

## Length–volume relationship of lake phytoplankton

Udo Mittler <sup>1</sup>, Bernd Blasius <sup>1,2</sup>, Ursula Gaedke <sup>3</sup>, Alexey B. Ryabov <sup>1\*</sup>

<sup>1</sup>Institute for Chemistry and Biology of the Marine Environment (ICBM), University of Oldenburg, Oldenburg, Germany

<sup>2</sup>Helmholtz-Institute for Functional Marine Biodiversity at the University Oldenburg (HIFMB), University of Oldenburg, Oldenburg, Germany

<sup>3</sup>Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany

### Abstract

The shapes of phytoplankton units (unicellular organisms and colonies) are extremely diverse, and no unique relationship exists between their volume,  $V$ , and longest linear dimension,  $L$ . However, an approximate scaling between these parameters can be found because the shape variations within each size class are constrained by cell physiology, grazing pressure, and optimality of resource acquisition. To determine this scaling and to test for its seasonal and interannual variation under changing environmental conditions, we performed weighted regression analysis of time-dependent length–volume relations of the phytoplankton community in large deep Lake Constance from 1979 to 1999. We show that despite a large variability in species composition, the  $V(L)$  relationship can be approximated as a power law,  $V \sim L^\alpha$ , with a scaling exponent  $\alpha = 3$  for small cells ( $L < 25 \mu\text{m}$ ) and  $\alpha = 1.7$  if the fitting is performed over the entire length range, including individual cells and colonies. The best description is provided by a transitional power function describing a regime change from a scaling exponent of 3 for small cells to an exponent of 0.4 in the range of large phytoplankton. Testing different weighted fitting approaches we show that remarkably the best prediction of the total community biovolume from measurements of  $L$  and cell density is obtained when the regression is weighted with the squares of species abundances. Our approach should also be applicable to other systems and allows converting phytoplankton length distributions (e.g., obtained with automatic monitoring such as flow cytometry) into distributions of biovolume and biovolume-related phytoplankton traits.

Phytoplankton can appear as unicellular organisms or colonies. The volume of these units is a complex function of their linear extents and shape. The volume of spherical cells increases as a cube of cell diameter (Menden-Deuer and Lesard 2000; Young and Ziveri 2000), while for flat cells it might scale as a second power and for needle-like cells as a first power (linearly) of the cell longest linear dimension (Sun and Liu 2003). Similar rules hold for phytoplankton colonies. We use for simplicity the term “cells” to address both unicellular organisms and colonies unless stated otherwise.

There are fundamental reasons which suggest a generic relation between the geometry and the size of phytoplankton cells. One reason is that both cell size and cell shape are important drivers of cell fitness through their effect on the surface to volume ratio, nutrient uptake (Karp-Boss and Boss 2016), grazing rates of zooplankton (Reynolds 2006), as well as sinking and diffusion rates (Padisák et al. 2003). Furthermore,

environmental, metabolic, and other constraints demand an adjustment of cell shape with size. Indeed, it is known that the shape of a phytoplankton cell varies systematically with its size (Niklas 2000). For instance, small cells are typically spherical, while larger cells are frequently found to be elongated. Thus, although there may be no formula which would describe a perfect one-to-one relationship between cell volume and linear dimensions in general, we can expect to find a mathematical expression which successfully approximates this dependency on average among dominant phytoplankton species.

The phytoplankton composition, and therefore also the dominant species and average cell sizes, permanently change depending on factors such as season, resource availability, and grazing pressure (Cermeño et al. 2006; Acevedo-Trejos et al. 2013). These changes may favor phytoplankton species of one specific cell shape over the others. Thus, the aggregated cell size–volume relationship in a given habitat may also change with time. Finding such a relationship has an important practical aspect: Numerous studies show that cell volume is a good predictor of cell metabolism (Marañón et al. 2012; Enquist et al. 2015) and cell abundance (Cermeño et al. 2006). Thus, identifying a relation between cell length and volume opens up

\*Correspondence: alexey.ryabov@uni-oldenburg.de

Additional Supporting Information may be found in the online version of this article.

a new perspective in determining dynamic changes in total phytoplankton biovolume and in cell volume-related phytoplankton traits, even if only information on cell length is available.

In this article, we determine the length–volume relationship for the phytoplankton community in large deep Lake Constance, using 893 samples of phytoplankton composition obtained during microscopic long-term monitoring from 1979 to 1999 (Weithoff and Gaedke 2017). The data contain time series of the abundance of 36 phytoplankton morphotypes, each characterized by a specific (time-independent) longest linear dimension and volume. The measurements were performed 2–8 times a month during each year, and thus allow to test for interannual and seasonal changes in the length–volume relationship. Finally, these results can be related to varying nutrient conditions, which have changed in Lake Constance from rather eutrophic to more oligotrophic conditions during the study period (Gaedke and Schweizer 1993; Gaedke et al. 1998).

To find an optimal fit to the relationship between the length and volume of phytoplankton cells, we compare various weighted fitting algorithms and fitting functions. As we show, most fitting methods provide a relatively good approximation to the observed length–volume relationship. However, the choice of the fitting method strongly affects the estimates of the total biovolume, calculated based on the obtained fitting functions. Thus, we rank the fitting algorithms by the accuracy for these estimates and show that the highest accuracy is obtained when the fitting uses the squares of morphotype abundances as weights. Finally, we study seasonal and interannual variations of the optimal fitting parameters for Lake Constance by comparing the median parameters calculated for each month, year, and the entire period of observations. As phytoplankton community of Lake Constance is typical for many freshwater lakes of intermediate trophic state, we expect our findings to be applicable to other freshwater systems.

## Materials and methods

### Data

We use data on abundance, average volume, and longest linear dimension of unicellular phytoplankton and colonies from 1979 to 1999 in Upper Lake Constance (<https://fred.igb-berlin.de/data/package/22>). The list of morphotypes and their cell sizes is found in Weithoff and Gaedke (2017). Upper Lake Constance (Bodensee) is a warm-monomictic, large (472 km<sup>2</sup>), and deep (water depth up to 253 m and 101 m on average) temperate lake north of the European Alps. It underwent reoligotrophication as total phosphorus concentrations, the most limiting nutrient for phytoplankton growth, declined from more than 2.8 μM P in 1979 to 0.5 μM P in 1998, resulting in a pronounced phosphorus depletion in the epilimnion during summer (Tirok and Gaedke 2006). Phytoplankton samples were taken weekly during the growing season (twice a week in 1981 and 1987, biweekly in 1999) and approximately every 2 weeks in winter. Species-specific phytoplankton cell counts and the

linear extends of each species were obtained using Utermöhl's (1958) inverted microscope technique. The cell volume was calculated by assuming a characteristic cell shape for each species. All measurements were provided per unit area and comprise the biovolume within the uppermost water layer from 0 m to 20 m depth, which roughly corresponds to the epilimnion and the euphotic zone. We considered the 36 most abundant morphotypes of phytoplankton (constituting on average 92% of total phytoplankton biomass) comprising individual species or higher taxonomic units that are morphologically similar (Rocha et al. 2011, 2012). This guaranteed a consistent resolution of phytoplankton counts during the 19 yr of sampling.

For each of the 36 morphotypes, we include in our analysis only one characteristic value for cell length and cell volume and focus on the variations in morphotypes abundance only. This means, we neglect intraspecific variability and potential temporal variations in morphotypes cell length and volume as the maximal range of these variations is typically constrained by a factor from 2 to 5. This range is much smaller than the range of interspecific variations, which extend up to 2 orders of magnitude for the longest linear dimension and 4 orders of magnitude for volume.

### Fitting procedure

For every morphotype  $i$ , let us denote  $L_i$  its cell longest linear dimension and  $V_i$  its volume, which we regard to be constant in time. As a first order approximation to the size–volume relationship, we apply a power law function

$$V(L) = V_0 L^\alpha \quad (1)$$

which on a log-log scale describes a straight line with slope  $\alpha$  and intercept  $\log V_0$ . Fitting this line on a log-log scale leads, however, to an uneven distribution of the residuals and systematic errors in the range of either small or large cells. To obtain a second-order approximation, we tested various different fitting functions and obtained the best fit to the empirical results with a transitional power function

$$V(L) = \frac{V_0 L^\alpha}{(L/L_0)^{\alpha-\beta} + 1}, \quad (2)$$

which smoothly changes its slope in a double-logarithmic plot from  $\alpha$  to  $\beta$  with increasing  $L$ . Thus, this function exhibits two distinct scaling regimes for small and large cells:

$$V(L) \approx V_0 L^\alpha, L \rightarrow 0 \quad \text{and} \quad V(L) \approx V_0 L_0^{\alpha-\beta} L^\beta, L \rightarrow \infty. \quad (3)$$

These two regimes intersect at the characteristic size  $L_0$ , which can be interpreted as a midpoint of the transition from one scaling exponent to the other.

To account for the difference in abundance among morphotypes, for each time instance  $t$ , we apply weighted regression analysis using time-dependent weights  $W_{i,t}$ . The fitting can,

however, be performed on a linear or logarithmic scale, and various definitions of weights can be used. Generally, for each sample, we need to find such time-dependent parameters  $\alpha_t$ ,  $V_{0,t}$  (and  $\beta_t$  for Eq. 2) that minimize the weighted mean square deviation between the measured and predicted cell volume for given cell longest linear dimensions. The mean square deviation can be calculated either on a normal or logarithmic scale:

$$F_{\text{lin}}(\alpha_n, V_{0,t}) = \sum_i W_{i,t} (V_i - V(L_i, \alpha_t, V_{0,t}))^2$$

$$F_{\text{log}}(\alpha_n, V_{0,t}) = \sum_i W_{i,t} (\log V_i - \log V(L_i, \alpha_n, V_{0,t}))^2$$

The choice of the weights and scale is crucial, as it strongly affects the best fit parameters and the prediction accuracy. For instance, fitting data on a logarithmic scale reduces the relative residuals across all size classes, while fitting on a linear scale minimizes the absolute residuals. Similarly, different definitions of weights affect to a greater or lesser extent the accuracy obtained for dominant morphotypes at the expense of the accuracy for rare morphotypes.

One possibility to define the weights and scale is to require that the resulting fitting function provides an unbiased prediction for the biovolume of the whole community. The measured and predicted total biovolume of the whole community at each time instance  $t$  can be expressed in terms of morphotype abundance  $A_{i,t}$  as

$$B_{\text{meas},t} = \sum_i A_{i,t} V_i \quad (4)$$

$$B_{\text{pred},t} = \sum_i A_{i,t} V(L_i, \alpha_t, V_{0,t})$$

Their difference equals

$$B_{\text{meas},t} - B_{\text{pred},t} = \sum_i A_{i,t} [V_i - V(L_i, \alpha_t, V_{0,t})].$$

The best agreement between predicted and measured biovolumes arises if the absolute value of each term in the sum approaches zero. We can achieve this by minimizing the sum of squares of these terms.

$$F_{\text{AW}}(\alpha_t, V_{0,t}) = \sum_i A_{i,t}^2 [V_i - V(L_i, \alpha_t, V_{0,t})]^2. \quad (5)$$

Thus, we need to minimize the mean square deviation from the cell volumes on a linear scale using the squares of morphotypes abundances as weights. Below we refer to this method as abundance-weighted (AW) fitting.

The differences in Eq. 5 can be written as  $[A_{i,t} V_i - A_{i,t} V(L_i, \alpha_t, V_{0,t})]^2$ , where the first product expresses the observed biovolume of morphotype  $i$  and the second product is the predicted biovolume of this morphotype (see Eq. 4). Thus, the AW-fitting

can be reinterpreted as a (nonweighted) least square regression to the total biovolume of each phytoplankton morphotype. As a consequence, the sensitivity of AW-fitting to variations in morphotype composition is much smaller than it might seem at first glance. While both the cell abundance  $A_{i,t}$  and volume  $V_i$  vary over many orders of magnitude separately, the variability of their product (i.e., the biovolume) is much smaller because cell densities tend to decrease with increasing cell volume (Mullin et al. 1966; Kruk et al. 2010).

To additionally test for the accuracy of the AW fitting, we compare it against two other fitting approaches, where we perform the weighted regression on a logarithmic scale, assuming that weights  $W_{i,t}$  either equal 1(0) for present (absent) morphotypes, or are proportional to morphotype biovolumes,  $B_i = A_i V_i$ . Below we refer to these methods as presence/absence (PA) and biovolume-weighted (BW) fitting. These approaches seem to be intuitively more robust than AW-fitting, but, as we show below, they lead to biased and less accurate estimations of the total biovolume.

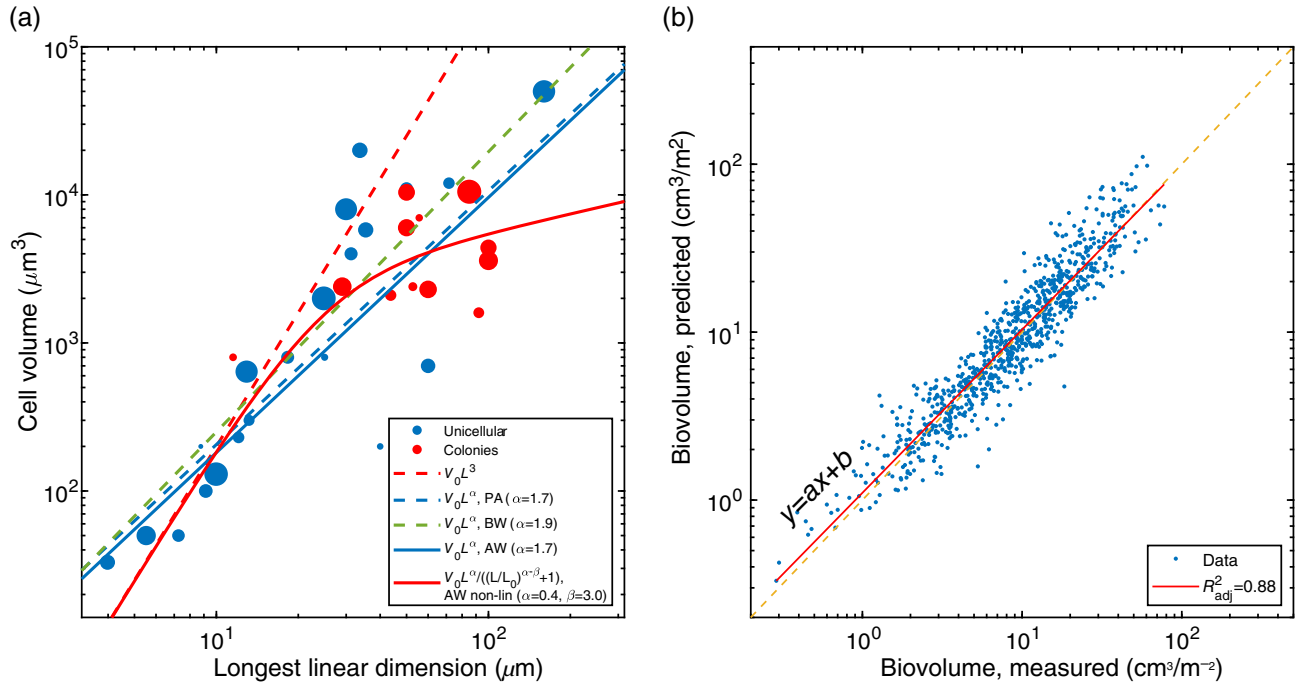
To evaluate the accuracy of the methods, we express the match between the measured and predicted total biovolume through the coefficient of determination  $R_{\text{adj}}^2$  and the coefficients of the linear regression  $y = ax + b$  between these quantities (see an example in Fig. 1b). The coefficient of determination characterizes the scatter of the data points around the diagonal (red dashed line, Fig. 1b), and the linear regression parameters show the tilt and vertical shift of the cloud of data points with respect to the diagonal, thus, indicating a systematic bias in the prediction. In an ideal case, all data points should collapse on the diagonal, so that  $R_{\text{adj}}^2 = 1$ ,  $a = 1$ , and  $b = 0$ .

To find the best fit parameters, we use the Levenberg-Marquardt weighted nonlinear least squares algorithm (fitnlm function in MATLAB 2016). The best fit parameters obtained for the community observed at time  $t$  are subsequently referred to as instantaneous parameters ( $V_{0,t}$ ,  $\alpha_t$ , etc.). These parameters carry the information about the community composition for a given date and provide the closest match between the modeled length–volume relationship and empirical data for each time instance (see Supporting Information Fig. S1 for an example). We further test for the seasonal dependence and interannual trends of the best-fit parameters by calculating their median values for each calendar month irrespective of the year (denoted as  $\tilde{\alpha}_m$ ,  $\tilde{V}_{0,m}$ , etc.), and annual median values ( $\tilde{\alpha}_y$ ,  $\tilde{V}_{0,y}$ , etc.). Finally, we determine the overall median values ( $\tilde{V}_0$ ,  $\tilde{\alpha}$ , etc.) over the entire period.

## Results

### Ranking the fitting methods

For the phytoplankton morphotypes observed in Lake Constance, the longest linear dimension ranges over 2 orders of magnitude and the volume over almost 4 orders of magnitude



**Fig. 1.** Fitting and evaluating the length–volume relationship. **(a)** Longest linear dimension  $L$  and volume  $V$  of unicellular phytoplankton (blue circulars) and colonies (red circulars), the area of the circulars is proportional to the morphotype time-averaged biovolume. Approximated scaling relationships plotted for the overall median parameters obtained with various fitting methods (lines, see Table 2 for parameters). The volume of small cells ( $L < 25 \mu\text{m}$ ) closely follows a cubic dependence on  $L$  (red dotted line). Fitting a power law (Eq. 1) over the entire range gives a slope  $\alpha$  from 1.7 to 1.9, in dependence on the method applied (blue dashed line – PA, green dashed line – BW, and blue solid line – AW), see Fig. 2 for the comparison of these results. The transitional power function (Eq. 2) provides the best fit over the entire range of cell length (red line is based on AW-fitting). **(b)** An example of the comparison of the measured total biovolume and total biovolume predicted by Eq. 4 for the  $V(L)$  relationship obtained with PA-fitting. In an ideal case, for every sample, the predicted total biovolume should equal the measured biovolume, and all data points should collapse on the diagonal (red dashed line). The deviation of data points from the diagonal is characterized by  $R_{\text{adj}}^2$  and by the linear regression parameters (solid line)  $a$ , indicating the tilt of the cloud of points with respect to the diagonal, and  $b$ , showing the bias along the vertical axis. In an ideal situation,  $R_{\text{adj}}^2 = 1$ ,  $a = 1$ , and  $b = 0$ .

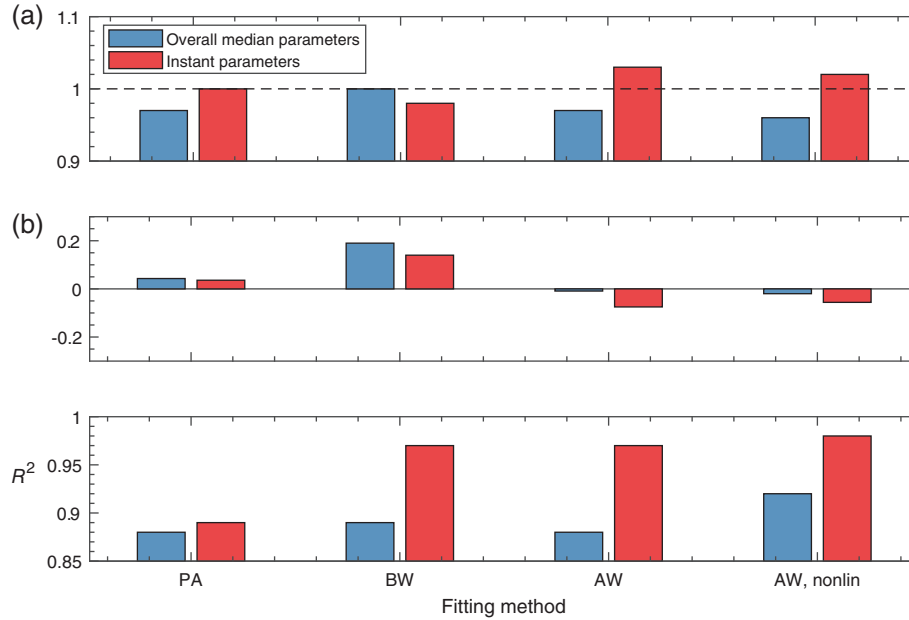
(Fig. 1a). To differentiate between common and rare morphotypes, the size of the data points in Fig. 1 indicates the morphotype biovolume averaged over the monitoring period. The longest linear dimension of most unicellular phytoplankton morphotypes (blue dots) is below  $40 \mu\text{m}$ , and all colonies are above this limit (red dots). We first focus on the comparison of  $L(V)$  relationships obtained with PA-weighted, BW, and squared AW fittings.

Fitting a power law over the entire range of cell lengths and volumes gives an overall median slope  $\bar{\alpha} = 1.73$  if the PA and AW-methods are applied (Fig. 1a, blue dashed and blue solid lines) and  $\bar{\alpha} = 1.9$  when BW-fitting is used (green dashed line). In all cases, we obtain a good visual match between the regression lines and data points. However, the difference in the results becomes evident when the obtained  $L(V)$  relationships are used to predict the total biovolume (see Eq. 4).

Compare, for instance, the measured biovolumes with biovolumes calculated based on the overall median parameters obtained with PA-method (Fig. 1b). Using this simple approach, we obtain a relatively good prediction of the total biovolume ( $R_{\text{adj}}^2 = 0.88$ ), and this prediction is unbiased ( $b = 0.043 \pm 0.012$  and  $a = 0.97 \pm 0.01$ ). However, the accuracy

of this prediction does not increase if instead of the median parameters we use their instantaneous values  $\alpha_t$  and  $V_{0,t}$  ( $R_{\text{adj}}^2 = 0.89$ , Table 1). Thus, the presence-absence fit is not sensitive enough to capture the current state of the community.

A comparison of the results obtained using PA, AW, and BW methods demonstrates the advantages of AW fitting (Fig. 2). First, only with AW and BW methods, we obtain a substantially better prediction of the total biovolume, when instantaneous parameters instead of the overall median values are used as indicated by an increase of  $R_{\text{adj}}^2$  (the bottom panel in Fig. 2 and Table 1). Second, only AW and PA methods provide an unbiased estimation of the total biovolume, as is evident from the low value of the linear regression parameter  $b$  (Fig. 2 middle panel). However, if the BW-method is applied, the linear regression gives  $\log B_{\text{pred}} = 0.98 \log B_{\text{meas}} + 0.14$ , meaning that the total biovolume is on average overestimated by a factor of  $10^{0.14} = 38\%$  (Supporting Information Fig. S2). Finally, note that the slope  $a$  of the linear regression is always around 1, indicating no tilt of the cloud of data points with respect to the diagonal in all cases (Fig. 2, top panel). Thus, only the fitting weighted by the squares of morphotype abundances



**Fig. 2.** Evaluation of  $L(V)$  relationships obtained for PA-weighted, BW, and squared AW fittings. The evaluation is done for overall median parameters (blue bars) and instantaneous best-fit parameters (red bars). (Top) All methods lead to relatively low tilt,  $a$ , of the data points. (Middle) All methods except of BW lead only to a small bias,  $b$ , in the estimation of the total biovolume. (Bottom) For all methods except of PA, instantaneous parameters in comparison with median parameters provide better estimates (greater  $R^2_{\text{adj}}$ ) for the total biovolume. See Table 1 for the values.

**Table 1.** Linear regression between the predicted and measured biovolume ( $\log B_{\text{pred}} = a \log B_{\text{meas}} + b$ ). The predicted biomass is calculated based either on instantaneous parameters  $\alpha_t$  and  $V_{0,t}$  or on overall median parameters.  $p$  values are calculated for a  $t$ -statistic test that the coefficient is zero.

		Instantaneous parameters		Overall median parameters	
		Estimate	$p$ value	Estimate	$p$ value
Power law, log scale, PA	$a$	$1.00 \pm 0.01$	$<10^{-6}$	$0.97 \pm 0.01$	$<10^{-6}$
	$b$	$0.036 \pm 0.013$	0.005	$0.043 \pm 0.012$	0.0005
	$R^2_{\text{adj}}$	0.89	—	0.88	—
Power law, log scale, BW	$a$	$0.98 \pm 0.01$	$<10^{-6}$	$1.00 \pm 0.01$	$<10^{-6}$
	$b$	$0.14 \pm 0.006$	$<10^{-6}$	$0.19 \pm 0.01$	$<10^{-6}$
	$R^2_{\text{adj}}$	0.97	—	0.89	—
Power law, linear scale, AW	$a$	$1.03 \pm 0.007$	$<10^{-6}$	$0.97 \pm 0.01$	$<10^{-6}$
	$b$	$-0.075 \pm 0.007$	$<10^{-6}$	$-0.009 \pm 0.012$	0.475
	$R^2_{\text{adj}}$	0.97	—	0.88	—
Power law, linear scale AW $2 \leq L \leq 25 \mu\text{m}$	$a$	$1.00 \pm 0.002$	$<10^{-6}$	$0.99 \pm 0.004$	$<10^{-6}$
	$b$	$-0.02 \pm 0.001$	$<10^{-6}$	$-0.033 \pm 0.003$	$<10^{-6}$
	$R^2_{\text{adj}}$	0.998	—	0.987	—
Transitional power function, linear scale, AW	$a$	$1.02 \pm 0.01$	$<10^{-6}$	$0.96 \pm 0.01$	$<10^{-6}$
	$b$	$-0.056 \pm 0.005$	$<10^{-6}$	$-0.020 \pm 0.010$	0.04
	$R^2_{\text{adj}}$	0.98	—	0.92	—

(AW method) gives best-fit parameters, which provide an unbiased estimation of the total biomass and allow to take advantage of the sample data if instantaneous parameters are used. Consequently, we include only this method in the further analysis.

As shown in Fig. 1a, the cloud of data points has a concave up shape. However, in the range of small cells ( $L < 25 \mu\text{m}$ ), the dependence is approximately linear on a logarithmic scale, and fitting of a power law gives the overall median slope

**Table 2.** The median values and 25–75% interquartile range (in brackets) of the fitting parameters calculated over the entire period for different fit functions and weights.

	Function	$\alpha$	$\beta$	$V_0$	$L_0$
Power law, log scale, PA	$\log V = \log V_0 + \alpha \log L$	1.73 (1.64, 1.79)	—	3.92 (3.27, 4.98)	—
Power law, log scale, BW	$\log V = \log V_0 + \alpha \log L$	1.85 (1.64, 2.14)	—	3.07 (1.54, 6.10)	—
Power law, linear scale, AW	$V = V_0 L^\alpha$	1.73 (1.47, 2.02)	—	3.42 (1.56, 5.55)	—
Power law, linear scale, AW $2 \leq L \leq 25 \mu\text{m}$	$V = V_0 L^\alpha$	2.97 (2.48, 3.00)	—	0.14 (0.13, 0.62)	—
Transitional power function, linear scale, AW	$V = \frac{V_0 L^\alpha}{(L/L_0)^{\alpha-\beta} + 1}$	3.00	0.42 (0.03, 1.10)	0.20 (0.16, 0.24)	25 (24, 26)

$\tilde{\alpha} = 2.97$  (Fig. 1a, red dotted line; Table 2) with relatively small variations of  $\alpha_t$  among samples. Thus, in this range of cell length, we can neglect the seasonal dependence of the fitting parameters and assume that the cell volume scales approximately as a cube of cell length.

Fitting the nonlinear power function requires the definition of four parameters (see Eq. 2). However, these parameters are interdependent and, for instance, an increase of  $L_0$  can be compensated by decreasing  $\alpha$ . This results in a large variability and low significance of the best-fit values. To constrain the parameter variability, we assume that the initial slope  $\tilde{\alpha} = 2.97$ , as shown before. We also assume that the midpoint-parameter  $L_0$  is the same for all samples and only  $\beta_t$  and  $V_{0,t}$  change among the samples. We also assume that the cell volume on average increases with cell length, and therefore,  $\beta_t > 0$ . We tested different  $L_0$  and found the best match between the predicted and measured biovolume for  $L_0 = 25 \pm 1 \mu\text{m}$ , with a vanishing dependence of the predicted total biovolume on the value of  $L_0$ . Thus, we chose  $L_0 = 25 \mu\text{m}$ , which corresponds to the upper border of the small cell range considered above (see Table 2 for the values of the other parameters). The resulting function with a median final slope of  $\tilde{\beta} = 0.4$  reflects better the nonlinear shape of the cloud of data points (Fig. 1a, red line), and allows for an unbiased estimation of the total biomass. The comparison of the measured and predicted biomasses gives  $R_{\text{adj}}^2 = 0.92$  if the overall median parameters are used and shows an excellent agreement ( $R_{\text{adj}}^2 = 0.98$ ) if the instant parameters are used (Fig. 2; Table 1).

### Seasonal and interannual dynamics

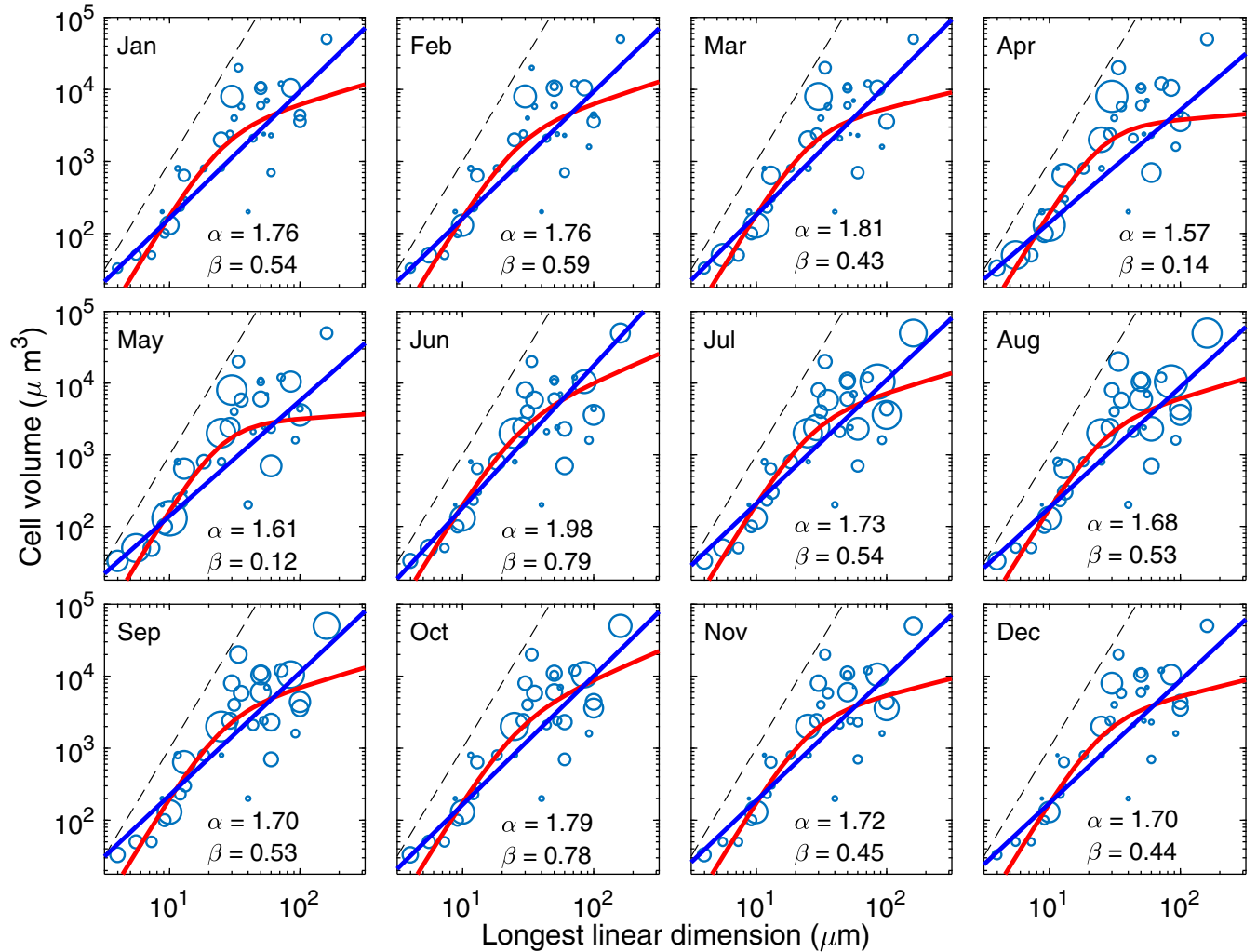
To gain an idea about the seasonal dependence of  $L(V)$  relationships, consider the fitting functions plotted for the median parameters calculated for each calendar month without regard to the year of sampling (Fig. 3). For most calendar months, the  $L(V)$  relationship varies little: the power law has a median slope  $\tilde{\alpha}_m \approx 1.7$  (blue lines) and the nonlinear power function (red lines) has the final slope  $\tilde{\beta}_m \approx 0.5$ . Only in April and May, the exponents drop for both models below these stationary levels, and in June and October rise above them. However, in spite of the significance of these changes, they are relatively small compared to the random variability among

samples obtained in different years (see Fig. 4 for more details). Applying the monthly median parameters instead of the overall medians does not provide any increase of the accuracy of the total biovolume estimates (details not shown).

To find potential long-term trends in  $V(L)$  relationships, we performed a linear regression of the instantaneous values  $\alpha_t$  and  $\beta_t$  grouped by months (indicated by the colored dots and colored lines in Fig. 5) and of all data points (black lines). The  $p$  values show the probability that the null-model with no positive or negative trend holds. The regression of monthly aggregated data gives in most cases insignificant trends. However, most trends are positive (for 9 out of 12 calendar months) and for three months, the slope  $\alpha_t$  shows a significant increase during the observation period (shown by solid regression lines). A superposition of these trends leads to a significant ( $p \leq 0.01$ ) positive interannual trend in the instant slopes  $\alpha_t$  of the power law, indicating a significant increase in the steepness of the  $V(L)$  relationship with time, which is however small in absolute terms (Fig. 5a black line). For the nonlinear power function, we did not find any significant changes in the final slope  $\beta_t$  (Fig. 5b).

### Discussion

The longest linear dimension,  $L$ , and volume,  $V$ , of phytoplankton cells and colonies vary over a wide range. To better understand how  $V$ , which defines the total biomass, scales with  $L$ , we performed a weighted regression analysis of long-term phytoplankton data obtained in Lake Constance. We used various fitting approaches and ranked the obtained  $V(L)$  dependencies by the accuracy with which these dependencies allow to estimate the total biovolume based on observed cell length distributions. We found that in the entire range of cell sizes the  $V(L)$  dependence can be approximated by a power law with slope  $\alpha \approx 1.7$ , or by a transitional power function which changes its slope from 3 in the range of small cells ( $< 25 \mu\text{m}$ ) to 0.4 in the range of large cells and colonies ( $> 25 \mu\text{m}$ ). Both fitting functions provide good estimates of the total biovolume even based on the overall median parameters, which neglect the between-sample variability (power law:  $R_{\text{adj}}^2 = 0.88$ ; transitional power function:  $R_{\text{adj}}^2 = 0.92$ ). The application of instantaneous best-fit parameters (obtained for each

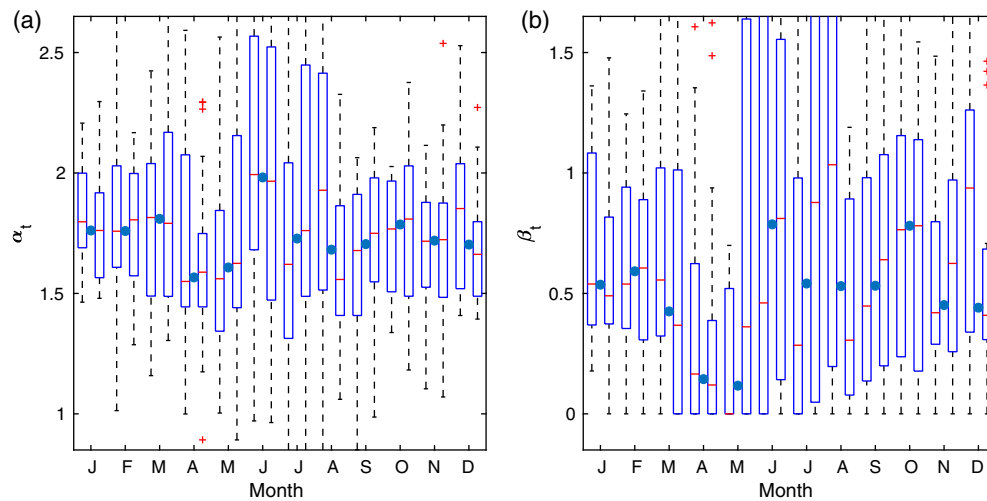


**Fig. 3.** Scaling relationships between length and volume of phytoplankton morphotypes for each calendar month averaged over the entire observation period. Different morphotypes are shown by blue circulars with area proportional to the morphotype average biovolume for a given month. The median slope  $\alpha$  of a power function (blue line) and the final slope  $\beta$  of a nonlinear power function (red line) drop in April and May and rise in June and October. During the other months, the scaling functions vary only little. The black dashed line is a reference line with a slope of 3.

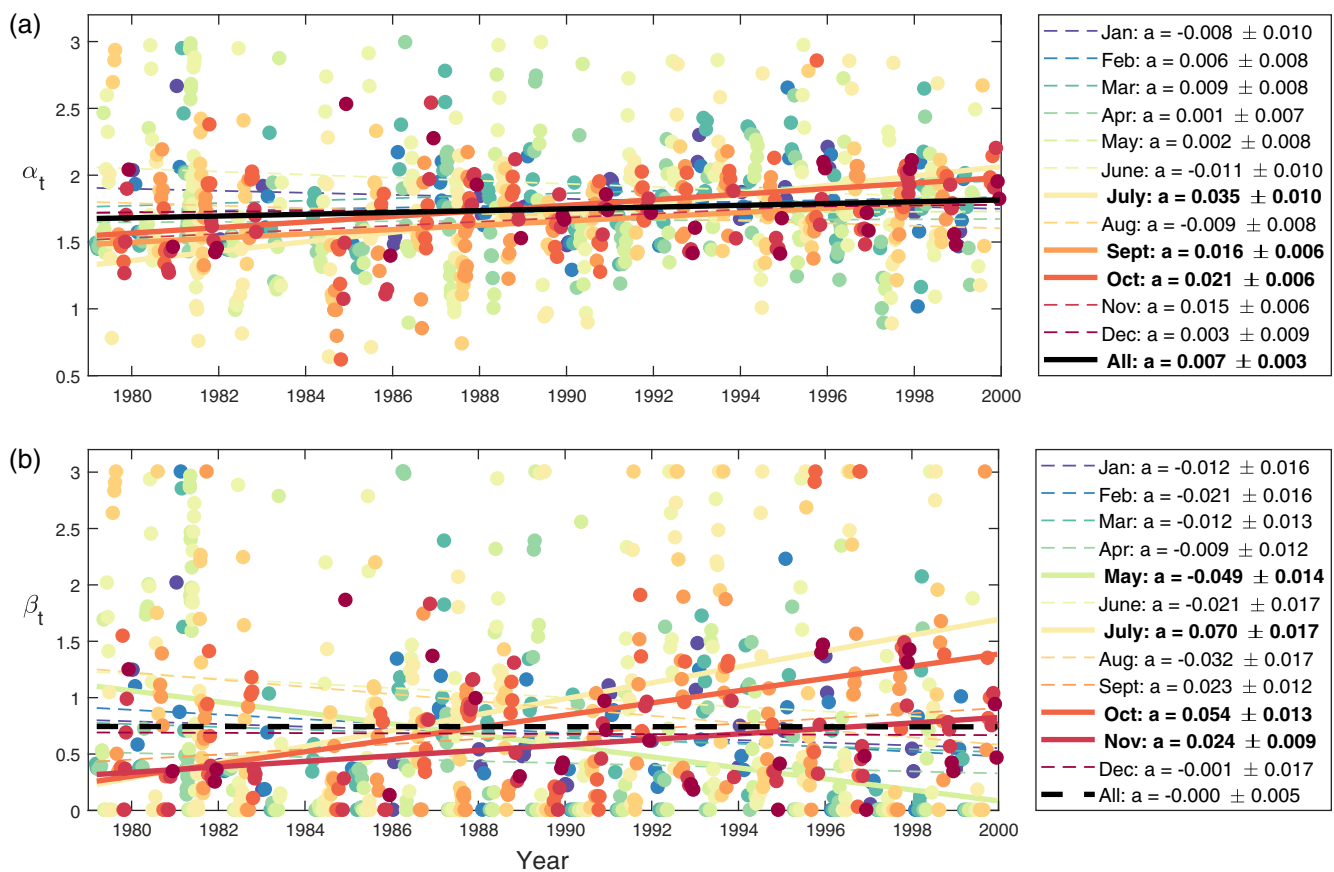
sample) further improves these estimates and leads to an excellent match between the measured and predicted biovolume with  $R_{\text{adj}}^2 = 0.97$  for the power law and  $R_{\text{adj}}^2 = 0.98$  for the transitional power function. Finally, note that for the ecologically most relevant range with  $L < 25 \mu\text{m}$  comprising highly productive and well edible morphotypes, we found an approximately cubic scaling of volume with cell length which provides a nearly perfect estimation of total biomass for this range with  $R_{\text{adj}}^2 = 0.998$ .

What is the biological meaning of the obtained slope  $\alpha$  and intercept  $V_0$  of a power law,  $V(L) = V_0 L^\alpha$ ? It is important to stress that the slope does not indicate the most common shapes or dimensionality of cells. That is, a value of  $\alpha = 2$  does not mean that the cells are flat, just as  $\alpha = 3$  does not mean the cells are spherical or cubic. The meaning of  $\alpha$  can be easily seen in the range of small cells. The cubic slope in this range

indicates that phytoplankton cells scale approximately isometrically with cell length, i.e., we should find approximately the same proportional abundance of spherical, cylindrical, needle-like, and so on, cells in each length class (see Appendix for detail). By contrast, a smaller slope  $\alpha$  (as found in the range of large cells and cell colonies) indicates that, with increasing longest linear dimension and assuming isometric scaling (i.e., keeping the geometry and linear proportions of each morphotype), the shape distribution monotonically shifts toward the dominance of complex and elongated structures. This can occur because, in particular for large cells, a prolate shape may be more advantageous than a spherical, disk, or oblate shape, as it enhances, e.g., the nutrient transport due to an increased surface to volume ratio (Karp-Boss and Boss 2016). Because the slope  $\alpha$  is small in the range of large cells, we conclude that with increasing cell length the fraction of elongated phytoplankton units monotonically increases.



**Fig. 4.** Seasonal variations in the instantaneous (measured for each sample) best-fit exponent  $\alpha_t$  of a power law function (Eq. 1) and the instantaneous final exponent  $\beta_t$  of a nonlinear power function (Eq. 2). For a better representation, the values are grouped in a boxplot by each two calendar weeks. The boxes include 50% of data points and show the 25% and 75% quantiles across all values in the given group, and the whiskers extend to the most extreme value not considered as an outlier (red cross). The blue dots show monthly median values (cf. Fig. 3).



**Fig. 5.** Trends of the exponent  $\alpha_t$  (a power law function) and the final exponent  $\beta_t$  (a transitional power function) for the entire monitoring period. The instantaneous values of the exponents are grouped by calendar months (dots of various colors). The linear regression through the instantaneous parameters calculated for different months (lines of various colors) thick solid lines and bold font in the legend indicate significant positive or negative trends ( $p < 0.01$ ). The linear regressions of all data points (black thick lines) show a significant increase of the exponents  $\alpha_t$  and no interannual change of the median exponent  $\beta_t$ .



In contrast to  $\alpha$ , the intercept  $V_0$  might give an idea about an average cell shape as it shows the average volume of cells with  $L = 1 \mu\text{m}$ . Fitting the power function  $V = V_0 L^\alpha$  in the range of small cells, we found  $V_0 = 0.14 \mu\text{m}^3$  (Table 2). This volume is less than the volume of a sphere with diameter of  $1 \mu\text{m}$  ( $V_{\text{sphere}} = 0.52 \mu\text{m}^3$ ). If we assume that the cells are prolate (the longest dimension is larger than the other two), then we obtain an approximately equal volume for a cylinder with 1 : 0.4 ratio of height to diameter or a prolate spheroid (a stretched sphere) with an aspect ratio of 1 : 0.5. Thus, the dominating cells in this range have an aspect ratio of around 2. Note that this aspect ratio matches with one of two peaks in the aspect ratio distribution found by Gibson et al. (2007), for freshwater phytoplankton from five different classes including diatoms, chlorophyceae, cryptophytes, dinoflagellates, and cyanobacteria.

We have shown that various fitting methods can provide a good visual match between data points and fitting curves (Fig. 1a), but small variations in the best-fit parameters caused by changes in weights or fitting scales can lead to large systematic errors in the calculation of an integral quantity such as total biovolume. How should one select the most appropriate fitting method? This decision depends on the purpose for which the obtained  $V(L)$  dependence is used. Here, we focused on calculating the total biovolume and showed that the best unbiased estimates are obtained, if the fitting is performed on a linear scale and weights equal to the squares of morphotype abundances (AW fitting). Contrasting these results with other fitting approaches we found that regression on a log-log scale weighted by morphotype biovolumes (BW fitting) results in a systematic overestimation of the total biovolume. Unexpectedly, even the regression weighted by PA data (PA-fitting) provides better results. Although such PA-fitting is insensitive to fluctuations in morphotype composition and does not provide an advantage when instantaneous parameters are used instead of their median values, the overall best-fit parameters obtained with PA-fitting are close to those obtained with AW-fitting.

At first, it might seem that the AW fitting (Eq. 5) is overly sensitive to changes in morphotype composition, as this procedure uses cell volumes (which vary over 4 orders of magnitude), cell lengths (2 orders of magnitude), and squares of morphotype abundance (up to 14 orders of magnitude). Intuitively, fitting a line to logarithms of cell dimensions and using morphotype biovolume as weights seems to be a more reliable approach. However, as explained in the “Materials and methods” section, the actual variability in Eq. 5 is not as great as it might seem since cell volumes, lengths, and abundances are not independent. In fact, it has been often observed that densities of phytoplankton cells tend to decrease approximately proportionally with cell volume (Mullin et al. 1966; Gaedke 1992; Kruk et al. 2010). As a consequence, the variability of biovolumes is much smaller than that of cell volumes and abundances separately—thus rendering the AW fitting, which minimizes the differences between the observed and

predicted total biovolumes, quite robust. This also explains why the AW fitting gives the best estimate for the total biovolume of the community.

Note that our approach neglects the variability of cell volumes within each length class. This variability can be caused either by interspecific differences when, for instance, two species with the same length have different shapes and, therefore, different volumes, or by intraspecific differences when individuals of the same species fall into different size classes. A further improvement in defining the  $V(L)$  relationship can be achieved by including the volume variability into the variance of a probability density function providing the probability to obtain a certain volume for a given length. Although we do not anticipate a substantial increase in the accuracy of the total biomass calculated based on such a function, this approach might be beneficial for estimating a cell volume distribution based on the distribution of the cells longest linear dimensions.

The periodic seasonal oscillations and monotonic interannual trends in the  $V(L)$  relationship were small compared to the random between-sample variability in the best-fit parameters. During the calendar year, the  $V(L)$  relationship was relatively stable and only deviated from this stationary state in April, May, June, and to a lesser extent in October. Thus, these deviations occurred only for 3–4 of 12 months in a year and were relatively small in comparison with interannual fluctuations in the phytoplankton community (see an example in Supporting Information Fig. S1). This likely explains the negligible increase in the accuracy of the total biovolume estimates if the seasonal variation is taken into account.

Regarding changes in the  $V(L)$  relationship over longer time spans, we found a positive interannual trend of the scaling exponent  $\alpha$  in the 20 yr monitoring period (Fig. 5). When the scaling exponents were aggregated by months, we saw in most cases a positive but insignificant trend. This led, however, to a significant positive trend in the entire time series of  $\alpha_t$ , which increased on average from approximately 1.67 to 1.8 during the observation period. Further data analysis revealed that this increase was caused by a significant positive trend in average cell volumes during the observation time, whereas we neither found an increasing nor a decreasing trend in the average length of cells. As the distribution of small cells was relatively stable, this increase in the average volume and in the slope of  $V(L)$  relationship likely arose from a shift in the shape distribution in the range of large cells ( $L > 25 \mu\text{m}$ ) from more elongated cells with relatively small volumes to more spherical cells with larger volume and smaller aspect ratio. To conclude, we established scaling laws between phytoplankton cell volume and longest linear dimension delivering an excellent agreement between the measured and predicted total biovolume ( $R^2$  up to 0.998). We anticipate that our results are applicable to other freshwater systems for the following reasons. Lake Constance has a phytoplankton community typical for

many freshwater lakes of intermediate trophic state (Sommer et al. 1986). It exhibits a pronounced seasonality and the phosphorous concentrations decreased threefold during the monitoring period, resulting in distinct seasonal and long-term taxonomic changes (Weithoff and Gaedke 2017). Nevertheless, our analysis did not reveal any strong seasonal signal and the fitting based on the time-averaged median parameters provided already very good estimates over the entire period of observations. This suggests that the established relationship between cell volume and longest linear dimension can be directly transferred to roughly comparable systems. Furthermore, our detailed comparison and mechanistic understanding of the performance of the different fitting procedures can be used to establish corresponding relationships for phytoplankton communities with strongly deviating properties.

Our proposed method of establishing a relation between phytoplankton cell length and volume has important practical aspects. This is due to the different roles played by these two quantities. While cell volume is frequently taken as a predictor of cell metabolism (Marañón et al. 2012; Enquist et al. 2015) and cell abundance (Cermeño et al. 2006), its direct microscopic measurements of cell volume are complex and time-consuming (Olenina 2006; Weithoff and Gaedke 2017). In contrast, cell longest linear dimension, even though it does not have similar straightforward ecological and physiological implications, can be determined via automatic measurement techniques. One prominent example is flow cytometry, which can provide a high spatio-temporal resolution and allows for detecting rare species (Pomati et al. 2013). Thus, having established a relation between cell longest linear dimension and volume opens up a new perspective in estimating dynamic changes in total phytoplankton biovolume and in cell volume related phytoplankton traits based on automatic in situ monitoring.

## Appendix

### Isometric scaling

Here, we show that an isometric scaling of a set of shapes leads to a cubic relationship between length and volume, if the shape distribution (i.e., the proportion of morphotypes with different geometries) is preserved. Consider a simple example assuming that for the smallest length class the shape distribution includes  $p_1$  cylinders with height  $h$  and radius  $r_1$  and  $p_2$  balls with radius  $r_2$ , and all other length classes are obtained by isometric scaling of these shapes with the factor  $\alpha$ . For the smallest length class, the volume of cylinders will be  $V_{c,0} = p_1 h \pi r_1^2$  and the volume of spheres  $V_{sph,0} = p_2 4/3 \pi r_2^3$ , leading to an average volume of  $V_{t,0} = (p_1 h \pi r_1^2 + p_2 4/3 \pi r_2^3) / (p_1 + p_2)$ . Applying isometric scaling, the dimensions of other length classes will be  $H = \alpha h$ ,  $R_1 = \alpha r_1$ ,  $R_2 = \alpha r_2$ , giving rise to the average volume of  $V_t = (p_1 H \pi R_1^2 + p_2 4/3 \pi R_2^3) / (p_1 + p_2) = (p_1 \alpha^3 h \pi r_1^2 + p_2 4/3 \pi \alpha^3 r_2^3) / (p_1 + p_2) = \alpha^3 r_2^3 \pi (p_1 \frac{hr_1^2}{r_2^3} + p_2 \frac{4}{3}) / (p_1 + p_2)$ . Thus, the average volume of this set of shapes scales with the cube

of the class characteristic size  $\alpha r_2$ , unless  $p_1$  and  $p_2$  also depend on  $\alpha$ . The proportions of cylinders and spheres as well as the cylinders aspect ratio will change only the factor  $(p_1 \frac{hr_1^2}{r_2^3} + p_2 \frac{4}{3}) / (p_1 + p_2)$ , which defines the intercept of the power law in double logarithmic axes. This derivation can be easily extended to include other shapes and an arbitrary number of shapes.

## References

- Acevedo-Trejos, E., G. Brandt, A. Merico, and S. L. Smith. 2013. Biogeographical patterns of phytoplankton community size structure in the oceans: Phytoplankton biogeographical patterns. *Glob. Ecol. Biogeogr.* **22**: 1060–1070. doi:10.1111/geb.12071
- Cermeño, P., E. Marañón, D. Harbour, and R. P. Harris. 2006. Invariant scaling of phytoplankton abundance and cell size in contrasting marine environments. *Ecol. Lett.* **9**: 1210–1215. doi:10.1111/j.1461-0248.2006.00973.x
- Enquist, B. J., J. Norberg, S. P. Bonser, C. Violle, C. T. Webb, A. Henderson, L. L. Sloat, and V. M. Savage. 2015. Chapter nine-scaling from traits to ecosystems: Developing a general trait driver theory via integrating trait-based and metabolic scaling theories. *Adv. Ecol. Res.* **52**: 249–318. doi:10.1016/bs.aacr.2015.02.001
- Gaedke, U. 1992. The size distribution of plankton biomass in a large lake and its seasonal variability. *Limnol. Oceanogr.* **37**: 1202–1220. doi:10.4319/lo.1992.37.6.1202
- Gaedke, U., and A. Schweizer. 1993. The first decade of oligotrophication in Lake Constance. *Oecologia* **93**: 268–275. doi:10.1007/BF00317681
- Gaedke, U., D. Ollinger, E. Bäuerle, and D. Straile. 1998. The impact of the interannual variability in hydrodynamic conditions on the plankton development in Lake Constance in spring and summer. *Adv. Limnol.* **53**: 565–585.
- Gibson, R. N., R. J. A. Atkinson, and J. D. M. Gordon. 2007. Inherent optical properties of non-spherical marine-like particles—from theory to observation. *Oceanogr. Mar. Biol. Annu. Rev.* **45**: 1–38. doi:10.1201/9781420050943
- Karp-Boss, L., and E. Boss. 2016. The elongated, the squat and the spherical: Selective pressures for phytoplankton shape, p. 25–34. *In*, Aquatic microbial ecology and biogeochemistry: A dual perspective. Glibert, P. M., and T. M. Kana [eds.], Springer. doi:10.1007/978-3-319-30259-1\_3
- Kruk, C., V. L. Huszar, E. T. Peeters, S. Bonilla, L. Costa, M. Lüring, C. S. Reynolds, and M. Scheffer. 2010. A morphological classification capturing functional variation in phytoplankton. *Freshw. Biol.* **55**: 614–627. doi:10.1111/j.1365-2427.2009.02298.x
- Marañón, E., P. Cermeño, D. C. López-Sandoval, T. Rodríguez-Ramos, C. Sobrino, M. Huete-Ortega, J. M. Blanco, and J. Rodríguez. 2012. Unimodal size scaling of phytoplankton growth and the size dependence of nutrient

- uptake and use. *Ecol. Lett.* **16**: 371–379. doi:[10.1111/ele.12052](https://doi.org/10.1111/ele.12052)
- Menden-Deuer, S., and E. J. Lessard. 2000. Carbon to volume relationships for dinoflagellates, diatoms, and other protist plankton. *Limnol. Oceanogr.* **45**: 569–579. doi:[10.4319/lo.2000.45.3.0569](https://doi.org/10.4319/lo.2000.45.3.0569)
- Mullin, M. M., P. R. Sloan, and R. W. Eppley. 1966. Relationship between carbon content, cell volume, and area in phytoplankton. *Limnol. Oceanogr.* **11**: 307–311. doi:[10.4319/lo.1966.11.2.0307](https://doi.org/10.4319/lo.1966.11.2.0307)
- Niklas, K. J. 2000. The evolution of plant body plans—a biomechanical perspective. *Ann. Bot.* **85**: 411–438. doi:[10.1006/anbo.1999.1100](https://doi.org/10.1006/anbo.1999.1100)
- Olenina, I. 2006. Biovolumes and size-classes of phytoplankton in the Baltic Sea. HELCOM, Baltic Sea Environ. Proc. **106**: 1–4.
- Padisák, J., É. Soróczki-Pintér, and Z. Rezner. 2003. Sinking properties of some phytoplankton shapes and the relation of form resistance to morphological diversity of plankton—an experimental study. *Hydrobiologia* **500**: 243–257. doi:[10.1023/A:1024613001147](https://doi.org/10.1023/A:1024613001147)
- Pomati, F., N. J. B. Kraft, T. Posch, B. Eugster, J. Jokela, and B. W. Ibelings. 2013. Individual cell based traits obtained by scanning flow-cytometry show selection by biotic and abiotic environmental factors during a phytoplankton spring bloom. *PLoS One* **8**: e71677. doi:[10.1371/journal.pone.0071677](https://doi.org/10.1371/journal.pone.0071677)
- Reynolds, C. S. 2006. *Ecology of phytoplankton*. Cambridge Univ. Press. doi:[10.1017/CBO9780511542145](https://doi.org/10.1017/CBO9780511542145)
- Rocha, M. R., D. A. Vasseur, M. Hayn, M. Holschneider, and U. Gaedke. 2011. Variability patterns differ between standing stock and process rates. *Oikos* **120**: 17–25. doi:[10.1111/j.1600-0706.2010.18786.x](https://doi.org/10.1111/j.1600-0706.2010.18786.x)
- Rocha, M. R., D. A. Vasseur, and U. Gaedke. 2012. Seasonal variations alter the impact of functional traits on plankton dynamics. *PLoS One* **7**: e51257. doi:[10.1371/journal.pone.0051257](https://doi.org/10.1371/journal.pone.0051257)
- Sommer, U., Z. M. Gliwicz, W. Lampert, and A. Duncan. 1986. The PEG-model of seasonal succession of planktonic events in fresh waters. *Arch Hydrobiol* **106**: 433–471.
- Sun, J., and D. Liu. 2003. Geometric models for calculating cell biovolume and surface area for phytoplankton. *J. Plankton Res.* **25**: 1331–1346. doi:[10.1093/plankt/fbg096](https://doi.org/10.1093/plankt/fbg096)
- Tirok, K., and U. Gaedke. 2006. Spring weather determines the relative importance of ciliates, rotifers and crustaceans for the initiation of the clear-water phase in a large, deep lake. *J. Plankton Res.* **28**: 361–373. doi:[10.1093/plankt/fbi121](https://doi.org/10.1093/plankt/fbi121)
- Utermöhl, H. 1958. Zur vervollkommnung der quantitativen phytoplankton-methodik. *Mitt. Int. Ver. Theor. Angew. Limnol.* **9**: 1–38. doi:[10.1080/05384680.1958.11904091](https://doi.org/10.1080/05384680.1958.11904091)
- Weithoff, G., and U. Gaedke. 2017. Mean functional traits of lake phytoplankton reflect seasonal and inter-annual changes in nutrients, climate and herbivory. *J. Plankton Res.* **39**: 509–517. doi:[10.1093/plankt/fbw072](https://doi.org/10.1093/plankt/fbw072)
- Young, J. R., and P. Ziveri. 2000. Calculation of coccolith volume and its use in calibration of carbonate flux estimates. *Deep-Sea Res. Part II Top. Stud. Oceanogr.* **47**: 1679–1700. doi:[10.1016/S0967-0645\(00\)00003-5](https://doi.org/10.1016/S0967-0645(00)00003-5)

#### Acknowledgments

ABR acknowledges the Lower Saxony Ministry for Science and Culture, project POSER. UM was funded by funded by DFG project BL 772/6-1 within the DFG Priority Programme 1704 DynaTrait.

#### Conflict of Interest

None declared.

Submitted 21 March 2018

Revised 09 November 2018

Accepted 17 November 2018

Associate editor: Tammi Richardson