among predictor variables (Elith et al. 2008) and performed better than other regression approaches such as GAM and PDM. The BRTs accommodated interactions among predictors in a multidimensional setting, allowing complex interactions among water quality variables to be understood. We therefore recommend using further nonlinear modelling and learning techniques that can incorporate complex variable interactions.

Predicting chlorophyll a concentration in perialpine and central Balkan mountain lakes

Prediction performances of BRTs were evaluated for 13 selected perialpine and central Balkan mountain lakes. The RMSE differences between using LSWT or altitude seem minor (Chl-*a* 1.62–1.76 μg/L), a conclusion that can also be inferred by comparing the predicted and observed trophic states. Three of 13 lakes (Ammer, Biel, and Chiem) had minor differences between the observed and predicted trophic state, which may be related to the absence of lakes with similar characteristics within the calibration dataset. Other reasons for lower accuracy of predictions can be attributed to the generalization abilities of the models (i.e., better data coverage) than precise predictions for exceptionally low or high Chl-*a* concentrations (see also Suleiman et al. 2016). Similarly, other well-performing water quality modelling approaches, such as the artificial neural network model of Lu et al. (2016), led to underestimates of high values and overestimates of low values. However, for the unusual lake systems of Lakes Ohrid and Prespa, which are characterized by a connection through karst channels (Matzinger et al. 2006), the predicted trophic states agreed well with the observations. As such, the regression tree approach yielded accurate predictions for these unique lake systems, which usually can be seen as limits to transferability of general water quality models. Nonetheless, we emphasize the need for further studies on applications of general water quality models for unique lakes from mountain regions and thus for further improvement of the models. Although the dataset included some lakes that covered longer time periods than others, overall Chl-*a* predictions of BRT models for perialpine and central Balkan mountain lakes in general provided satisfactory results. Further improvement potential exists for balancing data availability across space and time.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Current research is funded by the DFG Grant MA 6593/2-1 and the EU funded ECOPOTENTIAL (Horizon 2020 ref. 641762) project.

ORCID

Oskar Kärcher http://orcid.org/0000-0002-9418-6074
Christopher T. Filstrup http://orcid.org/0000-0003-3812-2831
Niels Hellwig http://orcid.org/0000-0003-4517-1746

References

Abell JM, Ozkundakci D, Hamilton DP, Jones JR. 2012. Latitudinal variation in nutrient stoichiometry and chlorophyll–nutrient relationships in lakes: a global study. Fundam Appl Limnol. 181:1–14.

Bachmann RW, Bigham DL, Hoyer MV, Canfield DE Jr. 2012. Factors determining the distributions of total phosphorus, total nitrogen, and chlorophyll *a* in Florida lakes. Lake Reserv Manag. 28: 10–26.

Bachmann RW, Hoyer MV, Canfield DE. 2003. Predicting the frequencies of high chlorophyll levels in Florida lakes from average chlorophyll or nutrient data. Lake Reserv Manag. 19:229–241.

Battarbee RW, Kernan M, Rose N. 2009. Threatened and stressed mountain lakes of Europe: assessment and progress. Aquat Ecosyst Health Manag. 12:118–128.

Bracken MES, Hillebrand H, Borer ET, Seabloom EW, Cebrian J, Cleland EE, Elser JJ, Gruner DS, Harpole WS, Ngai JT, et al. 2015. Signatures of nutrient limitation and co-limitation: responses of autotroph internal nutrient concentrations to nitrogen and phosphorus additions. Oikos. 124:113–121.

Breiman L. 2001. Random forests. Mach Learn. 45:5–32.

Cardinale BJ, Hillebrand H, Harpole WS, Gross K, Ptacnik R. 2009. Separating the influence of resource 'availability' from resource 'imbalance' on productivity-diversity relationships. Ecol Lett. 12:475–487.

Carvalho L, Solimini AG, Phillips G, Pietiläinen OP, Moe J, Cardoso AC, Solheim AL, Ott I, Søndergaard M, Tartari G, et al. 2009. Site-specific chlorophyll reference conditions for lakes in Northern and Western Europe. Hydrobiologia. 633:59–66.

Croissant Y, Millo G. 2008. Panel data econometrics in *R*: the plm package. J Stat Softw. 27(2):1–43.

Dillon PJ, Rigler FH. 1974. The phosphorus–chlorophyll relationship in lakes. Limnol Oceanogr. 19:767–773.

Elith J, Leathwick JR, Hastie T. 2008. A working guide to boosted regression trees. J Anim Ecol. 77:802–813.

Elliott JA, Jones ID, Thackeray SJ. 2006. Testing the sensitivity of phytoplankton communities to changes in water temperature and nutrient load, in a temperate lake. Hydrobiologia. 559:401–411.

Elser JJ, Bracken MES, Cleland EE, Gruner DS, Harpole WS, Hillebrand H, Ngai JT, Seabloom EW, Shurin JB, Smith JE. 2007. Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. Ecol Lett. 10:1135–1142.

Filstrup CT, Downing JA. 2017. Relationship of chlorophyll to phosphorus and nitrogen in nutrient-rich lakes. Inland Waters. 7(4):385–400.

Filstrup CT, Wagner T, Soranno PA, Stanley EH, Stow CA, Webster KE, Downing JA. 2014. Regional variability among nonlinear chlorophyll–phosphorus relationships in lakes. Limnol Oceanogr. 59(5):1691–1703.

Friedman JH, Meulman JJ. 2003. Multiple additive regression trees with application in epidemiology. Stat Med. 22:1365–1381.

Friedman JH, Popescu BE. 2008. Predictive learning via rule ensembles. Ann Appl Stat. 2:916–954.

Greene WH. 2012. Econometric analysis, 7th ed. Boston (MA), London (UK): Pearson.

Guildford SJ, Hecky RE. 2000. Total nitrogen, total phosphorus, and nutrient limitation in lakes and oceans: Is there a common relationship? Limnol Oceanogr. 45:1213–1223.

Gunkel G, Casallas J. 2002. Limnology of an equatorial high mountain lake – Lago San Pablo, Ecuador: the significance of deep diurnal mixing for lake productivity. Limnologica. 32:33–43.

Hijmans RJ, Phillips S, Leathwick J, Elith J. 2016. dismo: species distribution modeling. R package version 1.1-1. https://CRAN.R-project.org/package=dismo

Hollister JW, Milstead WB, Kreakie BJ. 2016. Modeling lake trophic state: a random forest approach. Ecosphere. 7(3):e01321.

Huber UM, Bugmann HKM, Reasoner MA. 2005. Global change and mountain regions: an overview of current knowledge. Dordrecht (Netherlands): Springer.

Kalff J. 2002. Limnology – inland water ecosystems. Upper Saddle River (NJ): Prentice Hall.

Kasprzak P, Padisák J, Koschel R, Krienitz L, Gervais F. 2008. Chlorophyll *a* concentration across a trophic gradient of lakes: an estimator of phytoplankton biomass? Limnologica. 38:327–338.

Kraemer BM, Mehner T, Adrian R. 2017. Reconciling the opposing effects of warming on phytoplankton biomass in 188 large lakes. Sci Rep. 7:10762.

Lampa E, Lind L, Lind PM, Bornefalk-Hermansson A. 2014. The identification of complex interactions in epidemiology and toxicology: a simulation study of boosted regression trees. Environ Health. 13:57.

Londe LR, Novo EMLM, Barbosa C, Araujo CAS. 2016. Water residence time affecting phytoplankton blooms: study case in Ibitinga Reservoir (São Paulo, Brazil) using Landsat/TM images. Braz J Biol. 76(3):664–672.

Lu F, Chen Z, Liu W, Shao H. 2016. Modeling chlorophyll-*a* concentrations using an artificial neural network for precisely eco-restoring lake basin. Ecol Eng. 95:422–429.

Long JS, Ervin LH. 2000. Using heteroscedasticity consistent standard errors in the linear regression model. Am Stat. 54(3):217–224.

MacCallum SN, Merchant CJ. 2013. ARC-Lake v2.0. 1991–2011. University of Edinburgh, School of GeoSciences / European Space Agency [accessed on April 2017]. http://hdl.handle.net/10283/88

Magumba D, Maruyama A, Takagaki M, Kato A, Kikuchi M. 2014. Relationships between chlorophyll-*a*, phosphorus and nitrogen as fundamentals for controlling phytoplankton biomass in lakes. Environ Control Biol. 51(4):179–185.

Markovic D, Carrizo SF, Kärcher O, Walz A, David JNW. 2017. Vulnerability of European freshwater catchments to climate change. Glob Chang Biol. 23:3567–3580.

Matzinger A, Jordanoski M, Veljanoska-Sarafiloska E, Sturm M, Müller B, Wüest A. 2006. Is Lake Prespa jeopardizing the ecosystem of ancient Lake Ohrid? Hydrobiologia. 553:89–109.

McCauley E, Downing JA, Watson S. 1989. Sigmoid relationships between nutrients and chlorophyll among lakes. Can J Fish Aquat Sci. 46:1171–1175.

McCulloch WS, Pitts W. 1943. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys. 5:115–133.

McPhearson PT, Wallace OC. 2008. Remote sensing applications to biodiversity conservation [accessed on April 2017]. http://ncep.amnh.org

Peterson DL, Parker VT. 1998. Ecological scale: theory and applications. New York (NY): Columbia Universtiy Press.

Phillips G, Pietiläinen OP, Carvalho L, Solimini A, Solheim AL, Cardoso AC. 2008. Chlorophyll–nutrient relationships of different lake types using a large European dataset. Aquat Ecol. 42:213–226.

Poikane S, Portielje R, van den Berg M, Phillips G, Brucet S, Carvalho L, Mischke U, Ott I, Soszka H, Van Wichelen J. 2014. Defining ecologically relevant water quality targets for lakes in Europe. J Appl Ecol. 51:592–602.

Prairie Y, Duarte CM, Kalff J. 1989. Unifying nutrient–chlorophyll relationships in lakes. Can J Fish Aquat Sci. 46:1176–1182.

Premazzi G, Chiaudani G. 1992. Ecological quality of surface waters. Quality assessment schemes for European community lakes. Ispra (Italy): Commission of the European Communities, Joint Research Centre.

Rapport DJ, Costanza R, McMichael AJ. 1998. Assessing ecosystem health. Trends Ecol Evol. 13:397–402.

Ridgeway G [with contributions from others]. 2017. gbm: Generalized Boosted Regression Models. R package version 2.1.3. https://CRAN.R-project.org/package=gbm

Riffler M, Lieberherr GD, Wunderle S. 2015. Lake surface water temperatures of European Alpine lakes (1989–2013) based on the Advanced Very High Resolution Radiometer (AVHRR) 1 km data set. Earth Syst Sci Data. 7(1):1–17.

Sakamoto M. 1966. Primary production by phytoplankton community in some Japanese lakes and its dependence on lake depth. Arch Hydrobiol. 62:1–28.

Scheffer M. 1998. Ecology of shallow lakes. Dordrecht (Netherlands): Kluwer Academic.

Suleiman A, Tight MR, Quinn AD. 2016. Hybrid neural networks and boosted regression tree models for predicting roadside particulate matter. Environ Model Assess. 21(6):731–750.

Turner W, Spector S, Gardiner N, Fladeland M, Sterling E, Steininger M. 2003. Remote sensing for biodiversity science and conservation. Trends Ecol Evol. 18(6):306–314.

Weithoff G, Lorke A, Walz N. 2000. Effects of water-column mixing on bacteria, phytoplankton, and rotifers under different levels of herbivory in a shallow eutrophic lake. Oecologia. 125(1):91–100.

Wood SN. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J R Stat Soc Series B Stat Methodol. 73(1):3–36.