LIMNOLOGY
and
OCEANOGRAPHY: METHODS

Limnol. Oceanogr.: Methods 17, 2019, 320–329

doi: 10.1002/lom3.10316

Modeling vertical gradients in water columns: A parametric autoregressive approach

Diego G. F. Pujoni,[1]* Ludmila S. Brighenti,[1] José F. Bezerra-Neto,[1] Francisco A. R. Barbosa,[1] Renato M. Assunção,[2] Paulina M. Maia-Barbosa[1]

[1]Laboratório de Limnologia, Ecotoxicologia e Ecologia Aquática, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

[2]Departamento de Ciência da Computação, Instituto de Ciências Exatas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

Abstract

Vertical gradients (e.g., water temperature and density) are a measurable expression of the potential energy of a water body, which is a fundamental driver for biogeochemical processes in aquatic ecosystems. Seasonal vertical stratification is typically described by the mixed layer and thermocline depths, and these metrics are often estimated through visual assessment of graphical plots or using numerical methods. The most widely used numerical method estimates the derivative of temperature (or density) along the depth, but it is sensitive both to profile data resolution and presence of nonconforming observations. In this study, we propose a new method of modeling vertical gradients using a four-parameter sigmoidal function, including temporal autocorrelation. The parameters were estimated through Bayesian nonlinear regression with conditional autoregressive errors. The proposed method provides a quantitative and automated way to estimate the mixed layer and thermocline depths even for data profiles with a poor resolution. It also showed good performance against high-frequency measurement data.

A fundamental determinant of the functioning of waterbodies is vertical gradients in environmental conditions, a primary one being water-column gradients in thermal regimes produced by the unequal heating of water by solar radiation (Wetzel 2001). Vertical thermal variability in deeper waterbodies is defined by marked density gradients that usually divide the waterbody into three vertical compartments: epilimnion, a well-mixed upper region high-influenced by changes in wind and air temperature; metalimnion, a region with sharp changes in temperature/density; and hypolimnion, a deep and relatively stable compartment (Monismith and MacIntyre 2010). Chemical gradients are also generated as a consequence of thermal stratification, for instance, density gradients define vertical gradients in dissolved oxygen, often leading to anoxic conditions in the hypolimnion of nutrient-enriched waterbodies (Wetzel 2001). Anoxic conditions near the bottom may be associated with the production of methane that is transported upward by diffusion and ebullition (Bastviken et al. 2004). In turn, the emissions and oxidation of methane and other greenhouse gases across the air–water interface are moderated by turbulence generated due to variable thermal conditions at the water surface (MacIntyre et al. 2010).

Physical and chemical compartmentalization of the water column exerts a strong influence also on the biological processes. Seasonal variations in thermal stratification can influence phytoplankton (Reynolds 1976; Barros et al. 2006) and zooplankton population dynamics (Eckert and Walz 1998; Brandão et al. 2012). The vertical extent of the epilimnion (i.e., mixed layer depth) and the strength of the thermal gradient in the water column also regulate light penetration and the internal loading of nutrients, both affecting plankton growth and primary production (Vincent et al. 1984; O'Brien et al. 2003; Brighenti et al. 2015). Taken together, robust approaches to the modeling of temperature/density gradients remain critical for advancing our understanding of the structure and function of aquatic ecosystems.

Among the many physical indices describing the thermal structure of lakes, reservoirs, and oceans, the thermocline depth and the mixed layer depth have been consistently used in limnological and oceanographic studies (e.g., Hambright et al. 1994; Perez-Fuentetaja et al. 1999; Pedlosky 2006; Cantin et al. 2011; Zhang et al. 2014). Even though these concepts are well established in the literature, a general automated method for the estimation of these indices has only recently been developed (Read et al. 2011). Many studies in the past estimated the thermocline and mixed layer depths through visual analysis of the temperature or density profiles, a methodology that is subjective and not tractable to apply to large datasets (e.g., thousands of profiles). Some recent studies have used numerical algorithms to

*Correspondence: diegopujoni@gmail.com

estimate thermocline depths using the gradient criterion (GC), which defines the thermocline by the highest derivative of water temperature (or density) in relation to depth (e.g., Hambright et al. 1994; Zhang et al. 2014) and estimates the depth of the mixed layer as the depth above the thermocline where the gradient exceeds a prespecified threshold (Read et al. 2011). This approach has the advantage of being less subjective and can be automated for large datasets, using freely available packages such as those in the Lake Analyzer library for the R programming software (Winslow et al. 2018). One disadvantage of this method, however, is that the estimates of the rate of change are made locally without using information from the entire gradient profile. Therefore, these estimates are substantially sensitive to the data resolution of the profile, which is obtained at discrete depth intervals. An improvement to this approach involves adding weights to adjacent measurements (Read et al. 2011). Because no assumptions are made regarding the form of the vertical gradient and the probability distribution of the measured values, this method is considered nonparametric. Other nonparametric or semiparametric methods, such as piecewise linear regression and cubic spline, also have been used to smooth the discrete profile to better model vertical gradients and estimate derivatives (Fiedler 2010).

An additional class of methods makes use of parametric functions (equations with a fixed number of coefficients) to model environmental gradient profiles. Many functions have been used, including the sigmoidal function (King et al. 1997; Alvera-Azcárate et al. 2011), the cumulative function of generalized F distribution (Chan and Matthews 2005), and a modified version of the van Genuchten (1980) function widely used to model hydraulic conductivity in soils (Rimmer et al. 2005). The main advantage of this broad suite of approaches is to use information from the entire environmental profile, thus producing estimates that are not as sensitive to data resolution of the profile. The main drawbacks are that parametric methods impose a more rigid modeling function compared to nonparametric approaches, and they may be computationally more intensive, requiring good nonlinear optimization algorithms like Gauss–Newton. Some constraints to the parameter space may be provided to avoid unrealistic parameter values (Alvera-Azcárate et al. 2011).

An optimal approach would consider not only the whole profile for the estimation of the thermocline and mixed layer depths but also take advantage of the temporal or spatial correlation of neighboring profiles. This is possible due to the nature of water temperature or density variables, which are spatially and temporally explicit. Many statistical approaches have been developed to handle this type of autocorrelated variable, among them the autoregressive models, such as the simultaneous and conditional autoregressive models (SAR and CAR; Cressie 1993). These methods derive their autoregressive part from stationary time series models, using the Markov property to adjacent spatial measurements. This class of autoregressive models has been used in many areas such as econometry, epidemiology, sociology, and ecology, due to its flexibility and robustness (Anselin 1988; Potter et al. 2010; Ver Hoef et al. 2018).

Ver Hoef et al. (2018) strongly advocate the use of these models in ecological studies and list many practical uses such as model selection, estimation of autocorrelation and related connectivity parameters, spatial regression, smoothing, and prediction. In relation to the latter, this method is a powerful tool for handling missing values and making good predictions, because including autocorrelation creates unbiased estimators (Miller et al. 2007; Ver Hoef et al. 2018). The spatial smoothing is analogous to increasing sample size, because information from neighbors is borrowed, thus decreasing the chance to obtain noisy or erroneous estimates (Potter et al. 2010).

The objective of this study is to evaluate the performance of a CAR model in modeling vertical gradients of water column and compare the estimates of thermocline and mixed layer depths with estimates made by the GC approach (i.e., the most commonly used numerical approach). To our knowledge, we present the first attempt to use nonlinear regression with temporal autocorrelation of the parameters for the estimation of thermocline and mixed layer depths. In this article, we make the following contributions:

1. Propose a parametric model for the vertical gradient of water density in which the parameters vary smoothly in time according to a CAR model. We also propose a Bayesian estimation procedure for these parameters.
2. Compare the performance of both methods (i.e., CAR and GC) on profiles that have decreasing resolution in space (i.e., depth) and show that our proposed methodology provides more consistent estimates of thermocline and mixed layer depths even with low numbers of measurements in space (i.e., depth) and time.
3. Assess the performance of both methods using a high-frequency dataset, with low resolution in space (i.e., depth), and show that our proposed methodology is more robust and less variable.

Material and procedures

Study lakes

The field study was conducted in one tropical natural lake—Lake Dom Helvécio (19°46′54″S, 42°35′31″W)—located in the Atlantic rainforest of southeast Brazil. The lake belongs to a unique lake system, the Middle Rio Doce, with more than 300 waterbodies, which have been monitored by the International Long-Term Ecological Research (ILTER) project since 1999. This lake is warm monomictic, with a short circulation period (from May to August) during the dry season. It is classified as oligomesotrophic (Maia-Barbosa et al. 2010) with a perimeter of 37.7 km, an area of 5.3 km^2, a volume of 5.94×10^7 m^3, and a maximum and mean depth of 39.3 and 11.2 m, respectively (Bezerra-Neto and Pinto-Coelho 2008). The second is a temperate lake—Lake Mendota (43°06′N, 89°25′W)—located in an urban area of Madison, Wisconsin, and has been monitored by the

North Temperate Lakes Long-Term Ecological Research (NTL-LTER) program since 1995. It is a dimictic and eutrophic lake with a perimeter of 33.8 km, a surface area of 39.4 km^2, a volume of about 5.0×10^8 m^3, and a maximum and mean depth of 25.3 and 12.7 m, respectively.

High-resolution profile

We performed a downward and immediately thereafter an upward high-resolution temperature profile in Lake Dom Helvécio in each month of 2012 using a Hydrolab® DS 5 (Hach) multiparameter probe. Temperature and depth values were collected automatically at 3 s intervals while the probe was retrieved slowly through the water column (the average number of measurements were 136 per profile, with a mean distance of 11 cm between adjacent points). Because the measurements were taken at distinct depths in each month, the profiles were smoothed by a local polynomial regression (with span parameter equal to 0.5 and fitted surface computed exactly), and the fitted values of a 10 cm equidistant profile were extracted. The downward and the upward profiles of the same month were examined graphically to look for possible inconsistencies, and then they were averaged, producing one unique profile per month. From the high-resolution 10-cm profile, we created six downscaled (i.e., downsampled) profiles by systematically sampling the high-resolution profile with distances between samples equal to 20, 50, 100, 150, 200, and 400 cm. The depths of the thermocline and the mixed layer in each month were estimated using the GC implemented in Lake Analyzer software following Read et al. (2011), and CAR logistic model estimated through a Bayesian nonlinear estimation using software WinBUGS 1.4.3 (*see* below). For the estimates, measurements of the first meter of the profile were discarded due to the presence of secondary daily thermoclines very common in lakes of this system (Barbosa and Padisák 2002). Furthermore, the measurements below 20 m were also discarded because of the lack of variability in temperature throughout the year. This study was conducted only in Lake Dom Helvécio.

High-frequency dataset

The high-frequency dataset for Lake Mendota was obtained from the NTL-LTER site (https://lter.limnology.wisc.edu/). Specifically, we used the "North Temperate Lakes LTER: High-Frequency Water Temperature Data - Lake Mendota Buoy 2006 - current" (dataset ID: 130 [Magnuson et al. 2012]). For Lake Dom Helvécio, data were collected by a thermistor chain of 12 sensors installed at the deepest point at depths 0.5, 2, 3.5, 4.5, 5.5, 6.5, 7.5, 9, 10, 12, 15, and 21 m. These sensors recorded water temperature, with an accuracy of $\pm$ 0.2°C, every 15 min between May 2011 and December 2012. The raw data collected were validated, excluding unrealistic measurements and corrected for sensor drift by comparing to monthly data from the high-resolution profile and by field calibrations. For the estimation, we restricted our analysis within the stratified period (between 01 November 2011 until 31 May 2012) and used a time resolution of 2 h, which was systematically sampled from the high-frequency dataset. We also did not use the subsurface (0.5 m) measurements, in order to exclude secondary thermoclines, as described above.

Meteorological variables and water column stability

Meteorological data for lake Mendota were obtained from the Wisconsin State Climatology Office, Madison, Wisconsin (https://www.aos.wisc.edu/~sco). We used the daily values of mean air temperature and accumulated rainfall from three stations located in the Mendota's surroundings, with values averaged to yield one time series. Wind speed data were obtained from the NTL-LTER site (https://lter.limnology.wisc.edu/). Specifically, we used the "North Temperate Lakes LTER: High Frequency Data: Meteorological, Dissolved Oxygen, Chlorophyll, Phycocyanin - Lake Mendota Buoy 2006 - current" (dataset ID: 129 [Magnuson et al. 2012]). Meteorological data for Lake Dom Helvécio were obtained from the Instituto Nacional de Meteorologia (http// www.inmet.gov.br) and from Sistema Brasileiro de Coleta de Dados (http://sinda.crn2.inpe.br/PCD/). We used daily values of mean air temperature and accumulated rainfall obtained from three weather stations located within a radius of 10 km from the center of the lake. Wind speed was measured at 1 m above the lake surface from a station (Global Water® WE550) located in the central deep region of the lake.

The water column physical stability was assessed by the Schmidt stability (Idso 1973) and lake number (Imberger and Patterson 1989). These two indexes were estimated using the time series functions (ts.schmidt.stability and ts.lake.number, respectively) from the R package rLakeAnalyzer (Winslow et al. 2018). Details of index equations and calculations can be found in Read et al. (2011). Hypsographic data necessary to calculate the stability indexes were obtained from Brock (1985) and Bezerra-Neto and Pinto-Coelho (2008).

Lake Analyzer software

The thermocline and mixed layer depths were estimated using the function "thermo.depth" with S_{min} parameter equal to 0.1 and the function "meta.depths" with *slope* parameter equal to 0.05. We tested two S_{min} values equal to 0.1 and 0.05, and the difference was negligible in the estimation of the thermocline depth, so we choose to use the function default value which is 0.1 (results not shown). The same two values were tested for the *slope* parameter, and we found a large difference in the estimated mixed layer depths. We choose the slope value equal to 0.05 by visual inspection. It is important to note that water temperature values are internally converted to water density using the function "water.density" before mixed layer and thermocline depths are estimated. Thus, before applying the CAR method (described below), water temperature measurements were also converted to density using the same function. As we are working with densities, it would be better to use the term pycnocline instead of thermocline, however, we decided to keep the latter term because it is more widely used.

CAR logistic model

We assumed that the density D at depth z in time t has a normal distribution.

$$(D \mid z, t) \underset{iid}{\sim} N\left(\mu_{z,t}, \sigma^2\right) \tag{1}$$

with constant σ^2 and

$$\mu_{z,t} = D_{\text{bottom}_t} + \frac{D_{\text{top}_t} - D_{\text{bottom}_t}}{1 + e^{[\beta_t(z - \tau_t)]}} = F(z, t) \tag{2}$$

where D_{bottom_t} and D_{top_t} are, respectively, the densities of the bottom and the top of the thermistor chain used in the analysis at time t, τ_t is the depth of thermocline at time t, and β_t is proportional to the derivative of density change in relation to depth at the thermocline depth (Fig. 1).

We included a temporal autocorrelation for each one of the four parameters (P) with lag equal to 1 and −1. This means that the parameters had a unique mean (α) with a random error correlated with estimates of its two adjacent neighbors (ε_i).

$$P = \alpha + \varepsilon_i \tag{3}$$

$$f(\varepsilon_1, \ldots, \varepsilon_N \mid \lambda) \propto \exp\left(-\frac{\lambda}{2}\sum_{i-j}\left(\varepsilon_i - \varepsilon_j\right)^2\right) \tag{4}$$

where i–j are two adjacent periods (for the high-resolution profile, i–j are two adjacent months and for the high-frequency dataset, i–j are two adjacent measures taken at a 2-h frequency), and N is the number of periods used. We used flat improper priors for the mean of each of the four parameters in the logistic regression

$$f(\alpha_k) \propto 1,\ k = 1 \ldots 4 \tag{5}$$

The variance was modeled as the reciprocal of the precision and we used a gamma (0.01; 0.01) distribution. The same gamma distribution was used as a priori distribution of hyperparameters (λ) in Eq. 4. Estimation was made using software WinBUGS 1.4.3 together with the R package "rbugs" (Yan and Prates 2013). We simulated 10,000 samples and neglected the first 5000.

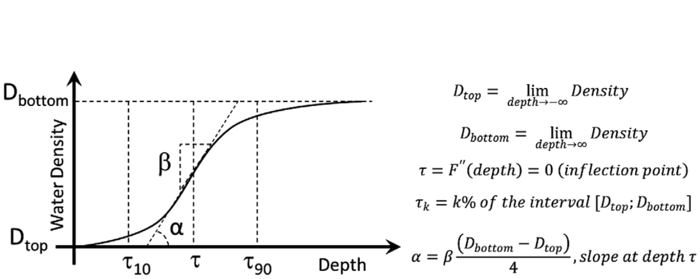

Fig. 1. The four-parameter logistic function to describe the interpretation of its parameters.

The thermocline depth is estimated by the τ parameter in the formula, and the mixed layer depth can be estimated through the formula $\tau - \ln(p)/\beta$. The value p is chosen to make this parameter represent the depth at which the density increases by $1/(p + 1)$% of its top value. It is arbitrary and can be changed by the user, similarly to the slope parameter in "thermo.depth" function. For the high-resolution profile, we used a $p = 199$, which corresponds to a density decrease of 0.5%. For the high-frequency dataset, we used $p = 9$, which corresponds a density decrease of 10%.

Comparing model performance

We visually compared the thermocline and the mixed layer depths estimated by both methods from the 10 cm interval profiles by plotting the estimated values against the temperature profiles (Fig. 2). We also used a linear regression analysis to check the concordance between both methods. After assessing the concordance between methods in this ideal scenario, we quantified the performance of each method separately to downsampling through the root-mean-squared deviation (RMSD) between the respective value estimated using the 10 cm interval profiles and that using the downsampled profile.

$$\text{RMSD}_{ij} = \sqrt{\frac{\sum_{t_{ij}=1}^{n_{ij}}\left(d_{t_{ij}} - \delta_{t_j}\right)^2}{n_{ij}}} \tag{6}$$

where $d_{t_{ij}}$ is the estimation of mixed layer or thermocline depth in month t based on the downsampled profile i (i = 20, 50, 100, 150, 200, and 400 cm) for method j (j = GC or CAR), and $\delta_{t_{ij}}$ is the mixed layer or thermocline depth of month t based on the ideal scenario (i.e., 10 cm interval profile) for each method j. We excluded from the analysis the estimates for the months June, July, and August, as the gradient profiles were very small in these months, and the lake is considered completely mixed.

Assessment

Comparing the estimation with the high-resolution profile (10 cm)

According to our expectations, both methods demonstrated strong concordance for estimating both mixed layer and thermocline depths in the ideal scenario with the high-resolution profiles (i.e., distance between adjacent measurements equal to 10 cm; Fig. 2). The largest differences between methods were observed for the mixed layer depth in December (0.97 m) and for thermocline depth in January (1.6 m), with a mean difference equal to 0.55 m. In September, the CAR method estimated a negative value for the mixed layer depth that was converted to zero. The intercept and the slope between estimates from both methods did not differ from the 1 : 1 line (intercept equal to zero

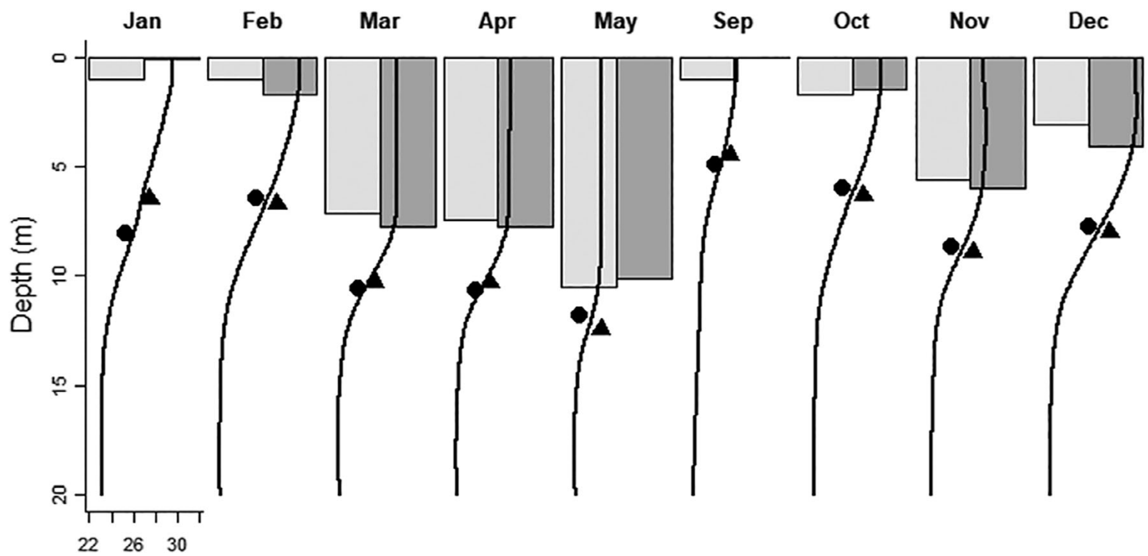

Fig. 2. Comparison between the mixed layer and thermocline depths estimated by two distinct methods in Lake Dom Helvécio. The gray bar represents the mixed layer. For each month, the light gray bar is the mixed layer estimated by the GC method, and the dark gray bar is the mixed layer estimated by the CAR method. The circle and the triangle symbols represent the estimated thermocline depths according to the GC and the CAR methods, respectively. The estimates for months June, July, and August are not shown, because the gradient profiles were very small in these months, and the lake was considered completely mixed.

and slope equal to one), indicating a significant agreement (Table 1; Figs. 3–4).

Performance to decreasing resolution of profile

The RMSD values increased for both methods when comparing very refined profiles to coarser profiles (Figs. 5–6). The CAR method had the lowest RMSD compared to the GC for both estimates (i.e., mixed layer and thermocline depths) in every comparison, except for the mixed layer depth with a 20 cm interval profile (Fig. 5). These results show that the GC method had a higher sensitivity to downsampling of water profile data and indicates that the CAR method is more efficient and may be preferred when data from a smaller number of thermistors are available. Our findings are supported by a previous study which revealed similar results when comparing the thermocline estimation using the GC method with the estimation of thermocline by fitting individual temperature profiles to sigmoid curves (Alvera-Azcárate et al. 2011).

Performance of methods on a high-frequency dataset

The performance of both methods on a high-frequency dataset was assessed qualitatively. The GC method showed a highly erratic oscillation in the estimated mixed layer and thermocline depths, because this method is highly sensitive to small perturbations in the profile (Figs. 7A, 8A). As an example, we analyzed the estimates made by both methods for Lake Dom Helvécio on 29 December 2011, for hours 2:00 and 4:00 h. The maximum temperature difference between these two profiles was 0.56°C (i.e., they are very similar profiles), but the thermocline and mixed layer depths estimated by GC method for both periods were 4.4 and 3.13 m apart, respectively. By contrast, the CAR method

Table 1. Estimates for the slope and intercept between the mixed layer and thermocline depths (m) estimated by both methods (i.e., GC and CAR) for the high-resolution profile dataset in Lake Dom Helvécio. In parenthesis is the 95% confidence interval.

	Mixed layer depth (m)	Thermocline depth (m)
Intercept	−0.13 (−1.09; 0.82)	−0.35 (−2.40; 1.79)
Slope	1.04 (0.87; 1.22)	1.03 (0.78; 1.28)

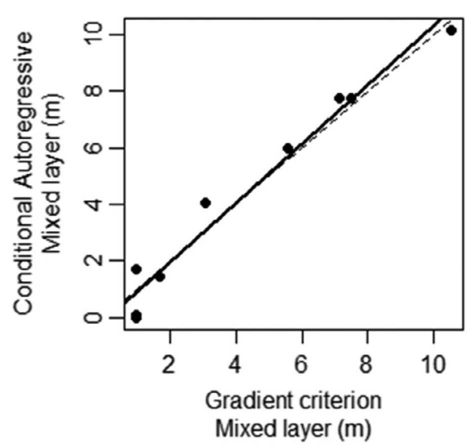

Fig. 3. Estimates of the mixed layer depth according to the GC and CAR methods in Lake Dom Helvécio. The dashed line shows the 1 : 1 diagonal, and the solid line is the regression fit.

produced estimates that were less than 0.04 m apart. The inclusion of temporal autocorrelation of the parameters avoided abrupt variations in the estimates when small oscillations of water

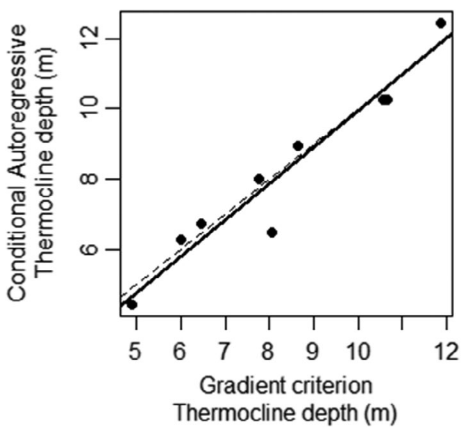

Fig. 4. Estimates of thermocline depth according to the GC and CAR methods in Lake Dom Helvécio. The dashed line shows the 1 : 1 diagonal, and the solid line is the regression fit.

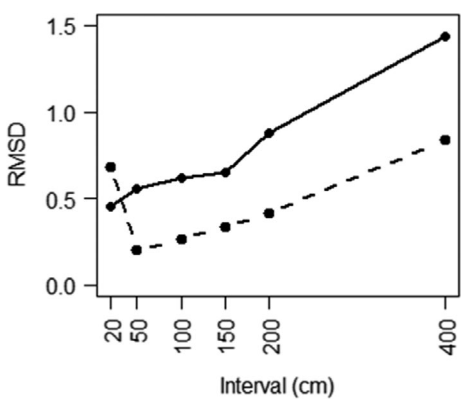

Fig. 5. RMSD for each profile interval (20, 50, 100, 150, 200, and 400 cm) for the mixed layer depth estimation in Lake Dom Helvécio. The solid line is the GC method, and the dashed line is the CAR method.

temperature were present (Figs. 7B, 8B). This is also a desirable feature when outliers are present. Chan and Matthews (2005) discussed the lack of convergence problem in nonlinear fitting due to the presence of nonconforming observations and proposed switching the nonconforming point by a local interpolation. Using the CAR method, this adjustment is not necessary because the nonconforming point would have a small effect on the estimation of the mixed layer and thermocline depths due to the fact that the neighboring profiles help achieve appropriate model convergence.

Although the estimates made by CAR method are robust to small measurement oscillations (i.e., noise), they are still capable of detecting "real" abrupt changes, such as extreme wind-driven mixing events. This can be observed very clearly at the end of September for Lake Mendota, where the wind velocity reached nearly 12 m s^{-1}, inducing water column mixing and deepening the thermocline more than seven meters in just 18 h (Fig. 8). Other wind-driven mixing events could also be

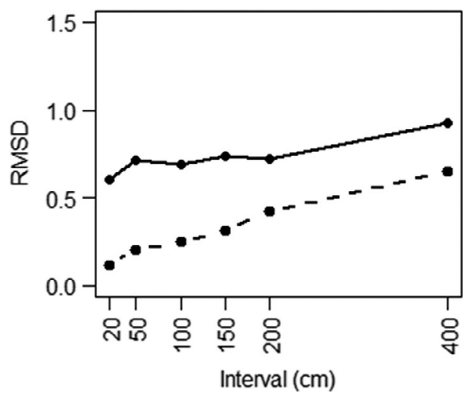

Fig. 6. RMSD for each profile interval (20, 50, 100, 150, 200, and 400 cm) for the thermocline depth estimation in Lake Dom Helvécio. The solid line is the GC method, and the dashed line is the CAR method.

seen at the end of August. For Lake Dom Helvécio, the period between November 2011 and January 2012 has the majority of storms that strongly influence the heat balance of the lake, bringing cold water from high rainfall. This unstable period was correctly captured by both methods, showing an oscillation in both mixed layer and thermocline depths (Fig. 7).

For the CAR method, there were negative estimates of the mixed layer at the end of February and June 2012, when the thermocline was shallower. This drawback could be easily resolved by transforming the negative values to zero. A similar approach is used by the Lake Analyzer software, which does not estimate the mixed layer when the difference between temperature measurements is below a pre-established threshold.

Correlation between water column stability and the beta parameter

The β parameter is related to the steepness of the metalimnion profile, which is related to water column stability. Another measure would be the slope at depth τ, which is related to the β parameter (Fig. 1). In order to check this, we calculated the Spearman correlation between β and slope parameters and two indices for measuring water column stability: Schmidt stability and lake number. The highest correlation found was between slope at thermocline depth and Schmidt stability for Lake Dom Helvécio (Table 2). For Lake Mendota, all the correlations were low (Table 3). Some contrasting results can be seen, such as the negative correlation between slope and lake number for Lake Mendota and a positive correlation for the same pair of variables for Lake Dom Helvécio. This result shows us that the β and the slope parameters may not be a good measurement for lake stability. A similar conclusion was found for oceanic profiles by Fiedler (2010), who concluded that the slope of the temperature-depth profile within the thermocline does not give a good measure of the degree of stratification, interpreted as water column stability.

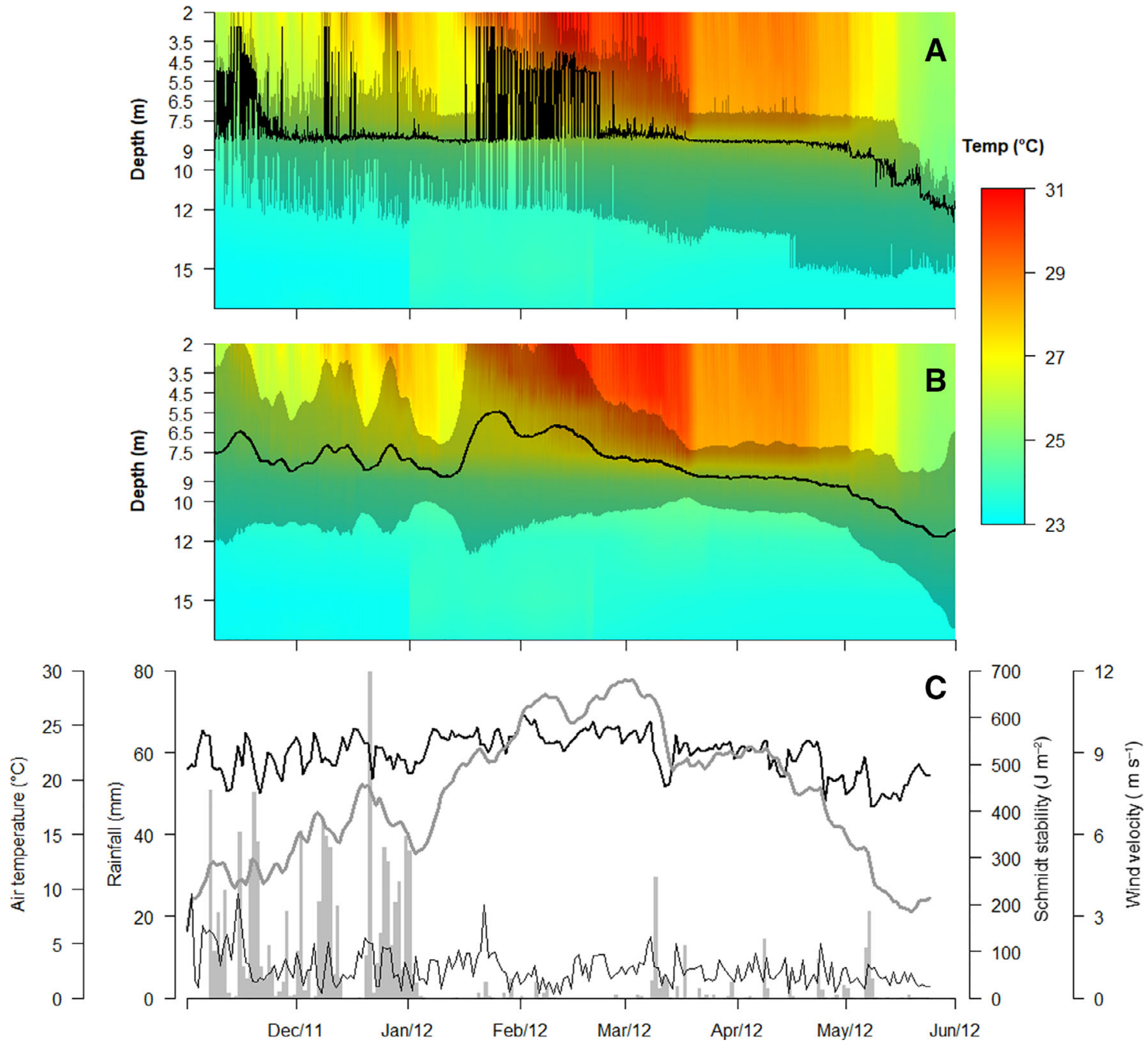

Fig. 7. Estimations of thermocline depths (black line) and metalimnion (shaded region) for Lake Dom Helvécio by both methods, (**A**) the GC method and (**B**) the CAR method. The thermal structure of the lake is plotted using the high-frequency data with a 2-h frequency. The depths where the thermistors were installed correspond to the numerical values marked in the *y*-axis. Measurements above 2 m and below 17 m are not shown. (**C**) Air temperature (thick black line), Schmidt stability (thick gray line), wind velocity (thin black line), and rainfall (gray bars) for the period between January 2011 and June 2012.

Discussion

Our study described a new approach to estimate the mixed layer and thermocline depths and compared the performance of this method with the commonly used GC approach. We found that the GC method, which estimates the parameters for each profile individually and locally calculates the rate of change of density relative to depth, is particularly sensitive to the resolution of the profile, and to the presence of small oscillations in water temperature measurements. We can overcome this limitation by using a statistical technique that considers both the temporal autocorrelation of adjacent profiles and the information from the entire profile to estimate parameters. In this sense, the use of our methodology has the same result as increasing sample size as the estimates have lower variability. For budget-limited projects where the number of probes or thermistors available is scarce, this method presents a very useful and desirable feature.

The CAR function is a Bayesian hierarchical model primarily used to model spatial processes (Besag 1974). However, this method is also useful to model any autocorrelated structures like time series, because it has the same Markov properties of temporal autocorrelation models (Cressie 1993). Some studies with vertical profiles data are spatially explicit (Fiedler 2010; Alvera-Azcárate et al. 2011) and they could take advantage of this spatial autocorrelation to make more accurate and precise estimates.

A potential limitation of CAR models is that they are computer intensive, and the estimation of very large datasets requires computers with high processing capability. Using an Intel i5 2.3 GHz processor with 8 GB of RAM, it took 23 min to calculate the estimates for 12 sensors collecting data every 2 h for 7 months (2556 time steps). However, new software for performing Bayesian inference is being developed which will include parallel computations (e.g., MultiBUGS), and we expect that the size of the dataset will no longer be an issue in the very near future.

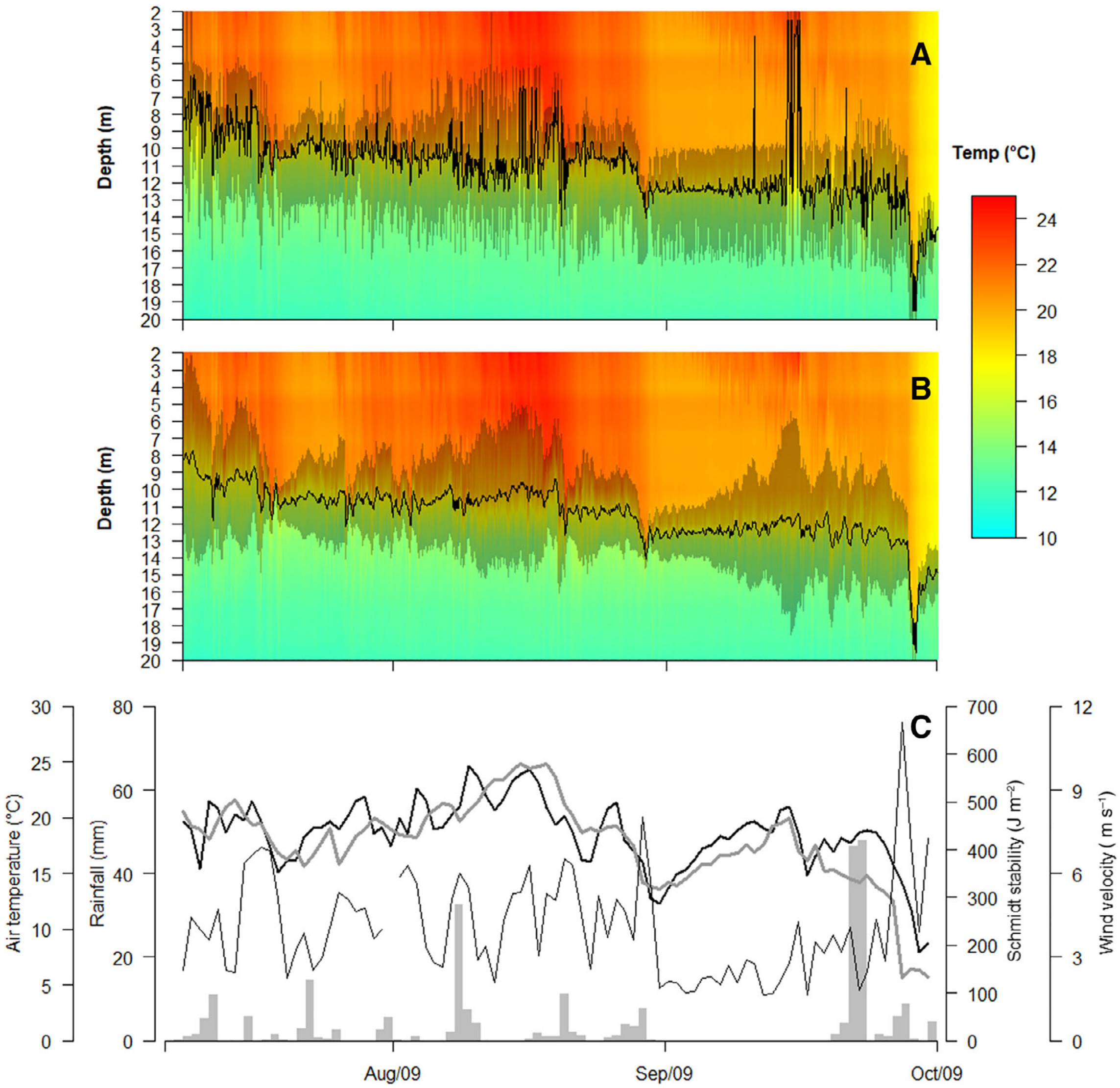

Fig. 8. Estimations of thermocline depths (black line) and metalimnion (shaded region) for Lake Mendota by both methods, (**A**) the GC method and (**B**) the CAR method. The thermal structure of the lake is plotted using the high-frequency data with an hourly frequency. The depths where the thermistors were installed correspond to the numerical values marked in the *y*-axis. Measurements above 2 m are not shown. (**C**) Air temperature (thick black line), Schmidt stability (thick gray line), wind velocity (thin black line), and rainfall (gray bars) for the period between July 2009 and October 2009. Note that the temperature scale is not the same as in Fig. 7.

Table 2. Spearman correlation between hourly estimates of stability indices and beta and slope values estimated by CAR method for Lake Dom Helvécio. Statistically significant correlation coefficients ($p < 0.05$) are in bold.

	Schmidt stability	Lake number
Lake number	**0.51**	
Beta CAR	**0.21**	−0.01
Slope CAR	**0.79**	**0.34**

Table 3. Spearman correlation between hourly estimates of stability indices and beta and slope values estimated by CAR method for Lake Mendota. All correlation coefficients are significant ($p < 0.05$).

	Schmidt stability	Lake number
Lake number	0.17	
Beta CAR	0.08	−0.32
Slope CAR	0.45	−0.36

Comments and recommendations

In this study, we used temperature/density measurements to study stratification. However, the same methodology could also be applied with any other variable that has a sigmoidal-like profile shape. For instance, the dissolved oxygen profile is often noisy, requiring some filtering in order to separate the signal from the noise, and our methodology could be a possible candidate. In addition, although we used lakes as our study system to demonstrate the efficiency of our method, it can readily be applied to profiles both in lakes and oceans.

It is well established that the vertical temperature profile in lakes and oceans is a result of interacting processes that include solar radiation, wind, currents, seiches, and so forth. Consequently, the real-vertical profile has a very complex microstructure, making it difficult to model using only a simple sigmoid function. The aim of our study was to demonstrate a new methodology to improve estimates of the thermocline and mixed layer depths, and we believe that this is the first step toward the improvement of vertical profile modeling strategies. For instance, in relation to secondary thermoclines, the GC implemented in Lake Analyzer package is capable of estimating not only the parent or seasonal thermocline but also the secondary thermocline. The four-parameter logistic function can model only the parent thermocline, but this function can be extended, including more parameters, to model both thermoclines. One candidate is the seven-parameter double-sigmoid function (Lipovetsky 2010). Extending this model to estimate also the secondary thermocline would increase computational time but would be very useful to study diel stratification patterns like atelomixis (Barbosa and Padisák 2002) and lake metabolism.

We demonstrated that by explicitly including temporal autocorrelated data from neighboring profiles, we could substantially improve estimates of the mixed layer and thermocline depths. These improvements are especially apparent for data profiles with a very poor resolution and irregular sampling schemes with the presence of missing values. The model fitting through Bayesian nonlinear regression with CAR errors (CAR model) was robust to noisy data and resulted in more precise estimates compared to the GC method. Read et al. (2011) advocate that the methodology implemented in Lake Analyzer software should be improved, promoting a more objective and efficient methodology that could be used by increasing the number of studies with high-frequency measurements. Although we do not imply that our method is necessarily more correct than the GC, it does allow more information to be leveraged in the estimation of vertical stratification characteristics.

References

Alvera-Azcárate, A., A. Barth, R. H. Weisberg, J. J. Castañeda, L. Vandenbulcke, and J. M. Beckers. 2011. Thermocline characterization in the Cariaco basin: A modelling study of the thermocline annual variation and its relation with winds and chlorophyll-a concentration. Cont. Shelf Res. **31**: 73–84. doi:10.1016/j.csr.2010.11.006

Anselin, L. 1988. Spatial econometrics: Methods and models. Kluwer Academic Publishers. doi:10.1007/978-94-015-7799-1

Barbosa, F. A. R., and J. Padisák. 2002. The forgotten lake stratification pattern: Atelomixis, and its ecological importance. SIL Proc. **28**: 1385–1395.

Barros, C., M. B. G. Souza, and F. A. R. Barbosa. 2006. Seasonal mechanisms driving phytoplankton size structure in a tropical deep lake (Dom Helvécio Lake, South-East Brazil). Acta Limnol. Bras. **18**: 55–66.

Bastviken, D., J. J. Cole, M. Pace, and L. Tranvik. 2004. Methane emissions from lakes: Dependence of lake characteristics, two regional assessments, and a global estimate. Glob. Biogeochem. Cycle **18**: GB4009. doi:10.1029/2004GB002238

Besag, J. 1974. Spatial interaction and the statistical analysis of lattice systems. J. R. Stat. Soc. B. **36**: 192–236. Stable URL: https://www.jstor.org/stable/2984812

Bezerra-Neto, J. F., and R. M. Pinto-Coelho. 2008. Morphometric study of Lake Dom Helvécio, Parque Estadual do Rio Doce (PERD), Minas Gerais, Brazil: A re-evaluation. Acta Limnol. Bras. **20**: 161–167.

Brandão, L., T. Fajardo, E. Eskinazi-Sant'Anna, S. L. Brito, and P. M. Maia-Barbosa. 2012. Fluctuations of the population of *Daphnia laevis* Birge 1878: A six-year study in a tropical lake. Braz. J. Biol. **72**: 479–487. doi:10.1590/S1519-69842012000300010

Brighenti, L. S., P. A. Staehr, L. M. Gagliardi, L. P. M. Brandão, E. C. Elias, N. A. S. T. Mello, F. A. R. Barbosa, and J. F. Bezerra-Neto. 2015. Seasonal changes in metabolic rates of two Tropical Lakes in the Atlantic Forest of Brazil. Ecosystems **18**: 589–604. doi:10.1007/s10021-015-9851-3

Brock, T. D. 1985. A eutrophic lake: Lake Mendota, Wisconsin. Ecol. Stud. **55**: 43–83.

Cantin, A., B. E. Beisner, J. M. Gunn, Y. T. Prairie, and J. G. Winter. 2011. Effects of thermocline deepening on lake plankton communities. Can. J. Fish. Aquat. Sci. **68**: 260–276. doi:10.1139/F10-138

Chan, V., and R. A. Matthews. 2005. Using the generalized F distribution to model limnetic temperature profile and estimate thermocline depth. Ecol. Model. **188**: 374–385. doi:10.1016/j.ecolmodel.2005.04.018

Cressie, N. 1993. Statistics for spatial data. Wiley Series in Probability and Mathematical Statistics. John Wiley.

Eckert, B., and N. Walz. 1998. Zooplankton succession and thermal stratification in the polymictic shallow Müggelsee (Berlin, Germany): A case for the intermediate disturbance hypothesis? Hydrobiologia **387**: 199–206. doi:10.1023/A:1017045927016

Fiedler, P. C. 2010. Comparison of objective descriptions of the thermocline. Limnol. Oceanogr.: Methods **8**: 313–325. doi:10.4319/lom.2010.8.313

Hambright, K. D., M. Gophen, and S. Serruya. 1994. Influence of long-term climatic changes on the stratification of a subtropical, warm monomictic lake. Limnol. Oceanogr. **39**: 1233–1242. doi:10.4319/lo.1994.39.5.1233

Idso, S. B. 1973. On the concept of lake stability. Limnol. Oceanogr. **18**: 681–683. doi:10.4319/lo.1973.18.4.0681

Imberger, J., and J. C. Patterson. 1989. Physical Limnology, p. 303–475. *In* Advances in applied mechanics. Elsevier. doi:10.1016/S0065-2156(08)70199-6

King, J. R., B. J. Shuter, and A. P. Zimmerman. 1997. The response of the thermal stratification of South Bay (Lake Huron) to climatic variability. Can. J. Fish. Aquat. Sci. **54**: 1873–1882. doi:10.1139/f97-093

Lipovetsky, S. 2010. Double logistic curve in regression modeling. J. Appl. Stat. **37**: 1785–1793. doi:10.1080/02664760903093633

MacIntyre, S., A. Jonsson, M. Jansson, J. Aberg, D. E. Turney, and S. D. Miller. 2010. Buoyancy flux, turbulence, and the gas transfer coefficient in a stratified lake. Geophys. Res. Lett. **37**: L24604. doi:10.1029/2010GL044164

Magnuson, J, J., S. R. Carpenter, and E. H. Stanley. 2012. North temperate lakes long term ecological research program. NSF, Center for Limnology, University of Wisconsin-Madison. Available from http://lter.limnology.wisc.edu. Accessed November 3, 2018

Maia-Barbosa, P. M., L. G. Barbosa, S. L. Brito, F. Garcia, C. F. A. Barros, M. B. G. Souza, N. A. S. T. Mello, A. . S. Guimarães, and F. A. R. Barbosa. 2010. Limnological changes in Dom Helvécio Lake (South-East Brazil): Natural and anthropogenic causes. Braz. J. Biol. **70**: 795–802. doi:10.1590/S1519-69842010000400010

Miller, J., J. Franklin, and R. Aspinall. 2007. Incorporating spatial dependence in predictive vegetation models. Ecol. Model. **202**: 225–242. doi:10.1016/j.ecolmodel.2006.12.012

Monismith, S. G., and S. MacIntyre. 2010. The surface mixed layer in lakes and reservoirs, p. 207–221. *In* Biogeochemistry of Inland Waters. Elsevier, San Diego.

O'Brien, K. R., G. N. Ivey, D. P. Hamilton, A. M. Waite, and P. M. Visser. 2003. Simple mixing criteria for the growth of negatively buoyant phytoplankton. Limnol. Oceanogr. **48**: 1326–1337. doi:10.4319/lo.2003.48.3.1326

Pedlosky, J. 2006. A history of thermocline theory, p. 139–152. *In* M. Jochum and R. Murtugudde [eds.], Physical oceanography: Developments since 1950. Springer. doi:10.1007/0-387-33152-2_9

Perez-Fuentetaja, A., P. J. Dillon, N. D. Yan, and D. J. McQueen. 1999. Significance of dissolved organic carbon in the prediction of thermocline depth in small Canadian shield lakes. Aquat. Ecol. **33**: 127–133. doi:10.1023/A:1009998118504

Potter, J. E., C. P. Schmertmann, R. M. Assunção, and S. M. Cavenaghi. 2010. Mapping the timing, pace, and scale of the fertility transition in Brazil. Popul. Dev. Rev. **36**: 283–307. doi:10.1111/j.1728-4457.2010.00330.x

Read, J. S., D. P. Hamilton, I. D. Jones, K. Muraoka, L. Winslow, R. Kroiss, C. H. Wu, and E. Gaiser. 2011. Derivation of lake mixing and stratification indices from high-resolution lake buoy data. Environ. Model. Softw. **26**: 1325–1336. doi:10.1016/j.envsoft.2011.05.006

Reynolds, C. S. 1976. Succession and vertical distribution of phytoplankton in response to thermal stratification in a lowland mere, with special reference to nutrient availability. J. Ecol. **64**: 529–551. doi:10.2307/2258772

Rimmer, A., Y. Aota, M. Kumagai, and W. Eckert. 2005. Chemical stratification in thermally stratified lakes: A chloride mass balance model. Limnol. Oceanogr. **55**: 147–157. doi:10.4319/lo.2005.50.1.0147

Van Genuchten, M. T. 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci. Soc. Am. J. **44**: 892–898. doi:10.2136/sssaj1980.03615995004400050002x

Ver Hoef, J. M., E. E. Peterson, M. B. Hooten, E. M. Hanks, and M. J. Fortin. 2018. Spatial autoregressive models for statistical inference from ecological data. Ecol. Monogr. **88**: 36–59. doi:10.1002/ecm.1283

Vincent, W., P. Neale, and P. Richerson. 1984. Photoinhibition: Algal responses to bright light during diel stratification and mixing in a tropical alpine lake. J. Phycol. **20**: 201–211. doi:10.1111/j.0022-3646.1984.00201.x

Wetzel, R. G. 2001. Limnology: Lake and river ecosystems, 3rd Edition. Academic Press.

Winslow L., J. Read, R. Woolway, J. Brentrup, T. Leach, and J. Zwart. 2018. rLakeAnalyzer: Lake Physics Tools. R package version 1.11.4. URL: https://cran.r-project.org/web/packages/rLakeAnalyzer/rLakeAnalyzer.pdf. Accessed March 29, 2019

Yan J., and M. Prates. 2013. rbugs: Fusing R and OpenBugs and beyond. R package version 0.5-9. Available from https://CRAN.R-project.org/package=rbugs

Zhang, Y., Z. Wu, M. Liu, J. He, K. Shi, M. Wang, and Z. Yu. 2014. Thermal structure and response to long-term climatic changes in Lake Qiandaohu, a deep subtropical reservoir in China. Limnol. Oceanogr. **59**: 1193–1202. doi:10.4319/lo.2014.59.4.1193

Acknowledgments

This study received financial support from the Conselho Nacional de Desenvolvimento Científico e Tecnológico/Ministério da Ciência e Tecnologia (MCT/CNPq), as part of the funding of the Programa de Pesquisas Ecológicas de Longa Duração-PELD (ILTER), Site 4 (grant 403698/2012-0) in addition to support from Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG; CRA-APQ-02623-10) and from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES; 88881.030499/2013-2101). We also thank the Forestry Institute of Minas Gerais (IEF-MG) for the sampling permit and all the staff of Parque Estadual do Rio Doce (Rio Doce State Park) for the logistic support. D.G.F.P. and L.S.B. received scholarships from CNPq and CAPES. We also thank the North Temperate Lakes Long-Term Ecological Research program for providing the data. We are grateful to Julian Olden, Marcos Oliveira Prates, and Derick Matheus de Oliveira for his valuable suggestions and support, as well as to the staff of the Laboratório de Limnologia, Ecotoxicologia e Ecologia Aquática (LIMNEA) of the Instituto de Ciências Biológicas of Universidade Federal de Minas Gerais (ICB/UFMG) for their contributions to the sampling and analysis. Finally, we would like to thank two anonymous reviewers for their detailed evaluation of the manuscript and valuable comments.

Conflict of Interest

None declared

Submitted 04 May 2018
Revised 20 February 2019
Accepted 08 March 2019

Associate editor: Craig Lee