Water Resources Research

RESEARCH ARTICLE

Key Points:

- A high frequency, long-duration field study reveals diurnal and seasonal trends in cyanobacteria vertical distribution dynamics
- In a stratified lake environment, magnitude of subsurface cyanobacteria peak concentration is driven primarily by the depth and temperature at the thermocline
- In a stratified lake environment, center of gravity of subsurface cyanobacterial biomass is driven primarily by the width and temperature of the surface mixed layer

Correspondence to:

J. Taylor,
tayl1562@umn.edu

Abiotic Drivers of a Deep Cyanobacteria Layer in a Stratified and Eutrophic Lake

J. Taylor[1,2] , **M. Hondzo**[1,2] , and **V. R. Voller**[1,2]

[1]Twin Cities – Department of Civil, Environmental, and Geo-Engineering, University of Minnesota, Minneapolis, MN, USA, [2]Twin Cities – St. Anthony Falls Laboratory, University of Minnesota, Minneapolis, MN, USA

Abstract Harmful algal blooms (HABs), in particular those consisting of the cyanobacteria *Microcystis*, are becoming increasingly more common across the globe. Despite the growing body of evidence that suggests vertical heterogeneity of *Microcystis* can be a precursor to HAB formation, the abiotic drivers of vertical distribution of *Microcystis* are poorly understood in the field environment. The prediction of subsurface cyanobacteria is also pertinent because subsurface concentrations are not easily recognizable to the public or lake system managers, creating a risk of exposure to harmful algal toxins. High-frequency temporal and vertical data were collected from a research station anchored in a stratified and eutrophic lake for five months. Using a combination of dimensional analysis and machine learning approaches, data show that the magnitude of the subsurface *Microcystis* concentration peak and the center of gravity of the deep cyanobacteria layer are statistically significantly mediated by the thermal structure of the lake. The peak subsurface cyanobacteria biovolume is related to the thermocline depth and temperature, whereas the center of gravity of the subsurface cyanobacteria biovolume is related to the mixed layer depth and temperature. Furthermore, our data suggest there is a seasonal evolution of the subsurface cyanobacteria center of gravity that could potentially indicate timing of HAB onset. Based on easily measured parameters associated with the vertical lake temperature profile and meteorological conditions, we provide evidence of predictable trends in subsurface cyanobacteria variables.

1. Introduction

Harmful algal blooms are one of the most imminent threats to freshwater quality across the globe (Huisman et al., 2018; O'Neil et al., 2012). Of the HAB-forming cyanobacteria species, *Microcystis* are of particular concern due to their ubiquity and their production of Microcystin toxins. There are numerous evolutionary advantages that allow *Microcystis* to thrive across the globe, and one such advantage in stratified lakes is the ability of vertical motility. Cell buoyancy is modulated by adjusting ballast weight through production or metabolism of dense carbohydrates to offset low density intracellular gas vesicles, allowing *Microcystis* to move up or down the water column (Reynolds et al., 1987). The speed of unicellular vertical motility can be greatly enhanced by forming colonies, which is a typical occurrence in natural environments (Xiao et al., 2018).

Traditionally, *Microcystis* vertical migration in natural environments has been hypothesized to be nutrient-driven chemotaxis (Fogg & Walsby, 1971; Ganf & Oliver, 1982). However, as lakes become more eutrophic and nutrients become a less limiting substrate, abiotic factors tend to dominate *Microcystis* vertical motility (Xiao et al., 2018). Most of the work on physical drivers has focused on light, wind, and temperature. Thomas and Walsby (1985, 1986) demonstrated experimentally that *Microcystis* cells will increase in density under high irradiance conditions to sink to a preferred low light intensity, but their ability to regain buoyancy was dependent on water temperature. In laboratory experiments, You et al. (2018) demonstrated that *Microcystis* vertical buoyant velocities were consistently and significantly faster at higher temperatures. Cao et al. (2006) suggested vertical distributions of different phytoplankton species, especially *Microcystis*, were largely correlated with wind events (and had no correlation with nutrients) in a field study of Lake Taihu, China. As a result, it has been suggested the relationship between *Microcystis* vertical motility and timing of HAB onset should be explored more in depth (Liu et al., 2019; Xiao et al., 2018; Zhu et al., 2018). There have been several models formulated to simulate *Microcystis* vertical motility as a function of abiotic factors (Medrano et al., 2013; Wallace et al., 2000; Zhu et al., 2018), and Yao et al. (2017) connected their

simulations of *Microcystis* motility to a hypothesis of necessary conditions for bloom formation, but none have had a long-term, high-frequency, in situ temporal data set for validation.

Aside from being active movers, *Micrycosystis* can also act as passive particles in a water column. Field investigations including Bormans et al. (1999) suggest surface dynamics play the largest role in determining vertical distribution of *Microcystis*. Marti et al. (2016) used relevant time scales of vertical transport in the surface layer and the metalimnion, vertical mixing in the surface layer and the metalimnion, and vertical migration of cyanobacteria to characterize when algae were acting as free movers or passive particles. It was suggested, through field work and 3-D simulations, that the time-scale hierarchy—how fast one process happens compared to another—determines success of particular algal species at specific locations in the lake. A field study by Hozumi et al. (2019) demonstrated that low turbulence levels in the surface layer of Lake Kinneret, Israel led to a thin, dense *Microcystis* scum layer, while higher levels of turbulence led to a thick, sparse *Microcystis* layer. These findings were corroborated in a mesocosm experiment conducted by X. Wu et al. (2019). Although they were concerned with cyanobacterial abundance instead of vertical heterogeneity, Nelson et al. (2018) used sophisticated unsupervised learning techniques on a low-frequency, long-duration data set to relate cyanobacterial abundance to both biotic and abiotic factors, determining that *Microcystis* is highly sensitive to flow conditions. To further complicate the relationship between lake hydrodynamics and cyanobacteria, Sommer et al. (2017) demonstrated that dense layers of motile algae can create a great enough density instability to modulate the mixed layer depth, as corroborated by the field study conducted by Sepúlveda Steiner et al. (2019).

The phenomenon of subsurface peaks in cyanobacterial biomass concentration has been studied extensively in both ocean and lake environments. This feature is often referred to as a deep chlorophyll maximum, or DCM (Cullen, 1982), and the zone in which this increase in biomass occurs is called a deep chlorophyll layer, or DCL (Brooks & Torke, 1977). The abiotic drivers of DCL formation have been extensively studied (Cullen, 2015). Huisman et al. (2006) demonstrated through numerical simulations that there exists a minimum turbulence level in the surface layer of the ocean in order to achieve a stable DCM. Scofield et al. (2017) conducted a field study on Lake Ontario and determined significant dependencies of DCM magnitude and location on the temperature profile of a lake; Sanful et al. (2017) demonstrated euphotic depth was the primary driver of DCM formation and maintenance in another field study. Somavilla et al. (2019) connected deep chlorophyll phenomena to surface blooms in an oceanic environment: field data demonstrated that cyanobacterial biomass was prevalent below the diurnal mixed layer, toward a DCL, when the net heat flux at the water surface became positive on a seasonal timescale. Lastly, despite being a major topic of study for four decades, the actual definition of a DCL or DCM is largely ambiguous. How deep is "deep"? What is the depth of the maximum biomass for profiles with two or more peaks in the DCL? W. Xu et al. (2019) developed a robust machine learning algorithm to help unify these definitions, but even parameters like mixed layer depth, which is used widely in limnological research, show rather dramatic inconsistencies (Gray et al., 2020).

The work highlighted above has brought many interesting insights to the field. However, as is almost always the case when dealing with microorganisms, discrepancies exist between experiments with short-term, high-frequency observations and field data with long-term, low-frequency observations. A wide range of modeling techniques have been utilized, from physically informed time-scale analyses to virtually assumption-free machine learning, based on the available data set. These varying methods and inconsistencies in DCM and DCL definitions and parameters (or lack thereof) make it difficult to draw cohesive conclusions between studies. To bridge these gaps, we anchored a research station from May to August of 2017 in a stratified and eutrophic lake with a history of *Microcystis* blooms. This research station recorded meteorological and water temperature conditions every 5 min and water quality variables every 2 h. Guided by knowledge of relevant physical and biological processes, data were then used to inform and validate a stepwise regression model to determine the relationship between the light and temperature profiles of a lake with the magnitude and shape of a deep cyanobacteria layer (DCL from here on out since we are investigating cyanobacteria specifically instead of chlorophyll in general), a potential precursor to a harmful algal bloom. These results were then compared to those of a random forest analysis, a more hands-off machine learning approach, to verify the intuitive models. Precise definitions and a new shape parameter for the DCL were introduced, as well as a potential method for predicting bloom formation when a DCL is present.
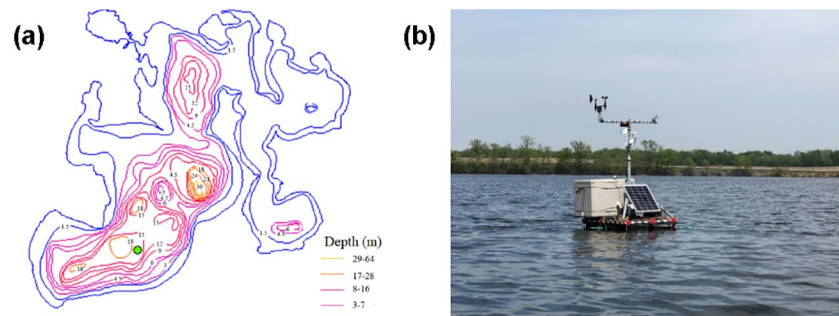
Figure 1. (a) Contour map of South Center Lake (depth contours in meters). Green dot indicates position of research station. (b) Research station.

2. Methods

2.1. Field Site

South Center Lake (45°22′17.8212″N, 92°49′39.18″W), Figure 1a, is a eutrophic and dimictic lake in Chisago County, Minnesota that stratifies from May to October. It has a surface area of approximately 3.3 km^2, a maximum depth of 33 m, an average depth of nearly 5 m, and its shoreline is mostly developed. Due to its status as a Minnesota Pollution Control Agency (MPCA) Sentinel lake, South Center Lake has a wealth of historical data, with the first lake survey occurring in 1942 and regular water quality data from 1997 to present day. This historical data, which indicate a reoccurring summer-time *Microcystis* bloom, were used to supplement measurements from our research station (Engel et al., 2011). For instance, historic phytoplankton assemblages indicate South Center Lake, in a location near where our measurements were taken, is almost entirely cyanobacteria dominant year-round, besides some diatoms and green algae in early summer. MPCA data were also used to corroborate grab sample measurements, detailed in Section 2.3.

A preliminary analysis was conducted that links the seasonal DCL progression to timing of HAB onset (Section 4.2). To corroborate the trend found in South Center Lake, data from a similar lake were used. Ramsey Lake (45°12′27″N, 94°59′43″W) is a deep, dimictic, and eutrophic lake in Wright County, Minnesota with a mostly developed shoreline. It has a surface area of 1.3 km^2, an average depth of 6.4 m, and a maximum depth of 24 m. The research station was placed in Ramsey Lake for May–October 2018. The MPCA has been monitoring Ramsey Lake since 1980, and data show regularly occurring summertime cyanobacteria blooms. Data for Ramsey Lake was obtained, prepared, and analyzed identically to South Center Lake data from the previous year.

2.2. Research Station

The research station was anchored in South Center Lake from May 12, 2017 to October 30, 2017 (Figure 1b). The data used in analysis extend to September 3, 2017—the end of *Microcystis* domination (Wilkinson et al., 2020). The lake depth was 14 m deep at the location of the water station. The research station records meteorological measurements—wind speed, wind direction, air temperature, relative humidity, photosynthetically active radiation (PAR), and rain depth—every 5 min. In addition, a thermistor chain records water temperature at depths of 0.1, 1, 2, 3, 4, 5, 6, 7, 8, 10, 12, and 14 m every 5 min. The water quality profiler takes measurements of dissolved oxygen concentration, pH, phycocyanin concentration, PAR, specific conductivity, and water temperature at depths of 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 6, 7, 8, 9, 10, 11, 12, 13, and 14 m every 2 h. For a full description of equipment used, see Wilkinson et al. (2020). Example profiles for temperature, biovolume, and PAR are shown in Figure A1.

2.3. Data Analysis

Water column parameters that describe the vertical thermal structure of the lake were determined from the Lake Analyzer and Lake Heat Flux Analyzer software in Matlab (Read et al., 2011; Woolway et al., 2015). Weekly grab samples were taken near the research station using a Van Dorn sampler to measure nutrient,

Table 1
Relevant Temperature Profile and Light Parameters

Parameter	Variable	Description
Mixed layer depth	h_{ML}	Depth of base of surface mixed layer (m)
Thermocline depth	h_T	Depth of maximum magnitude temperature gradient (m)
Surface PAR	PAR_{surf}	PAR intensity at the water surface ($\mu mol/s/m^2$)
Mixed layer PAR	PAR_{ML}	Depth-integrated PAR within the mixed layer ($\mu mol/s$)
Thermocline PAR	PAR_T	Depth-integrated PAR above the thermocline ($\mu mol/s$)
Mixed layer temperature	T_{ML}	Water temperature at mixed layer depth (°C)
Thermocline temperature	T_T	Water temperature at thermocline depth (°C)
Thermocline steepness	m_T	Temperature gradient at the thermocline depth (°C/m)

Abbreviation: PAR, photosynthetically active radiation.

phycocyanin, and biovolume concentrations at the same depths as the research station profiler. Following the protocol outlined in Wilkinson et al. (2020), nutrients and phycocyanin measurements were analyzed using a benchtop fluorometer, and phytoplankton species were identified and enumerated using microscopy. The in situ field measurements of phycocyanin, a protein found only in cyanobacteria, were linearly regressed to the laboratory measurements of the corresponding phycocyanin grab samples and the *Microcystis* biovolume estimates. Further details of the laboratory analysis can be found in Wilkinson et al. (2020).

The average mode-1 vertical seiche period of South Center Lake estimated from the Lake Analyzer software was approximately 4 h. Water temperature and research station data were averaged over this seiche period to determine diurnal and seasonal trends. The mixed layer depth, h_{ML}, was defined as the first depth with a temperature difference of at least −0.3°C from the surface water temperature. The thermocline depth, h_T, was defined as the depth at which the maximum temperature gradient occurs. Surface PAR intensity was recorded as PAR_{surf}. Depth-integrated PAR over the mixed layer and above the thermocline were labeled PAR_{ML} and PAR_T, respectively. The euphotic depth, h_{EP}, was defined as the depth at which the PAR intensity decreases to 1% of the surface PAR intensity. Mixed layer temperature, T_{ML}, was then defined as the water temperature at the mixed layer depth, with a corresponding definition for the thermocline temperature, T_T. The thermocline steepness, m_T, was defined as the temperature gradient at the thermocline. Since research station data were recorded at discrete depths (Section 2.2), a piecewise linearly interpolated line was fitted to research station profiler data to create pseudo-continuous profiles. For the convenience of the reader, key parameters derived from measurements are given in Table 1.

2.4. Wind Shear and Natural Convection

Physical processes at the air-water interface mediate the thermal structure of a lake and were thus hypothesized to play a role in cyanobacterial vertical heterogeneity. Turbulence can be introduced into the water column via surface shear induced by wind, or via natural convection when the surface water temperature is warmer than the surrounding air temperature. Two dimensionless parameters were used to measure the role of wind shear and natural convection. The classic Reynolds number was used to measure the contribution of wind shear, defined here as

$$Re = \frac{u_* h_{ML}}{\nu} \tag{1}$$

where ν is the kinematic viscosity of water (m^2/s) and u_* is the shear velocity at the air-water interface (m/s) as calculated by Lake Analyzer. A Reynolds number relevant to natural convection was derived, beginning with the introduction of the penetrative convective velocity:

$$w_* = (B h_{ML})^{\frac{1}{3}} \tag{2}$$

where B is the buoyancy flux in m^2/s^3 (Imberger, 1985). If we assume the air-water temperature difference is the dominant heat flux term, we can define the buoyancy flux at the water surface as

$$
\begin{aligned}
B &= \frac{g\alpha}{\rho C_P} H_Q \\
&\approx \frac{g\alpha}{\rho C_P}\left(k\frac{\partial T}{\partial z}\right)_{z=0} \\
&\approx \frac{g\alpha}{\rho C_P}\left(k\frac{T_{surf} - T_{air}}{\delta_t}\right)
\end{aligned}
\tag{3}
$$

where g is the gravitational acceleration (m/s^2), α is the coefficient of thermal expansion (1/K), ρ is the water density (kg/m^3), CP is the specific heat of water (J/kg/K), HQ is the total heat flux at the water surface (W/m^2) as calculated by Lake Heat Flux Analyzer, k is the thermal conductivity of water (W/m/K), and taking $T_{surf} \approx T_{ML}$ by assuming an infinitely small air-side boundary layer. A positive buoyancy flux, then, indicates the lake is undergoing surface cooling. J. Wu (1971) provides an estimation for δt, the thermal diffusivity layer thickness (m), to be

$$
\delta_t = 5.5\frac{\nu}{u_*}
\tag{4}
$$

We can now introduce a Reynolds number for natural convection at the water surface to be

$$
\begin{aligned}
Re_{NC} &= \frac{w_* h_{ML}}{\nu} \\
&= \frac{(Bh_{ML})^{\frac{1}{3}} h_{ML}}{\nu}
\end{aligned}
$$

Hence,

$$
Re_{NC} = \frac{\left(g\alpha D_T (T_{ML} - T_{surf})\right)^{\frac{1}{3}} h_{ML}}{\nu}
\tag{5}
$$

where $D_T = \dfrac{k h_{ML}}{\rho C_P \delta_t}$ is the thermal dispersion coefficient.

2.5. Deep Cyanobacteria Layer Definitions

A DCL is defined to be present if there exists a biovolume concentration, C, below the diurnal mixed layer depth that is greater than the maximum concentration within the diurnal mixed layer (Figure A1). If this condition is met, then the top of the DCL, z_{TOP}, is defined as the first depth of increasing phycocyanin concentration below the mixed layer. The bottom of the DCL, z_{BOT}, is defined as the first depth after z_{TOP} such that the phycocyanin concentration goes below the average concentration within the mixed layer. The dimensionless center of gravity of the DCL, z_{CG}, can then be defined as

$$
z_{CG} = \frac{1}{h_{max}}\left(\frac{\int_{z_{BOT}}^{z_{TOP}} zC dz}{\int_{z_{BOT}}^{z_{TOP}} C dz}\right)
\tag{6}
$$

where h_{max} is the maximum depth of the water column at the location of the research station ($h_{max} = 14m$). Dividing by h_{max} ensures not only that z_{CG} is dimensionless, but also that it is scaled from 0 to 1, from the water surface to the lake bed, respectively. This center of gravity term is similar to weighted mean depths (or mean residence depths) found in zooplankton literature (Bezerra-Neto & Pinto-Coelho, 2007; Y. Xu et al., 2010), but differs quantitatively and conceptually. Equation 6 gives the depth at which the center of gravity of a DCL is located–it is not the center of gravity of the entire cyanobacteria vertical profile. The variable is meant to describe the DCL only. Conceptually, weighted mean depths are typically used to describe

the active migration of zooplankton to that particular depth. As *Microcystis* is a microorganism capable of vertical motility but also highly subject to environmental conditions, the authors wish to make no assumptions of active or passive motion. Figure A1 demonstrates a typical profile with a DCL and where its center of gravity occurs.

C_{DCL} is a dimensionless variable that describes the relative magnitude of the DCL peak biovolume concentration to the biovolume concentration within the mixed layer and is defined as follows:

$$C_{DCL} = \frac{C_{DCL,max} - C_{ML,avg}}{C_{ML,avg}} \tag{7}$$

where $C_{DCL,max}$ is the maximum biovolume concentration within the DCL and $C_{ML,avg}$ is the average biovolume concentration in the mixed layer. C_{DCL} is bounded below by zero but is unbounded from above.

2.6. Modeling DCL Variables

We can ensure DCL definitions are physically meaningful if we can predict them with relevant forcings. While nitrogen was likely limiting growth the entire observation period, there is no evidence that nutrients were controlling cyanobacterial vertical heterogeneity (Appendix C, Wilkinson et al., 2020). It is hypothesized that the characteristics of the DCL are controlled by the vertical temperature structure and light conditions in the lake. That is, if a DCL is present, its center of gravity and magnitude can be expressed as functions of the variables in Table 1 to arrive at the following:

$$z_{CG} = f(h_{ML}, h_T, PAR_{surf}, PAR_{ML}, PAR_T, T_{ML}, T_T, m_T) \tag{8}$$

$$C_{DCL} = g(h_{ML}, h_T, PAR_{surf}, PAR_{ML}, PAR_T, T_{ML}, T_T, m_T) \tag{9}$$

The intent is to use dimensional analysis to reduce the number of variables in Equations 8 and 9 to physically meaningful dimensionless groups. To aid in our understanding of the dominant parameters in this system, we first utilize some methods from data science. A regression analysis was performed to determine what physical parameters and interactions of physical parameters had the most significant impact on the normalized maximum biovolume concentration in the DCL, C_{DCL}, and the normalized center of gravity of biovolume in the DCL, z_{CG}. C_{DCL}, a dimensionless ratio bounded below by zero with the majority of measurements at low values of C_{DCL} but with events of interest at large values, was log-transformed to normalize the data. Matlab's Statistical and Machine Learning Toolbox and, in particular, the Matlab function stepwiselm was used to generate a model that includes both linear and bi-linear (interaction) terms for C_{DCL} and z_{CG}. Daytime profiles were separated from nighttime profiles to investigate any potential differences between daytime and nighttime dynamics, but no discernible differences were determined. The stepwise regression algorithm assumes independent samples; however, with a high-frequency data set in particular, we expect non-negligible autocorrelation in both independent and dependent variables. While autocorrelated data would artificially inflate the statistical significance of the regression analysis and potentially lead to the erroneous inclusion of insignificant variables to the model, it is important to emphasize the statistics are not the main contribution of this project. Rather, the statistics serve as a guide for a dimensional analysis. The prediction of algae distribution as a function of the light profile in a lake also necessitates another potential caveat: a high abundance of algae causes self-shading, thereby impacting the light profile. Low light conditions could potentially be a consequence, rather than a driver, of high cyanobacteria abundance. These issues are analytically challenging, but the goal of this analysis is not to construct the most perfect model. The objective is to construct a model guided by physical intuition that utilizes a minimal number of easily measured variables. Equations 8 and 9 are designed to reveal how well cyanobacterial vertical distributions can be predicted by exploring temperature and light parameters alone. Once full stepwise regression models had been derived, results were paired with apparent observational trends to inform a simple dimensional analysis, with the goal of creating parsimonious models with the lowest number of dimensionless groups that simultaneously explain the data variability while providing insight to the relevant physical processes.

Because regression analysis results were only used as a starting point before combining with observational trends to guide the dimensional analysis, this method relied on using current knowledge of both physical
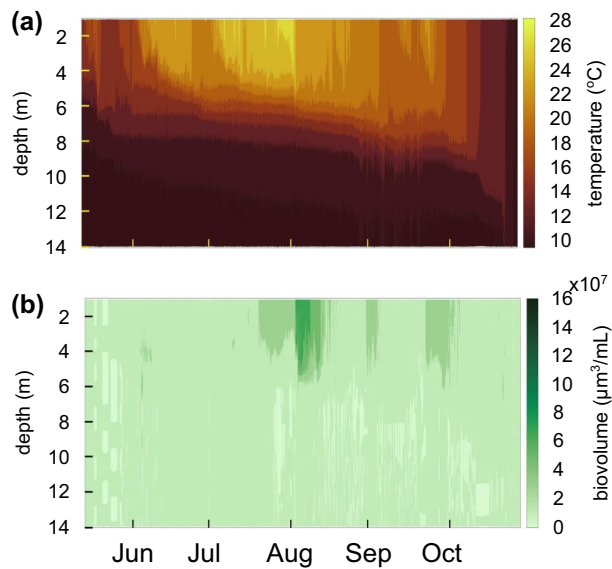
Figure 2. (a) Temperature contours and (b) biovolume contours at the location of the research station over the entire observation period. Air-water interface is at 0 m.

and biological processes in freshwater cyanobacteria dynamics to inform the final reduced-complexity models. Further, stepwise regression requires assumptions of model form; here we assume dependent variables will vary linearly with independent variables or with the interaction of independent variables. To ensure a lack of full understanding of the system was not impacting the potential accuracy of modeling results, predictor variable importance estimates were generated through a random forest analysis that was conducted using Matlab's TreeBagger function with 200 trees and the curvature algorithm. Random forest analysis is an ensemble, unsupervised machine learning technique that requires virtually no assumptions and is appropriate when dealing with unbalanced data, such as C_{DCL} (Breiman, 2001). Important variables as determined by the random forest analysis were compared to the final variables selected for the parsimonious model. Consistencies and discrepancies were investigated.

3. Results

3.1. Seasonal Trends of Water Temperature, Biological, and Meteorological Data

The temperature and biovolume profiles depicted seasonal and vertical patterns of thermal stratification and cyanobacterial accrual from June to October 2017 (Figure 2). Temperature stratification over the lake depth was established before June 2017 and the lake experienced thermal structure overturn shortly before November. Significant subsurface peaks in biovolume appeared at the beginning of June and July, and a surface bloom formed in early August. Grab samples (Section 2.3) show *Microcystis spp.* was the dominant biovolume genera up until the surface bloom in early August when *Planktothrix spp.* began to dominate the composition of cyanobacteria, however *Planktothrix spp.* and *Dolichospermum spp.* were present at low abundance throughout the observation period (Wilkinson et al., 2020). For the remainder of the analysis, we will focus only on the period of stratification from June to September, with a particular emphasis from June until the harmful algal bloom in early August when *Microcystis spp.* was dominant.

Time series of wind speed and air-water temperature difference are included to provide a sense of the meteorological forcings of the thermal and algal structure of the lake (Figure 3). Overall, surface water temperatures were higher than the air temperatures, thereby indicating prevalence of natural cooling due to the heat loss from the surface mixed layer (Figure 3b) The seasonal trend of four important water column depths: the mixed layer depth (h_{ML}), the center of gravity of the DCL biovolume (z_{CG}), the thermocline depth (h_T), and the euphotic depth (h_{EP}) are depicted in Figure 3c. The euphotic depth was essentially always well below the thermocline depth. The relative magnitude of the peak DCL biovolume concentration increases for a significant period of time in early June and early July (Figure 3d). The data demonstrate that a DCL is associated with low wind speeds and surface water temperatures warmer than airr temperatures. Both subsurface peaks eventually disperse, and do not form surface blooms. Equations 6 and 7 appear to appropriately characterize cyanobacterial vertical heterogeneity (see Appendix B for further discussion).

3.2. DCL Formation

Observational data suggest DCLs are formed during periods of little to no wind shearing and natural convection (Figures 3a and 3b). Using Equation 5 as a Reynolds number relevant to surface thermal cooling, profiles undergoing conditions such that $Re_{NC} > 5 \times 10^4$ virtually never formed a DCL for the entirety of the observation period (Figure 4). Further, profiles with a DCL tend to have shallow mixed layer depths (on average, $h_{ML} = 1.8$ m) and conditions close to $T_{surf} = T_{air}$. Conversely, profiles without a DCL had a wide range of mixed layer depths (on average, $h_{ML} = 3$ m) and tended toward conditions such that $T_{surf} > T_{air}$ (Figure 5).

These bulk trends can be seen in the diurnal evolution of biovolume profiles. To illustrate this point, we distinguish between two example days of interest: (a) a day with consistently low wind and T_{surf} switching
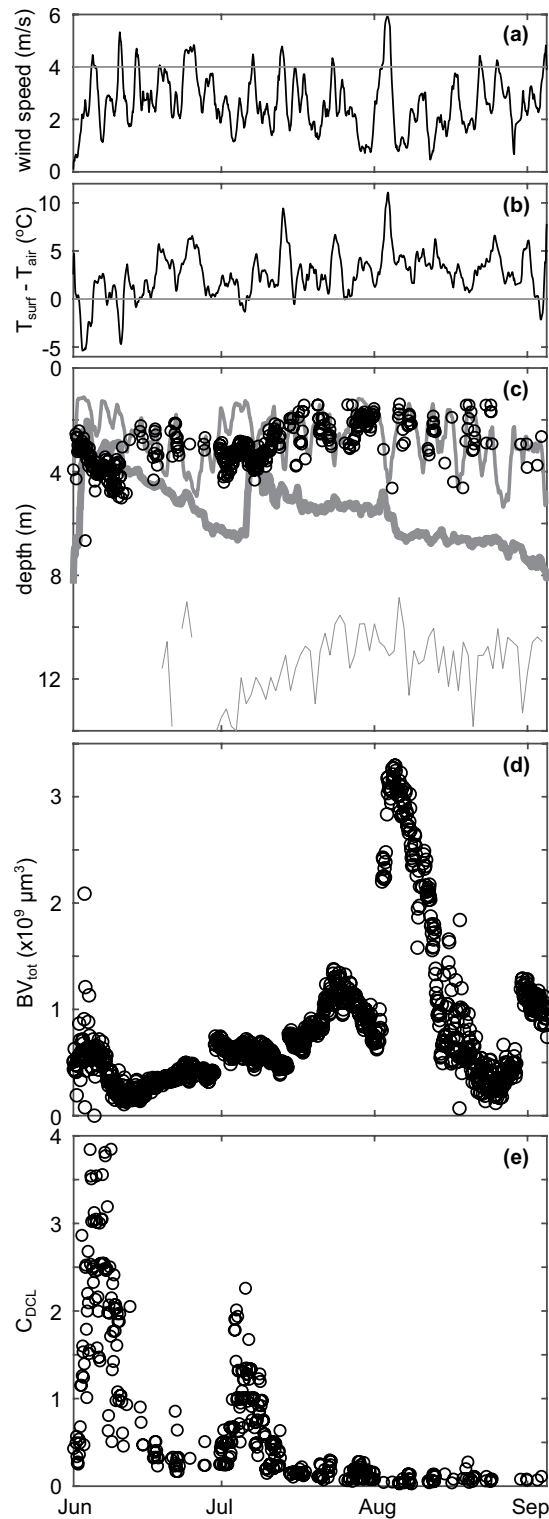
Figure 3. Time series of (a) wind speed (thin horizontal gray line indicates wind speed of 4 m/s), (b) $T_{surf}-T_{air}$ (thin horizontal gray line indicates a temperature difference of zero), (c) important depths (mixed layer depth as the medium thickness gray line, thermocline depth as the thick gray line, euphotic depth as the thin gray line, and center of gravity of the DCL—z_{CG}, as calculated by Equation 6—as the open black circles), (d) depth-integrated biovolume (BV_{tot}), and (e) relative peak magnitude of the DCL biovolume. Air-water interface is at 0 m. There are two periods with significantly large C_{DCL} values: June 1st–June 13th and July 1st–July 9th. A surface bloom occurred on August 3, 2017. Data were smoothed over a 24-h window in order to clearly show long-term trends, with the exception of z_{CG}, BV_{tot}, and C_{DCL}, which were left as calculated.
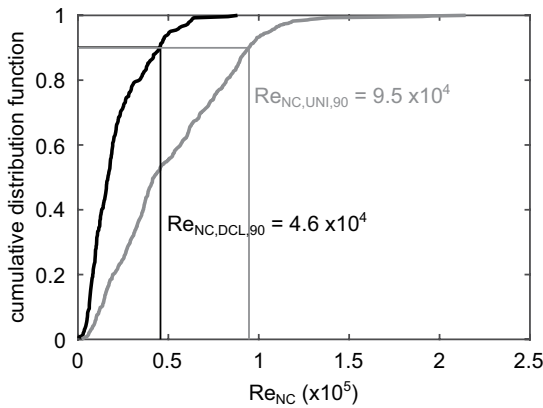
Figure 4. Surface thermal cooling Reynolds number cumulative distribution function (CDF). Black line is the CDF for profiles with a deep cyanobacteria layer (DCL), gray line is the CDF for profiles without a DCL. Ninety percentage of profiles with a DCL have concurrent Re_{NC} values of 4.6×10^4 or less; 90% of profiles without a DCL have concurrent Re_{NC} values of 9.5×10^4 or less.

between greater than and less than T_{air} (Figure 6), and (b) a day with consistently high wind and $T_{surf} > T_{air}$ (Figure 7). Comparing biovolume concentration profiles to temperature profiles in the context of surface processes like wind shear and natural convection seems to suggest a relationship between the stability of the thermal structure and heterogeneity of the vertical biovolume distribution. Thermal profiles with a poorly defined surface mixed layer, corresponding to low wind (small Re) and $T_{surf} > T_{air}$ (large Re_{NC}), are associated with biovolume profiles with a surface or near-surface peak (Figure 6, 00:00–10:00). However, when natural convection is not occurring and wind speeds remain relatively low, a subsurface peak begins to form (Figure 6, 12:00–18:00). Once natural convection is back on, the subsurface peak tends again toward a surface peak (Figure 6, 20:00–22:00). Thermal profiles with a well-defined and deep surface mixed layer, corresponding to high wind (large Re) and $T_{surf} > T_{air}$ (large Re_{NC}), are associated with biovolume profiles with all the biomass uniformly distributed within the surface mixed layer (Figure 7).

3.3. DCL Modeling Results

Since nutrient conditions did not appear to be a significant driver of cyanobacterial vertical heterogeneity throughout the observation period of South Center Lake (Appendix C), the protocol outlined in Section 2.6 was used to fit models to C_{DCL} and z_{CG} (Table 2). The euphotic depth was deeper than the DCL for most of the observation period, so there was an insignificant difference between daytime and nighttime biovolume profile behavior, and only results for all profiles are shown.

While the models outlined in Table 2 can predict values of C_{DCL} and z_{CG} to a relatively high degree, the models are rather cumbersome and difficult to assign physical meaning to. In an effort to achieve conceptually sound models that retain statistical significance, parsimonious models were generated using the full stepwise regression results as a first iteration and observational data as guides in the following manner. Upon inspection of Figures 3c and 3d, the difference in peak magnitude between the June and July periods could be due to the difference in thermocline depth: a deeper thermocline depth in July seems to have led to a lower magnitude relative peak. This is not to say that C_{DCL} is responding to the thermocline depth necessarily, but at least the processes that affect the thermocline depth also affect C_{DCL}. Further, location
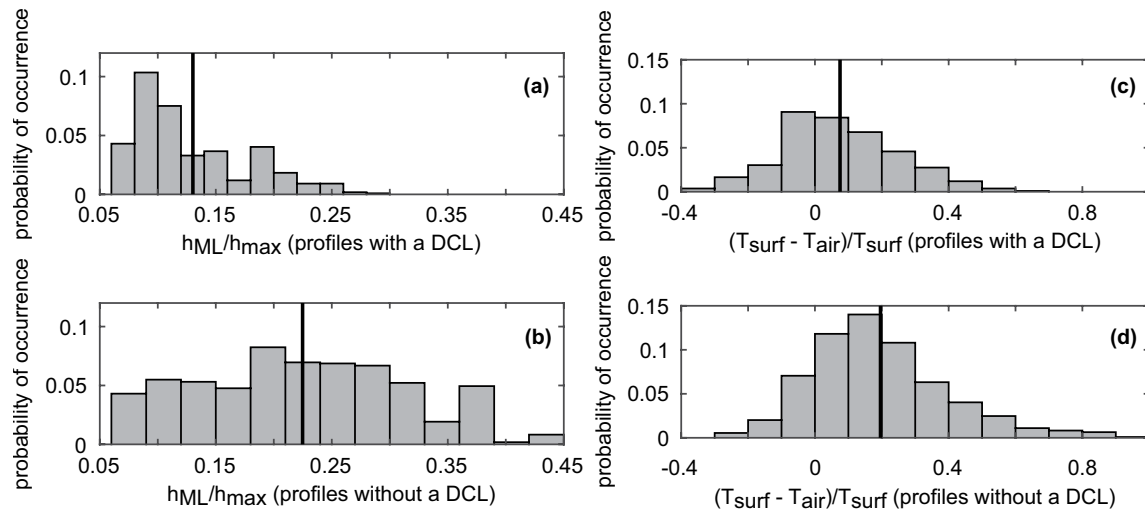

Figure 5. Histograms of (a) normalized mixed layer depth for profiles with a deep cyanobacteria layer (DCL), (b) normalized mixed layer depth for profiles without a DCL, (c) normalized air-water temperature difference for profiles with a DCL, and (d) normalized air-water temperature difference for profiles without a DCL. Vertical lines indicate mean values of (a) 0.13, (b) 0.22, (c) 0.076, and (d) 0.19.
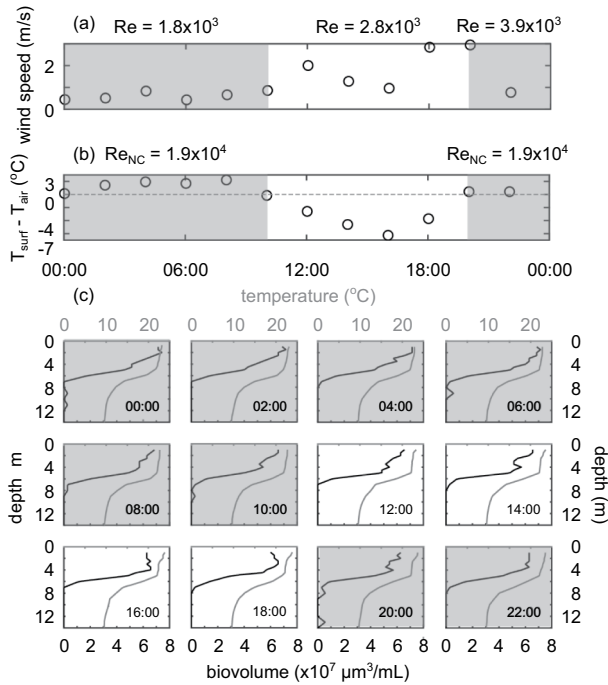
Figure 6. Biovolume and temperature profile evolution over the course of a convection-driven day (August 6, 2017). (a) Wind speed (m/s) time series. (b) $T_{surf}-T_{air}$ time series. Dashed gray line indicates $T_{surf}-T_{air} = 0$. If $T_{surf}-T_{air} > 0$, the water column is undergoing cooling and convective instabilities at the water surface. During this day, over half of the profiles were undergoing convective processes. (c) Biovolume (solid black lines, bottom x-axis) and temperature (solid gray lines, top x-axis) profiles measured over the course of the day. All axes are identical. See how the biovolume profiles tend toward a surface peak when $T_{surf}-T_{air} > 0$, but a subsurface biovolume peak starts to form during the period when $T_{surf}-T_{air} < 0$.

of the DCL center of gravity seems to be related to the mixed layer depth (Figures 6 and 7). Another way to explain these observations is through a time scale analysis. The thermocline depth deepens seasonally, with relatively little variation over the course of the day (Figure 3c). Likewise, we expect C_{DCL} to follow a more seasonal progression, since it depends on the (fairly slow) growth rate of algae. On the other hand, the mixed layer depth varies a great deal diurnally (Figure 3c). Since the center of gravity also responds to the fluid motion, it is expected for the center of gravity to vary on a similar time scale, and indeed it does. Using these observations, input parameters from Tables 1 and 2 were reduced to just thermocline depth parameters, h_T and T_T, for predicting C_{DCL} and mixed layer depth parameters, h_{ML} and T_{ML}, for predicting z_{CG}. The last step was non-dimensionalizing predictor variables. Because both models were able to be reduced to a single and unique length scale and temperature scale, other variables of similar units were not used in the dimensionalization process to prevent adding back in complexity. Instead, length scales were normalized by the depth of the water column at the location of the research station, $h_{max} = 14m$, and temperature scales were multiplied by the coefficient of thermal expansion, α (1/°C). In summary, $\tilde{h}_{ML} = \frac{h_{ML}}{h_{max}}$, $\tilde{h}_T = \frac{h_T}{h_{max}}$, $\tilde{T}_{ML} = \alpha T_{ML}$, and $\tilde{T}_T = \alpha T_T$, thereby reducing Equations 8 and 9 to $z_{CG} = \tilde{f}(\tilde{h}_{ML}, \tilde{T}_{ML})$ and $C_{DCL} = \tilde{g}(\tilde{h}_T, \tilde{T}_T)$, respectively. Using these reduced complexity inputs in a linear regression analysis generated the following expressions:

$$z_{CG} = 0.20 + 0.57\tilde{h}_{ML} - 0.67\tilde{T}_{ML} \tag{10}$$

which explains the data ($n = 417$ profiles) with an r^2 of 0.63 and a root-mean-squared error (RMSE) of 0.036 (Figure 8). Equation 9 reduces to

$$\ln C_{DCL} = 1.5 - 3.2\tilde{h}_T - 21\tilde{T}_T \tag{11}$$

which explains the data ($n = 417$ profiles) with an r^2 value of 0.69 and an RMSE of 0.68. These parsimonious models reduce input parameters to two while maintaining a large portion of the statistical significance (Figure 8). Essentially, Equation 10 states that the DCL center of gravity depends on mixed layer conditions, whereas Equation 11 states that the DCL relative peak magnitude depends on thermocline conditions. While it is likely the intercepts and even coefficients of both equations will vary from lake to lake, we anticipate the dependencies on the linear combination of mixed layer parameters for z_{CG} and the linear combination of thermocline parameters for C_{DCL} will remain for similar lakes.

Results from a random forest analysis supported the predictor variable decisions in the reduced complexity models (Figure 9). The two predictor variables chosen to describe z_{CG} were both in the top three most important predictor variables as determined by the random forest analysis. Similarly, the two predictor variables chosen to describe C_{DCL} were in the top three most important predictor variables as determined by the random forest analysis.

4. Discussion

4.1. DCL Dynamics

Equation 5 states that $Re_{NC} \sim h_{ML}\left(T_{surf} - T_{air}\right)^{\frac{1}{3}}$. These results reveal two important points. The first is that the deeper the mixed layer depth, the less likely it is that a DCL will form (Figures 5a and 5b). In fact, the two situations seem to have two entirely different distributions of mixed layer depths: profiles with a DCL follow a power law, whereas profiles wihtout a DCL follow a uniform distribution. Since the euphotic
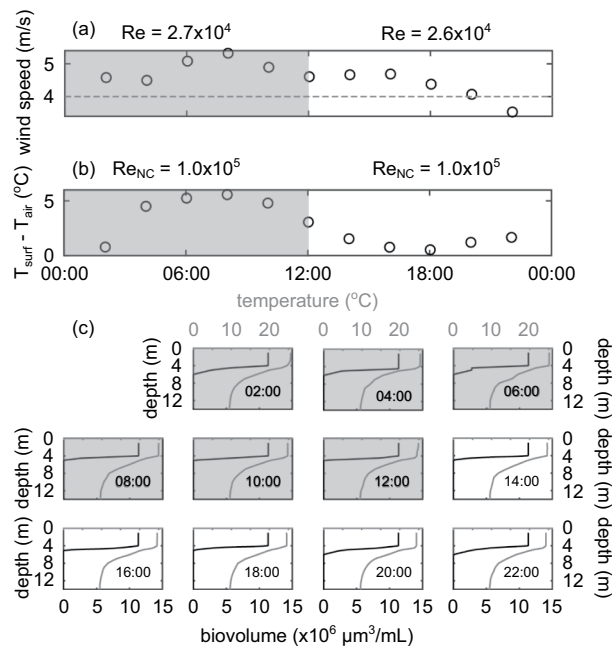
Figure 7. Biovolume and temperature profile evolution over the course of a convection-driven day (July 13, 2017). (a) Wind speed (m/s) time series. Dashed gray line indicates a wind speed of 4 m/s. (b) $T_{surf}-T_{air}$ time series. During this day, all of the profiles were undergoing convective processes. (c) Biovolume (solid black lines, bottom x-axis) and temperature (solid gray lines, top x-axis) profiles measured over the course of the day. All axes are identical. See how the biovolume profiles remain uniform throughout the course of the day.

depth was well below the thermocline for the entirety of the observation period (Figure 3c), light was not limiting the aggregation of *Microcystis* below the mixed layer depth. The lack of a DCL, then, is likely due to the physical drivers, like wind shearing (Pollard et al., 1972; Ushijima & Yoshikawa, 2020), of a deepening mixed layer depth. Second, from Equation 5, we can also infer that the greater the surface water temperature is relative to the air temperature, the less likely it is that a DCL will form (Figures 5c and 5d). This finding largely agrees with the work of Somavilla et al. (2019), who found that surface peaks of chlorophyll in the ocean occurred concurrently with heat loss events at the ocean surface. The temperature difference ($T_{surf} > T_{air}$) promotes natural convection-induced turbulence at the water surface. In effect, the data indicate the likelihood of a DCL occurring decreases as surface layer processes like wind shearing and natural convection increase. This point is made more apparent when considering the diurnal evolution of cyanobacterial vertical distributions (Figure 6). When thermal convection is occurring at the air-water interface, the biovolume profile exhibits a surface peak. As soon as the thermal convection is turned off, a subsurface peak begins to form. When thermal convection is turned on again, the subsurface peak is diminished. Although direct measurements of turbulent kinetic energy were not recorded, these results corroborate the findings of Zhu et al. (2018), who demonstrated that *Microcystis* colonies tend to aggregate at depths where the turbulent kinetic energy of the surrounding water is approximately equal to the critical vertical turbulent kinetic energy of the colony under steady conditions and only exhibited diurnal migration under similar conditions. With real world applications in mind, however, measuring water temperature profiles is more practical than measuring turbulence levels of the water column. Simple measurements make practical mitigation strategies. Our findings suggest thermistor chain measurements alone can indicate when and where subsurface biomass measurements are necessary.

Dimensional analysis results indicate the center of gravity of the DCL biovolume is controlled by mixed layer depth and temperature (Equation 10), whereas the magnitude of the DCL biovolume peak is controlled by thermocline depth and temperature (Equation 11). Because *Microcystis* moves up and down a water column by adjusting cell density, we expect the thermocline depth–the depth of the largest magnitude density gradient–to act as a boundary condition. Not necessarily impenetrable (see early June period in Figure 3c where the center of gravity of the DCL is actually below the thermocline depth), but more like a discontinuous step change from high to low diffusivity. Therefore, as the thermocline deepens, the width of habitable space for the algae increases. In this situation, we would expect the algae to diffuse throughout the entire habitable space, thereby decreasing peakiness. Following similar logic, increasing the temperature at the thermocline depth would also increase this habitable space, since temperature is monotonically decreasing and algae like it hot (Paerl & Huisman, 2008), so we would again expect a more diffuse and less peaky distribution of algae.

Next, since well-mixed conditions will likely impede algal aggregation, we expect the center of gravity to deepen as the mixed layer depth deepens. Conversely, higher temperatures (still below the lethal limit) increase the buoyant velocity of *Microcystis* (Thomas & Walsby, 1986; You et al., 2018), so we would expect the center of gravity to move shallower as the temperature at the mixed layer increases. The algae can be thought of as acting as both active and passive particles in a balance between hydrodynamic forcings and biological preferences. The cyanobacteria are subjected to environmental mixing conditions, but they are also able to recognize and move toward preferable depths. These results support a time-scale hierarchy analysis like the ones conducted by Marti et al. (2016) and Wallace et al. (2000) for determining vertical distributions of cyanobacteria.

Table 2
Stepwise Regression Results

DCL variable	Parameter	Coefficient estimate	t-stat	*p*-value
(a) z_{CG}, all profiles				
$n = 417$, $r^2 = 0.67$, RMSE = 0.034 (0.099 $\leq z_{CG} \leq$ 0.48)				
	intercept	-1.3	-6.1	2.7×10^{-9}
	h_{ML}	0.014	1.5	0.14
	h_T	0.14	8.4	7.5×10^{-16}
	T_{ML}	0.077	7.9	2.2×10^{-14}
	T_T	0.054	5.4	1.0×10^{-7}
	$h_{ML}h_T$	0.0056	3.0	0.0025
	$h_T T_{ML}$	-0.0065	-7.9	3.0×10^{-14}
	$h_T T_T$	-0.0016	-2.9	0.0038
	$T_{ML}T_T$	-0.0025	-6.8	4.4×10^{-11}
(b) $\ln C_{DCL}$, all profiles				
$n = 417$, $r^2 = 0.69$, RMSE = 0.678 ($-3.6 \leq \ln C_{DCL} \leq 1.3$)				
	intercept	-21	-4.9	1.6×10^{-6}
	h_{ML}	0.69	3.6	0.00031
	h_T	2.4	7.3	2.1×10^{-12}
	PAR_{ML}	7.3×10^{-5}	2.5	0.014
	T_{ML}	1.5	7.6	1.6×10^{-13}
	T_T	0.25	1.4	0.16
	$h_{ML}h_T$	-0.12	-3.1	0.0019
	$h_T T_{ML}$	-0.15	-9.5	1.3×10^{-19}
	$T_{ML}T_T$	44,000	-4.5	1.1×10^{-5}

Notes. Stepwise regression results for **(a)** the center of gravity of the DCL biovolume (F-statistic vs. constant model: 117, *p*-value: 1.2×10^{-100}) and **(b)** the relative peak magnitude of the DCL biovolume (F-statistic vs. constant model: 106, *p*-value = 1.1×10^{-94}). Full regression parameters listed in Table 1 and Equations 8 and 9.

Abbreviations: DCL, deep cyanobacteria layer; RMSE, root-mean-squared error.

Equations 10 and 11 predict recorded values of C_{DCL} and z_{CG} to a reasonable accuracy with statistical significance for all profiles (Figure 8). The physically informed decisions made by the authors were supported by the random forest analysis: the predictor variables determined to be the most important through intuition were indeed among the most important predictor variables as determined by the random forest analysis. Discrepancies could reflect nonlinear relationships. For example, the second most important predictor variable for z_{CG} as determined by the random forest analysis was the thermocline depth. However, if the thermocline depth was included in the input predictor variable matrix for the reduced complexity stepwise model of z_{CG}, it was not found to have a high enough statistical significance to include in the final model, and only $\tilde{h}_{ML}$ and $\tilde{T}_{ML}$ were included in the final model. This could be due to the linear and bi-linear constraint on the stepwise model. If predictor variables do not have a linear relationship with the response variable, they will not be included in the final model. No such restrictions are enforced in the random forest analysis. z_{CG}, then, could have a nonlinear dependence on h_T, including but not limited to threshold effects. In fact, Nelson et al. (2018) uncovered many threshold effects in their random forest analysis of *Microcystis* abundance, the most prominent of which was fluid flow.

Given a temperature profile, Equations 10 and 11 will output DCL relative peak magnitude and center of gravity. Lakes are complex ecosystems, and cyanobacteria are remarkably sensitive to all different kinds of forcings, hydrodynamic and biological alike. However, it seems that a significant portion of the hydrodynamic dependencies can be packaged into thermal structure parameters that are relatively easy to measure. Further, in the stratified and eutrophic conditions seen throughout the observation period, biological forcings, which are difficult and time-consuming to measure, are secondary controls of the vertical distrubtion of cyanobacterial biomass. Although definitions differed slightly, the findings presented in this paper corroborate the findings of Scofield et al. (2017) at Lake Ontario: thermal structure parameters can explain significant variability of where in the water column a DCL forms and how large the peak magnitude of the DCL gets. The machine learning algorithm developed by W. Xu et al. (2019) to determine key parameters and patterns of thermal and biological vertical profiles could be used to systemically determine C_{DCL} and z_{CG} from field data, in order to provide more consistent protocol and definitions between studies.

It's prudent to note that these modeling results only make sense for the range of input variables given. For example, if the temperature in the mixed layer ever got too hot to be lethal to algae, any increase in temperature would likely lead to algae moving to cooler temperatures at deeper depths in the water column, and the DCL center of gravity would increase. However, the hottest water temperatures recorded at South Center Lake over the summer of 2017 was approximately 28°C, which is right around the ideal temperature for *Microcystis*. Further, if the euphotic depth had not been below the thermocline for almost the entirety of the observation period, it is likely the euphotic depth would have had a statistically significant relationship with both C_{DCL} and z_{CG}, as predicted by a wealth of literature relating cyanobacteria abundance and motility to light conditions (see the introduction of Yao et al., 2017). Relevant time scales could also be a factor accounting for the low dependency on light. Perhaps light levels control DCL parameters on a diurnal basis, but that effect gets smoothed out when modeling seasonal trends. Lastly, nutrient concentrations had relatively low variability throughout the duration of the observation period (Appendix C). Changes in nutrient availability, either temporally or spatially, would likely lead to different results.
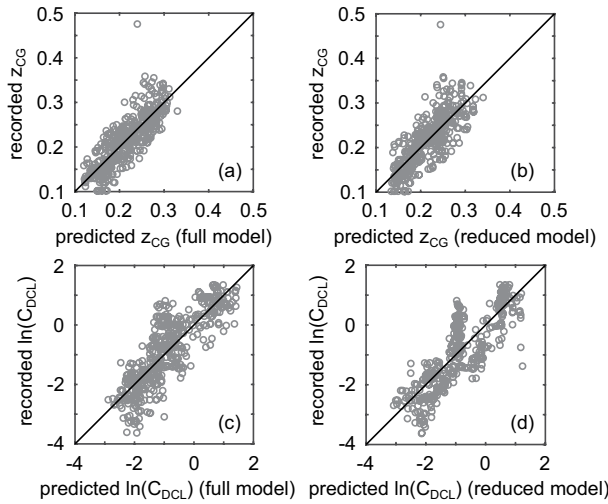
Figure 8. Goodness-of-fit results for models of (a) z_{CG} using full stepwise regression results in Table 2a ($n = 417$, $r^2 = 0.66$, RMSE = 0.034), (b) z_{CG} using the simplified Equation 10 ($n = 417$, $r^2 = 0.42$, RMSE = 0.044), (c) log-transformed C_{DCL} using full stepwise regression results in Table 2b, ($n = 417$, $r^2 = 0.65$, RMSE = 0.52), and (d) log-transformed C_{DCL} using the simplified Equation 11 ($n = 417$, $r^2 = 0.57$, RMSE = 0.57). Solid black lines show a one-to-one line.

4.2. Exploratory Observations on the Fate of DCLs

We have so far explored the environmental conditions that lead to DCL formation, but the intended future direction of this analysis is to understand the fate of DCLs. Does the subsurface peak disperse and form a uniform algae profile, or does it form a surface bloom? Although vertical heterogeneity has been qualitatively correlated to subsequent surface bloom formation, there is a lack of quantitative modeling to support this observation (Liu et al., 2019; Xiao et al., 2018; Zhu et al., 2018). While it is beyond the scope of this paper to make definitive claims on the topic, some observations from this data set could orient future work in the right direction. Both the June and July DCLs develop large peaks, although the June DCL center of gravity travels much deeper than the July DCL. During both periods, water temperatures in the surface layer are high enough to sustain cyanobacterial life, nutrient conditions are sufficient, and yet neither subsurface peaks form a surface bloom. To determine if the variables introduced previously could be put to use investigating the fate of these DCLs, we introduce a new term, the normalized DCL center of gravity, ζ, defined as

$$\zeta = \frac{z_{CG}}{\dfrac{1}{h_{max}} \dfrac{\int_0^{h_{max}} zCdz}{\int_0^{h_{max}} Cdz}} \tag{12}$$

Equation 12 is a ratio of the center of gravity of the DCL to the center of gravity of the whole profile. This will give a measure of the abnormality of the DCL peak. For example, if the entire biovolume profile followed a Gaussian distribution, then the center of gravity of the DCL would equal the center of gravity of the whole profile, and $\zeta = 1$. This would also be true if biovolume concentrations were zero everywhere except the DCL. However, if there is significant biovolume in the mixed layer, then this would shift the center of gravity of the whole profile shallower, relative to the center of gravity of just the DCL, resulting in $\zeta > 1$. Similarly, ζ will be less than one if the DCL fails to capture all of the biovolume below the mixed layer.

It was determined that, although the values of z_{CG} were significantly different for the June and July peaky period, the ζ values of each were approximately the same, $\zeta = 1$ (Figure 10). Further, there appears to be a sinusoidal seasonal trend in ζ that could indicate when a bloom will occur (thick black line in Figure 10). The bloom occurs at $\zeta = 0.64$, according to the fitted sine curve; this is also the value of ζ when averaged over the three days immediately preceding the bloom (thin gray vertical line demarcates when the bloom took place). To check this trend, 2018 data from Ramsey Lake—a deep, dimictic, and eutrophic lake with a history of *Microcystis* blooms—were investigated (Section 2.1). For profiles taken when the euphotic depth was deeper than the thermocline depth, we see a similar trend, with algal blooms occurring when ζ values dip below the approximate 0.6 threshold.

To further explore this, we consult individual profiles for $\zeta < 1$, $\zeta = 1$, and $\zeta > 1$ (Figure 11). When $\zeta = 1$ (Figure 11a), as is the case for the peaky periods in June and July, we see low biovolume concentrations in the mixed layer and below the DCL. In these situations, integrating biovolume over the entire water column is essentially the same as integrating biovolume just within the DCL, hence the center of gravity of the whole profile aligns with the center of gravity of just the DCL. When $\zeta > 1$ (Figure 11b), as is the case in between the two peaky periods in June and July, we see large concentrations of biovolume within the mixed layer relative to the biovolume concentration in the DCL. This brings the center of gravity of the whole profile shallower, when compared to the center of
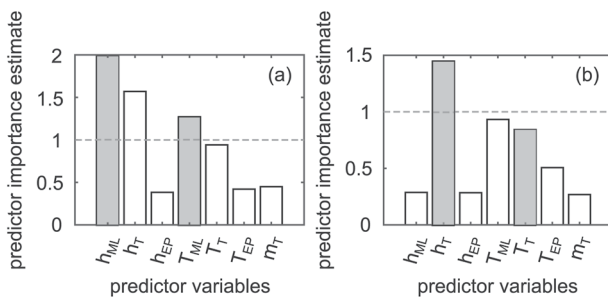

Figure 9. Predictor importance estimates for (a) z_{CG} and (b) C_{DCL}. An arbitrary cut-off of 1 (horizontal dashed gray line) was chosen for both variables to determine the three most important predictor variables for z_{CG}. Shaded bars indicate variables used in the parsimonious models, Equations 10 and 11.
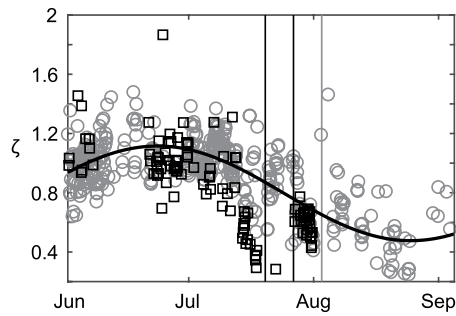
Figure 10. Seasonal trend of ζ from South Center Lake in 2017 (open gray circles) and Ramsey Lake in 2018 (open black squares, see Section 2.1). Thick black line indicates the best fit periodic function to South Lake determined to be $\zeta = 0.80 + 0.32\sin(0.05t + 7 \times 10^5)$, where t here is the Matlab serial date number ($r^2 = 0.56$, p-value $= 3.2 \times 10^{-75}$). Thin black vertical lines indicate dates of harmful algal blooms in Ramsey Lake (20 and 27 July 2018), and thin gray vertical line is the date of the harmful algal bloom in South Center Lake (August 3, 2017).

gravity of just the DCL. When $\zeta < 1$ (Figure 11c), as is the case just before the surface bloom, we see a narrow DCL but a wide distance between the mixed layer depth and the thermocline depth. Using the same logic that informed our model in the previous section, a deep thermocline gives algae a larger habitable space. So even though the DCL itself occupies a narrow band immediately below the mixed layer, not insignificant populations of algae are capable of living below the DCL, thus moving the center of gravity of the whole profile deeper compared to the center of gravity of just the DCL.

Upon qualitative inspection, profiles with $\zeta < 1$ appear to have a much easier journey from DCL to surface HAB compared to profiles with $\zeta \geq 1$. Recall also the phenomenon of bioconvection, in which the motion of dense algae introduces hydrodynamic instabilities into the water column (Sepúlveda Steiner et al., 2019; Sommer et al., 2017). Profiles with $\zeta \geq 1$ appear more likely to induce hydrodynamic instabilities, potentially changing the thermal structure of the lake, and inhibiting surface bloom formation.

The aim of this analysis is not to be conclusive, but rather provocative, in the hopes of fueling further thought and research. The center of gravity parameter as defined and used in this paper is new to the field of DCL research, but its usefulness appears hopeful. If the shape of the DCL biovolume relative to the shape of the entire biovolume profile does, in fact, exhibit a predictable seasonal trend, as is suggested in Figure 10, then this could be a missing link between predicting vertical distribution of algae and predicting harmful algal bloom formation. However, it is difficult to separate fact from coincidence with data from two lakes which experienced a total of three cyanobacteria blooms. We present our observations and qualitative analysis as an initial hypothesis which, we hope, can be tested more completely in the future.

5. Conclusions

A high frequency, long-duration research station was anchored in a eutrophic and dimictic lake for the entirety of summer stratification. This research station recorded meteorological measurements every 5 min and water quality profiles every 2 h. Two key parameters were introduced to describe cyanobacterial vertical heterogeneity: C_{DCL}, a measure of the relative peak biovolume concentration magnitude, and z_{CG}, the center of gravity of the biovolume concentration with the DCL. A stepwise regression analysis was performed to determine the dependence of these two variables on abiotic parameters of the lake, and results were compared to those of a random forest analysis, a more unsupervised approach.

A DCL was present for a large portion of the majority of the summer season. Results indicate the magnitude of the DCL peak depends on physical conditions at the thermocline, but the center of gravity of the DCL depends on physical conditions at the surface mixed layer. A deep mixed layer depth is correlated with a
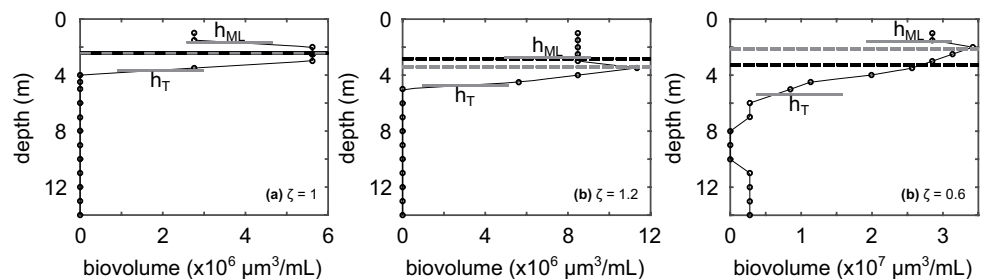

Figure 11. Examples of profiles for (a) $\zeta = 1$, profile taken July 1, 2017 04:00, (b) $\zeta > 1$, profile taken June 21, 2017 02:00, and (c) $\zeta < 1$, profile taken at July 31, 2017 12:00. Black dashed lines indicate center of gravity of the whole biovolume profile, and gray dashed lines indicate center of gravity of the DCL.

deep DCL, but a shallow thermocline depth is correlated with a DCL with a large biovolume peak. Increases in water temperature will promote a shallower DCL with a more diffuse biovolume population. It was also shown that a large Reynolds number related to surface cooling, $Re_{NC} > 5 \times 10^4$, inhibits the formation of a DCL. Conceptually, an increase in turbulence-inducing processes at the air-water interface, including wind shear and natural convection, decreases the occurrence of a DCL.

Situated neatly in a wealth of research connecting abiotic forcings to algal vertical distributions and surface bloom formation, this work connects results from low frequency, long-duration field studies and high frequency, short-duration experiments. The models and parameters presented in this research not only provide phenomenological insight toward *Microcystis* vertical distributions, but their utilitarian simplicity in both form and ease of measurement will also help inform cyanobacteria mitigation strategies available to water systems stakeholders. Re_{NC}, which can give an indication whether subsurface biomass measurements are necessary or not, can be calculated using meteorological and thermistor chain measurements alone. Similarly, water systems managers only need knowledge of a lake thermal profile to quantify the relative subsurface peak magnitude and approximate where the center of gravity of the subsurface biomass will occur. The findings will help inform the location of subsurface biomass measurements and consequently facilitate mitigation strategies.

Appendix A: Example Profiles

Figure A1 illustrates typical biovolume (BV) profiles, typical light and temperature conditions, and DCL concepts discussed in Sections 2.2–2.5. A DCL is only defined when there exists a concentration of biovolume below the mixed layer that is greater than the average biovolume within the mixed layer. This definition ensures insignificant increases in biovolume below the mixed layer do not get incorrectly labeled as a DCL. When a DCL does exist, the center of gravity of the DCL, z_{CG}, is the depth at which there is equal biovolume mass in the DCL specifically above and below that depth. For a normally distributed DCL, the center of gravity will be located at the depth of maximum biomass. This will not necessarily be true if there is a biovolume tail from above or below or if there are multiple biovolume peaks within the DCL.
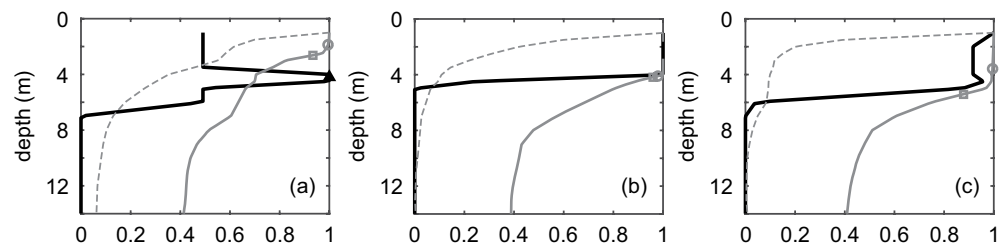

Figure A1. Example profiles from (a) June 10, 2017 12:00, showing a profile with a deep cyanobacteria layer (DCL), (b) July 13, 2017 14:00, showing a profile with a uniform biovolume (BV) profile, and (c) August 4, 2017 16:00, showing a profile that is not uniform but is not classified as a DCL because the subsurface BV peak is not larger than the average BV concentration in the mixed layer. Solid black lines are the BV profiles (normalized by the maximum BV concentration in the profile), solid gray lines are the temperature profiles (normalized by the surface temperature), and dashed gray lines are the photosynthetically active radiation (PAR) profiles (normalized by the surface PAR intensity). Gray circle markers indicated mixed layer depth, gray squares indicate thermocline depth, and the black triangle indicates the depth of the DCL, when a DCL is present. Water surface is at 0 m, and the lake bed is at 14 m. For each respective profile, we have (a) $T_{surface} = 23°C$, $BV_{max} = 2.8 \times 10^6 \, \mu m^3/mL$, $PAR_{surface} = 150 \, \mu mol/s/m^2$, (b) $T_{surface} = 25°C$, $BV_{max} = 1.1 \times 10^7 \, \mu m^3/mL$, $PAR_{surface} = 150 \, \mu mol/s/m^2$, (c) $T_{surface} = 23°C$, $BV_{max} = 6.9 \times 10^6 \, \mu m^3/mL$, and $PAR_{surface} = 100 \, \mu mol/s/m^2$.

Appendix B: Vertical Heterogeneity and the DCL

To ensure DCL definitions are physically meaningful and do not obscure biovolume vertical heterogeneities within the mixed layer, we introduce

$$C_{ML} = \frac{C_{ML,max} - C_{ML,avg}}{C_{ML,avg}} \tag{B1}$$

Equation B1 gives the magnitude of the biovolume concentration peak within the mixed layer relative to the average concentration in the mixed layer. Our assumption is that biovolume vertical heterogeneities can only occur below the mixed layer, because mixing processes within the mixed layer will dominate any depth-specific growth or migration of cyanobacteria, thus wiping out any possible biovolume aggregation. Values of C_{ML} close to zero will validate this assumption.

Results indicate an average value of $C_{ML,avg} = 0.058$, which is much lower than the average value of C_{DCL}, $C_{DCL,avg} = 0.79$ (Figure B1). Further, there appears to be no significant difference in the distribution of C_{ML} between profiles with and without a DCL. Meaning the existence of a DCL does not impact the biovolume heterogeneity, or lack thereof, within the mixed layer. For these reasons, the authors suggest that the definition of a DCL, its bounds, and its parameters detailed in Section 2.5 accurately describe biovolume vertical heterogeneities.
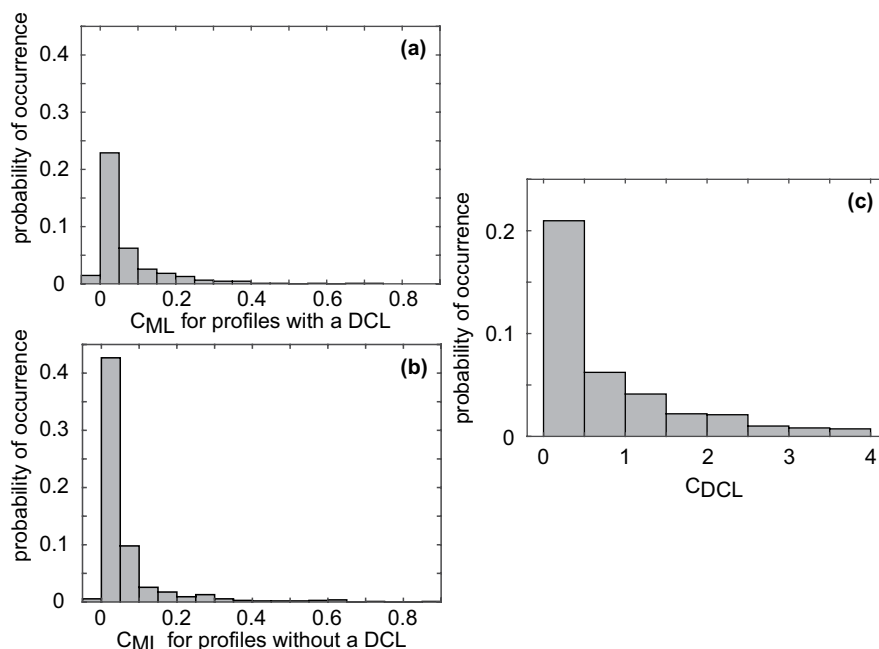

Figure B1. Histograms of (a) C_{ML} for profiles with a deep cyanobacteria layer (DCL), (b) C_{ML} for profiles without a DCL, and (c) C_{DCL}. There appears to be no discernible difference in C_{ML} distribution between profiles with and without a DCL. The largest value C_{ML} takes for all profiles is about 0.9, or a 90% increase in peak value from the mean value. This is much lower than the largest value of C_{DCL}, which is around 4.

Appendix C: Nutrients

Nutrient grab samples were collected on an approximately weekly basis to quantify nitrate and nitrite $(NO_3^- + NO_2^-)$ and phosphate (PO_4^{3-}) concentrations. Nitrate and nitrite concentrations were below the detection limit of 0.02 mg/L for all depths for the entire summer season, and phosphate concentrations at depths of 1, 3, 6, and 10 m are shown in Figure C1. Even at the lowest recorded phosphate concentration
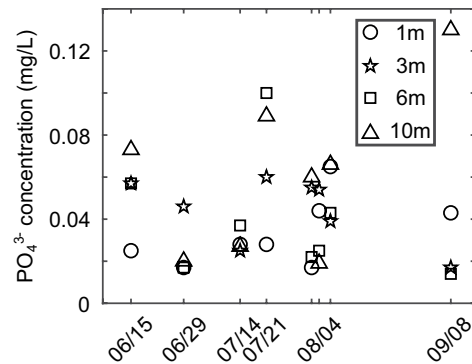
Figure C1. Phosphate concentrations (from grab samples) time series at location of research station. Circles indicate phosphate concentration at 1 m, stars at 3 m, squares at 6 m, and triangles at 10 m. Sampling dates shown on x axis.

of 0.02 mg/L, the highest N:P ratio was 1, indicating nitrogen was likely limiting *Microcystis* growth during the entire monitoring period (Marinho et al., 2007; Wurtsbaugh et al., 2019). Epilimnion concentrations of phosphate showed little variability and were comparable to hypolimnion concentrations throughout the observation period (Wilkinson et al., 2020). Surface level concentrations of phosphate (1 m data) remain consistently low until mid-September, well after the surface bloom of early August had formed and dissipated. The highest concentrations of phosphate occur in mid-July, after the July subsurface peaky period but before the August surface bloom. The phosphate profile changed from relatively uniform to almost monotonically increasing with depth in a matter of a week during this same time. However, this change in shape of the phosphate profile led to no distinguishable change in the shape of the biovolume profile. For these reasons, we suggest that nutrients played a secondary role in the vertical heterogeneity of *Microcystis* in South Center Lake for the summer 2017 season.

Data Availability Statement

Data archiving is underway at the Data Repository for University of Minnesota.

References

Bezerra-Neto, J. F., & Pinto-Coelho, R. M. (2007). Diel vertical migration of the copepod thermocyclops inversus (Kiefer, 1936) in a tropical reservoir: The role of oxygen and the spatial overlap with chaoborus. *Aquatic Ecology*, *41*(4), 535–545. https://doi.org/10.1007/s10452-007-9119-x

Bormans, M., Sherman, B., & Webster, I. (1999). Is buoyancy regulation in cyanobacteria an adaptation to exploit separation of light and nutrients? *Marine and Freshwater Research*, *50*(8), 897–906.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/a:1010933404324

Brooks, A. S., & Torke, B. G. (1977). Vertical and seasonal distribution of chlorophyll a in Lake Michigan. *Journal of the Fisheries Research Board of Canada*, *34*(12), 2280–2287. https://doi.org/10.1139/f77-306

Cao, H.-S., Kong, F.-X., Luo, L.-C., Shi, X.-L., Yang, Z., Zhang, X.-F., & Tao, Y. (2006). Effects of wind and wind-induced waves on vertical phytoplankton distribution and surface blooms of microcystis aeruginosain Lake Taihu. *Journal of Freshwater Ecology*, *21*(2), 231–238. https://doi.org/10.1080/02705060.2006.9664991

Cullen, J. J. (1982). The deep chlorophyll maximum: Comparing vertical profiles of chlorophyll a. *Canadian Journal of Fisheries and Aquatic Sciences*, *39*(5), 791–803. https://doi.org/10.1139/f82-108

Cullen, J. J. (2015). Subsurface chlorophyll maximum layers: Enduring enigma or mystery solved? *Annual Review of Marine Science*, *7*, 207–239. https://doi.org/10.1146/annurev-marine-010213-135111

Engel, L., Valley, R., & Vanderbosch, D. (2011). *Sentinel lake assessment report south center lake (13-0027) chisago county, Minnesota*. (Technical Report). Minnesota Pollution Control Agency and Minnesota Department of Natural Resources.

Fogg, G. E., & Walsby, A. E. (1971). Buoyancy regulation and the growth of planktonic blue-green algae. *SIL Communications, 1953-1996*, *19*(1), 182–188. https://doi.org/10.1080/05384680.1971.11903929

Ganf, G. G., & Oliver, R. L. (1982). Vertical separation of light and available nutrients as a factor causing replacement of green algae by blue-green algae in the plankton of a stratified lake. *Journal of Ecology*, *70*, 829–844. https://doi.org/10.2307/2260107

Gray, E., Mackay, E. B., Elliott, J. A., Folkard, A. M., & Jones, I. D. (2020). Wide-spread inconsistency in estimation of lake mixed depth impacts interpretation of limnological processes. *Water Research*, *168*, 115136. https://doi.org/10.1016/j.watres.2019.115136

Hozumi, A., Ostrovsky, I., Sukenik, A., & Gildor, H. (2019). Turbulence regulation of microcystis surface scum formation and dispersion during a cyanobacteria bloom event. *Inland Waters*, *10*, 1–20.

Huisman, J., Codd, G. A., Paerl, H. W., Ibelings, B. W., Verspagen, J. M. H., & Visser, P. M. (2018). Cyanobacterial blooms. *Nature Reviews Microbiology*, *16*(8), 471–483. https://doi.org/10.1038/s41579-018-0040-1

Huisman, J., Pham Thi, N. N., Karl, D. M., & Sommeijer, B. (2006). Reduced mixing generates oscillations and chaos in the oceanic deep chlorophyll maximum. *Nature*, *439*(7074), 322–325. https://doi.org/10.1038/nature04245

Imberger, J. (1985). The diurnal mixed layer1. *Limnology & Oceanography*, *30*(4), 737–770. https://doi.org/10.4319/lo.1985.30.4.0737

Liu, M., Ma, J., Kang, L., Wei, Y., He, Q., Hu, X., & Li, H. (2019). Strong turbulence benefits toxic and colonial cyanobacteria in water: A potential way of climate change impact on the expansion of harmful algal blooms. *The Science of the Total Environment*, *670*, 613–622.

Marinho, M. M., & de Oliveira e Azevedo, S. M. F. (2007). Influence of n/p ratio on competitive abilities for nitrogen and phosphorus by microcystis aeruginosa and aulacoseira distans. *Aquatic Ecology*, *41*(4), 525–533. https://doi.org/10.1007/s10452-007-9118-y

Marti, C. L., Imberger, J., Garibaldi, L., & Leoni, B. (2016). Using time scales to characterize phytoplankton assemblages in a deep subalpine lake during the thermal stratification period: Lake Iseo, Italy. *Water Resources Research*, *52*(3), 1762–1780. https://doi.org/10.1002/2015wr017555

Medrano, E. A., Uittenbogaard, R., Pires, L. D., Van De Wiel, B., & Clercx, H. (2013). Coupling hydrodynamics and buoyancy regulation in microcystis aeruginosa for its vertical distribution in lakes. *Ecological Modelling*, *248*, 41–56.

Nelson, N. G., Muñoz-Carpena, R., Phlips, E. J., Kaplan, D., Sucsy, P., & Hendrickson, J. (2018). Revealing biotic and abiotic controls of harmful algal blooms in a shallow subtropical lake through statistical machine learning. *Environmental Science & Technology*, *52*(6), 3527–3535. https://doi.org/10.1021/acs.est.7b05884

O'Neil, J., Davis, T., Burford, M., & Gobler, C. (2012). The rise of harmful cyanobacteria blooms: The potential roles of eutrophication and climate change. *Harmful Algae*, *14*, 313–334.

Paerl, H. W., & Huisman, J. (2008). CLIMATE: Blooms like it hot. *Science*, *320*(5872), 57–58. https://doi.org/10.1126/science.1155398

Pollard, R. T., Rhines, P. B., & Thompson, R. O. (1972). The deepening of the wind-mixed layer. *Geophysical & Astrophysical Fluid Dynamics*, *4*(1), 381–404.

Read, J. S., Hamilton, D. P., Jones, I. D., Muraoka, K., Winslow, L. A., Kroiss, R., et al. (2011). Derivation of lake mixing and stratification indices from high-resolution lake buoy data. *Environmental Modelling & Software*, *26*(11), 1325–1336. https://doi.org/10.1016/j.envsoft.2011.05.006

Reynolds, C. S., Oliver, R. L., & Walsby, A. E. (1987). Cyanobacterial dominance: The role of buoyancy regulation in dynamic lake environments. *New Zealand Journal of Marine & Freshwater Research*, *21*(3), 379–390. https://doi.org/10.1080/00288330.1987.9516234

Sanful, P. O., Otu, M. K., Kling, H., & Hecky, R. E. (2017). Occurrence and seasonal dynamics of metalimnetic deep chlorophyll maximum (DCM) in a stratified meromictic tropical lake and its implications for zooplankton community distribution. *International Review of Hydrobiology*, *102*(5–6), 135–150. https://doi.org/10.1002/iroh.201701899

Scofield, A. E., Watkins, J. M., Weidel, B. C., Luckey, F. J., & Rudstam, L. G. (2017). The deep chlorophyll layer in Lake Ontario: Extent, mechanisms of formation, and abiotic predictors. *Journal of Great Lakes Research*, *43*(5), 782–794. https://doi.org/10.1016/j.jglr.2017.04.003

Sepúlveda Steiner, O., Bouffard, D., & Wüest, A. (2019). Convection-diffusion competition within mixed layers of stratified natural waters. *Geophysical Research Letters*, *46*(22), 13199–13208. https://doi.org/10.1029/2019gl085361

Somavilla, R., Rodriguez, C., Lavín, A., Viloria, A., Marcos, E., & Cano, D. (2019). Atmospheric control of deep chlorophyll maximum development. *Geosciences*, *9*(4), 178. https://doi.org/10.3390/geosciences9040178

Sommer, T., Danza, F., Berg, J., Sengupta, A., Constantinescu, G., Tokyay, T., et al. (2017). Bacteria-induced mixing in natural waters. *Geophysical Research Letters*, *44*(18), 9424–9432. https://doi.org/10.1002/2017gl074868

Thomas, R. H., & Walsby, A. E. (1985). Buoyancy regulation in a strain of microcystis. *Microbiology*, *131*(4), 799–809. https://doi.org/10.1099/00221287-131-4-799

Thomas, R. H., & Walsby, A. E. (1986). The effect of temperature on recovery of buoyancy by microcystis. *Microbiology*, *132*(6), 1665–1672. https://doi.org/10.1099/00221287-132-6-1665

Ushijima, Y., & Yoshikawa, Y. (2020). Mixed layer deepening due to wind-induced shear-driven turbulence and scaling of the deepening rate in the stratified ocean. *Ocean Dynamics*, *70*, 505–512.

Wallace, B. B., Bailey, M. C., & Hamilton, D. P. (2000). Simulation of vertical position of buoyancy regulating microcystis aeruginosa in a shallow eutrophic lake. *Aquatic Sciences*, *62*(4), 320–333. https://doi.org/10.1007/pl00001338

Wilkinson, A., Hondzo, M., & Guala, M. (2020). Vertical heterogeneities of cyanobacteria and microcystin concentrations in lakes using a seasonal in situ monitoring station. *Global Ecology and Conservation*, *21*, e00838. https://doi.org/10.1016/j.gecco.2019.e00838

Woolway, R. I., Jones, I. D., Hamilton, D. P., Maberly, S. C., Muraoka, K., Read, J. S., et al. (2015). Automated calculation of surface energy fluxes with high-frequency lake buoy data. *Environmental Modelling & Software*, *70*, 191–198. https://doi.org/10.1016/j.envsoft.2015.04.013

Wu, J. (1971). An estimation of oceanic thermal-sublayer thickness. *Journal of Physical Oceanography*, *1*(4), 284–286. https://doi.org/10.1175/1520-0485(1971)001<0284:aeoots>2.0.co;2

Wu, X., Noss, C., Liu, L., & Lorke, A. (2019). Effects of small-scale turbulence at the air-water interface on microcystis surface scum formation. *Water Research*, *167*, 115091. https://doi.org/10.1016/j.watres.2019.115091

Wurtsbaugh, W. A., Paerl, H. W., & Dodds, W. K. (2019). Nutrients, eutrophication and harmful algal blooms along the freshwater to marine continuum. *Wiley Interdisciplinary Reviews: Water*, *6*(5). https://doi.org/10.1002/wat2.1373

Xiao, M., Li, M., & Reynolds, C. S. (2018). Colony formation in the cyanobacteriumMicrocystis. *Biological Reviews*, *93*(3), 1399–1420. https://doi.org/10.1111/brv.12401

Xu, W., Collingsworth, P. D., & Minsker, B. (2019). Algorithmic characterization of lake stratification and deep chlorophyll layers from depth profiling water quality data. *Water Resources Research*, *55*(5), 3815–3834. https://doi.org/10.1029/2018WR023975

Xu, Y., Cai, Q., Wang, L., Kong, L., & Li, D. (2010). Diel vertical migration of peridiniopsis niei, Liu et al., a new species of dinoflagellates in an eutrophic bay of three-gorge reservoir, china. *Aquatic Ecology*, *44*(2), 387–395. https://doi.org/10.1007/s10452-009-9298-8

Yao, B., Liu, Q., Gao, Y., & Cao, Z. (2017). Characterizing vertical migration of microcystis aeruginosa and conditions for algal bloom development based on a light-driven migration model. *Ecological Research*, *32*(6), 961–969. https://doi.org/10.1007/s11284-017-1505-9

You, J., Mallery, K., Hong, J., & Hondzo, M. (2018). Temperature effects on growth and buoyancy of microcystis aeruginosa. *Journal of Plankton Research*, *40*(1), 16–28. https://doi.org/10.1093/plankt/fbx059

Zhu, W., Feng, G., Chen, H., Wang, R., Tan, Y., & Zhao, H. (2018). Modelling the vertical migration of different-sized microcystis colonies: Coupling turbulent mixing and buoyancy regulation. *Environmental Science & Pollution Research*, *25*(30), 30339–30347. https://doi.org/10.1007/s11356-018-3041-8