



# Length, width, shape regularity, and chain structure: time series analysis of phytoplankton morphology from imagery

Virginie Sonnet ,<sup>1,2\*</sup> Lionel Guidi ,<sup>1</sup> Colleen B. Mouw ,<sup>2</sup> Gavino Puggioni ,<sup>3</sup>  
Sakina-Dorothee Ayata <sup>1,4</sup>

<sup>1</sup>CNRS, Laboratoire d'Océanographie de Villefranche, Sorbonne Université, Villefranche-sur-Mer, France

<sup>2</sup>Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island

<sup>3</sup>Department of Computer Science and Statistics, University of Rhode Island, Kingston, Rhode Island

<sup>4</sup>Laboratoire d'Océanographie et du Climat, Institut Pierre Simon Laplace (LOCEAN, SU/CNRS/IRD/MNHN), Sorbonne Université, Paris, France

## Abstract

Functional traits are increasingly used to assess changes in phytoplankton community structure and to link individual characteristics to ecosystem functioning. However, they are usually inferred from taxonomic identification or manually measured for each organism, both time consuming approaches. Instead, we focus on high throughput imaging to describe the main temporal variations of morphological changes of phytoplankton in Narragansett Bay, a coastal time-series station. We analyzed a 2-yr dataset of morphological features automatically extracted from continuous imaging of individual phytoplankton images (~ 105 million images collected by an Imaging FlowCytobot). We identified synthetic morphological traits using multivariate analysis and revealed that morphological variations were mainly due to changes in length, width, shape regularity, and chain structure. Morphological changes were especially important in winter with successive peaks of larger cells with increasing complexity and chains more clearly connected. Small nanophytoplankton were present year-round and constituted the base of the community, especially apparent during the transitions between diatom blooms. High inter-annual variability was also observed. On a weekly timescale, increases in light were associated with more clearly connected chains while more complex shapes occurred at lower nitrogen concentrations. On an hourly timescale, temperature was the determinant variable constraining cell morphology, with a general negative influence on length and a positive one on width, shape regularity, and chain structure. These first insights into the phytoplankton morphology of Narragansett Bay highlight the possible morphological traits driving the phytoplankton succession in response to light, temperature, and nutrient changes.

Morphology is the first characteristic we notice in any organism, whether it is a human, a cat, a shellfish, or a microorganism. Cell morphology is also a primary criterion to classify unicellular organisms like phytoplankton into species, both with light microscopy and from continuous imagery. Phytoplankton are unicellular and colonial photosynthetic

microorganisms that account for approximately half of Earth's primary productivity (Field et al. 1998). Their morphology directly affects their fitness through its impact on nutrient acquisition, light perception, cell division, defense against grazing by herbivores, and buoyancy (Litchman and Klausmeier 2008; Naselli-Flores and Barone 2011; Marañón 2015). Morphological characteristics are, as such, functional traits that can impact growth, reproduction, and survival (Litchman and Klausmeier 2008; Martini et al. 2021). Although hard to measure and to define in a universal way, cell size is considered a master trait among functional traits. In two dimensions, it can be estimated from surface area, major axis, or perimeter, but it can also be replaced by estimations of cell volume for relative abundances (e.g.: Hillebrand et al. 1999), more accurate for biomass and carbon analyses. Cell size is a well-defined metric in microscopy but only includes the part of the cell within the cell wall; if used alone, it thus occults the more complex morphological characteristics. Extensions and protuberances, for instance, play a major

\*Correspondence: [virginie\\_sonnet@uri.edu](mailto:virginie_sonnet@uri.edu)

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

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

**Author Contribution Statement:** C.B.M. provided the data, S-D.A. and V.S. designed the study, V.S. carried out the data processing and analysis with guidance from G.P., S-D.A. and L.G. V.S. and S-D.A. wrote the first draft. All authors reviewed and approved the final version of the manuscript.

role in modulating the sinking velocity of phytoplankton (Naselli-Flores et al. 2020), and an increase in size will affect differently the sinking velocity of an elongated or a spherical cell (Durante et al. 2019). At sub-saturating light, elongated cells also have higher growth than spherical or ellipsoidal ones, likely due to a better exposition of chloroplasts (Naselli-Flores and Barone 2011).

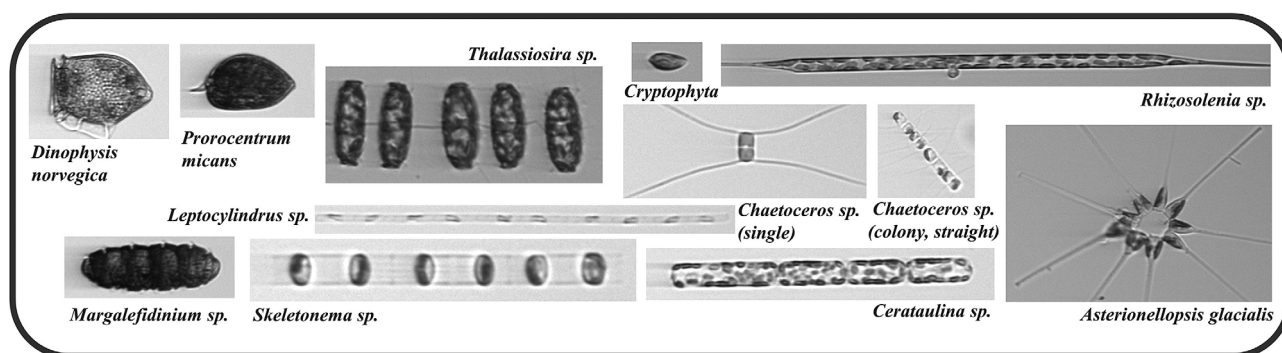
Differences in morphological characteristics drive the classification of phytoplankton into different phyla, families, genera, and species. Traditionally, water was collected with bottles or plankton nets and later analyzed under light microscopy, marking the creation of the first long-term plankton time series. Imaging devices have improved our capacity to monitor phytoplankton on broader temporal and spatial scales using automatic classification (Irisson et al. 2022) but the focus remains on using the morphology to infer the taxonomy. The degree of precision at the taxonomic level can also differ widely between datasets depending on both the quality of the annotated images and the algorithm used for classification. Obtaining accurate data both with light microscopy and imagery thus requires time and taxonomic expertise, and the ecological interpretation of the data relies on the quality of the identification. However, since morphology is the main criteria for discriminating taxa, it also holds, by itself, important additional ecological value independent from the quality of the taxonomic information.

Morphological analysis reduces high-dimensional datasets of multiple species to a few selected morphological traits characteristic of the community. By taking into account functional redundancy among taxa, it groups species with similar features without reducing the ecological information (Abonyi et al. 2018). Simplifying phytoplankton communities to a set of common functional traits also enables comparing environments based on phytoplankton functional roles in the ecosystems even when species might differ (Salmaso et al. 2015). Freshwater phytoplankton morpho-based classifications, for instance, can be better explained by environmental variables than phylogenetic and species-based classifications (Kruk et al. 2011). Here, we investigate cell morphology that can now be automatically studied using high throughput imaging systems. Imaging devices such as the CytoSense (Cytobuoy b. v., Dubelaar et al. 2000) or the Imaging FlowCytobot (IFCB, Olson and Sosik 2007) combine flow cytometry and imagery, targeting phytoplankton populations respectively from  $< 1 \mu\text{m}$  up to 4 mm (CytoSense) and from  $\sim 10$  to 150  $\mu\text{m}$  (IFCB). For each image, associated software (e.g., CytoClus for the CytoSense) or open-source processing codes (Sosik and Olson 2007, for the IFCB) provide easy, fast, reliable, and quantitative measurements of cell morphological features. For example, Fragoso et al. (2019) used eight morphological features derived from discrete water samples analyzed with a CytoSense to characterize trait variability of phytoplankton and plastid ciliates in the subpolar North Atlantic. Dunker (2020) combined taxonomic and morphological information from discrete samples run with an ImageStream<sup>®</sup>X Mk II to

characterize the seasonal changes of cyanobacteria and green algae in a pond. In parallel to the recent development of these instruments, several morphology-based approaches, mainly in freshwater ecosystems, have been developed to estimate the functional and morphological heterogeneity of phytoplanktonic communities (Weithoff and Beisner 2019). Focusing on freshwater phytoplankton, Reynolds et al. (2002) formalized decades of studies of functional groups into 31 associations of organisms sharing similar morphological features, habitats, or tolerances to nitrogen, light and carbon. On the other hand, Kruk et al. (2010) focused on morphology and developed morphology-based functional groups using traits determined from microscopy including volume, surface area, maximum linear dimension, aerotopes, flagella, mucilage, heterocysts, and siliceous exoskeletal structures. These approaches, developed for freshwater lakes, have paved the way for automated and solely morphological approaches for marine phytoplankton. However, they either make a detour through inferring morphological features from taxonomy, require time consuming light microscopy measurements, or include parameters that are not as easily detectable by imagery and automatic classification (e.g., mucilage, flagella), reducing the possible application to large temporal and spatial scales.

Narragansett Bay has been studied via weekly sampling since 1959 and the general taxonomical seasonal pattern is thus well known, offering a knowledge foundation to build upon and expand towards morphology. This coastal estuary is characterized by diatom domination in winter–spring and late summer months and flagellate domination in late spring–early summer and autumn (Pratt 1959). The winter–spring bloom supports fast-growing species with sizes greater than 20  $\mu\text{m}$  and consists in the succession of different populations; a dominant taxon, morphologically similar or not, replacing one another within a few weeks (Gowen et al. 2012). Karentz and Smayda (1998) detailed the succession, starting with winter species with low temperatures for optimal growth such as *Thalassiosira nordenskiöldii* and *Detonula confervacea*, moving to spring species such as *Leptocylindrus danicus*, often highest in abundance in May, then to summer species like *Heterosigma akashiwo* and *Prorocentrum* sp., motile and occurring at higher temperatures, and eventually fall species such as *Thalassiosira rotula*. Other important species can have both a warm-water and cold-water peak (e.g., *Skeletonema costatum* and *Rhizosolenia setigera*) or may, in the case of *Asterionellopsis glacialis* certain years, dominate the population almost every month except during the summer.

All species mentioned above have common morphological characteristics as well as unique features that can be captured by IFCB images (Fig. 1). Looking only at morphology, we might expect similar signals in elongation from communities dominated by chain-diatoms, while the width and shape of individual cells would allow for the taxonomic differentiation. Diatoms also tend to decrease in size throughout their life cycle due to asexual reproduction (Amato et al. 2005).



**Fig. 1.** Imaging FlowCytobot images of some of the most common phytoplanktonic organisms observed in Narragansett Bay at the GSO Pier. Included are chain-forming diatoms (*Asterionellopsis glacialis*, *Skeletonema* sp., *Cerataulina* sp., *Leptocylindrus* sp.), diatoms that can be found as single cells or as chains (*Thalassiosira* sp., *Chaetoceros* spp., *Rhizosolenia* sp.), the dinoflagellates *Dinophysis norvegica*, *Prorocentrum micans*, and *Margalefidinium* sp., and the image of a Cryptophyte.

Therefore, as one dominant chain-diatom replaces the other during the winter–spring blooms, there might be a small decrease in size. If dominant, species such as *Chaetoceros* spp. or *A. glacialis* should increase measures of perimeter and complexity due to their structure, spines, or setae while cryptophytes, dinoflagellates, and flagellates, on the contrary, would be expected to have a more homogeneous shape. Approaching these signals by the prism of morphology allows us to focus on the changes that are significant for the community. For instance, the replacement of a chain-diatom dominated population by another chain-diatom may not affect the predator–prey relationship of phytoplankton with zooplankton. Dominant phytoplankton groups are highly variable from 1 yr to the next, it is thus difficult to identify recurring yearly patterns in taxonomy. On the other hand, the morphological succession and functional role in the ecosystem may remain the same if taxa are replaced by others with a similar morphology (Gowen et al. 2012). By considering the whole community, we avoid focusing on only a small number of species selected based on their abundance—a possible bias towards bloom periods with high abundances—and include cells like nanophytoplankton, which are harder to detect and enumerate via light microscopy.

Here, we want to characterize the main morphological variations of these coastal phytoplankton communities by combining a trait-based approach with an automatic imaging system. Based on the winter–spring bloom usually detected with taxonomic analysis, we hypothesize that diatoms, and specifically chains, would dominate the morphological signal in winter and spring with consistent domination of elongated communities throughout this period. There may however be more changes in other morphological characteristics due to other differences between the successively dominant chain species and as environmental conditions change. On the contrary, we hypothesize that summer would show a very different morphological signal mainly dominated by smaller organisms. To study these variations of the morphological

signal, we use morphological features automatically extracted from individual cell images collected using an IFCB at a coastal time-series station over a 2-yr period. Using multivariate analysis, we create a morphological space capturing the main morphological variations of phytoplanktonic cells. Finally, we characterize the temporal dynamics of phytoplankton morphology within this morphological space in relation to environmental changes.

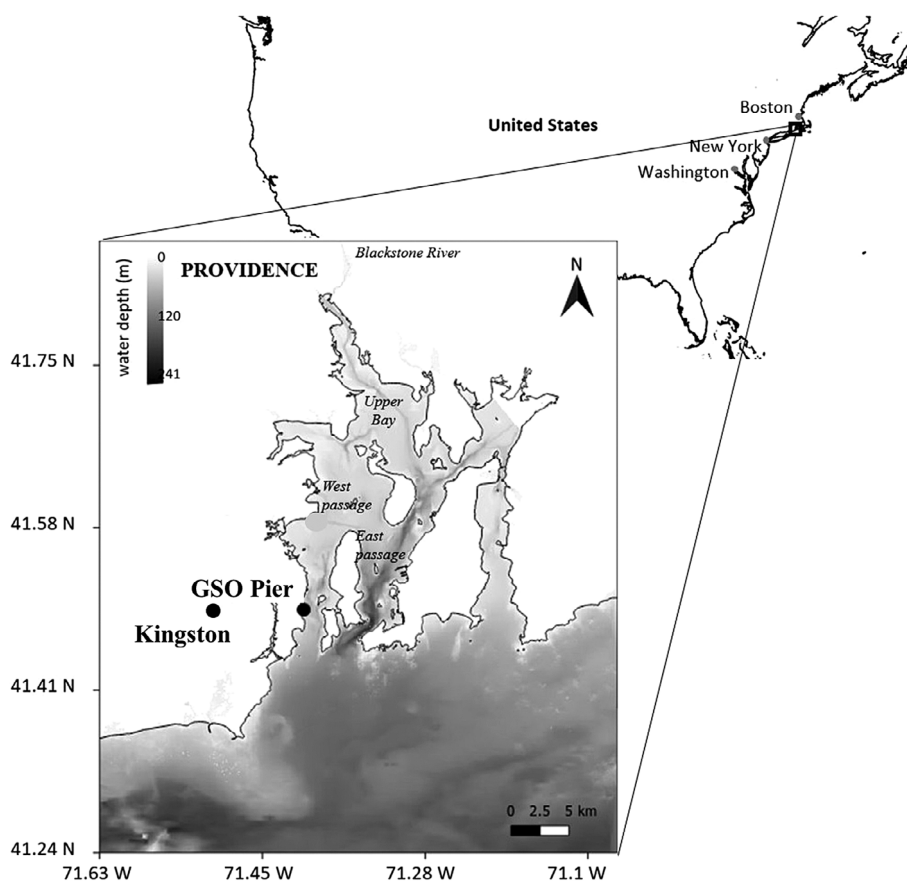
## Methods

### Sampling

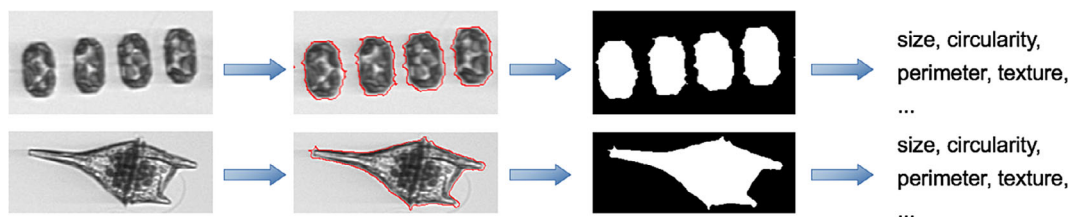
Water samples were continuously collected at the University of Rhode Island, Graduate School of Oceanography Pier (GSO Pier, total depth: 2 m, Lat: 41.4923, Lon: -71.1489, 11 November 2017 to 30 October 2019) using an IFCB. This station is located on the west shore of a temperate and shallow estuary, Narragansett Bay, on the northeast coast of the United States (Fig. 2). Narragansett Bay is New England's largest estuary and a populated area subject to anthropogenic pressures (Roman et al. 2000; Narragansett Bay National Estuarine Research Reserve 2009). The Bay is well mixed with similar salinity and nutrient concentration at the bottom and surface. Nutrients peak in fall and are close to depletion in spring–summer (Heffner 2009). Large islands complexify the circulation of water masses in the southern half of the Bay with the water flowing in through the East passage and out the West passage even though deeper water may also flow north in or out of the Bay depending on the tides (Kincaid et al. 2008).

### Phytoplankton imaging

The IFCB combines a flow cytometer and a camera to record images of phytoplankton between 10 and 150  $\mu\text{m}$  (Olson and Sosik 2007). The water is pumped from 1.5 m deep into a small research building located on the pier and filtered with 400 and 150  $\mu\text{m}$  screens before entering the IFCB. Triggered by fluorescence, the instrument records a file of images approximately every 20 min and 236 image descriptors are automatically



**Fig. 2.** Bathymetric map of Narragansett Bay with sampling stations. Sampling locations for phytoplankton imaging, physical variables, nutrients (GSO Pier), and light (Kingston) are indicated with a black dot and label. The level of gray represents the water depth with darker tones indicating greater depth.



**Fig. 3.** Example of IFCB image processing for a chain (top) and an individual cell (bottom). The MATLAB® image processing code detects the edges of organisms on the image and converts the image into blobs (see Sosik and Olson 2007). The blobs are the organisms detected on the image (i.e., white part of the black and white image), it can be a single cell or several cells depending on the morphology. Based on the original gray-scale and processed image, the code computes 236 image descriptors covering different measures of size, contrast, perimeter, circularity, and so on (see Supporting Information Table S1 for the morphological descriptors that were used in this analysis).

measured for each individual image (Sosik and Olson 2007, <https://github.com/hsosik/ifcb-analysis/wiki>) (Fig. 3).

**Environmental variables**

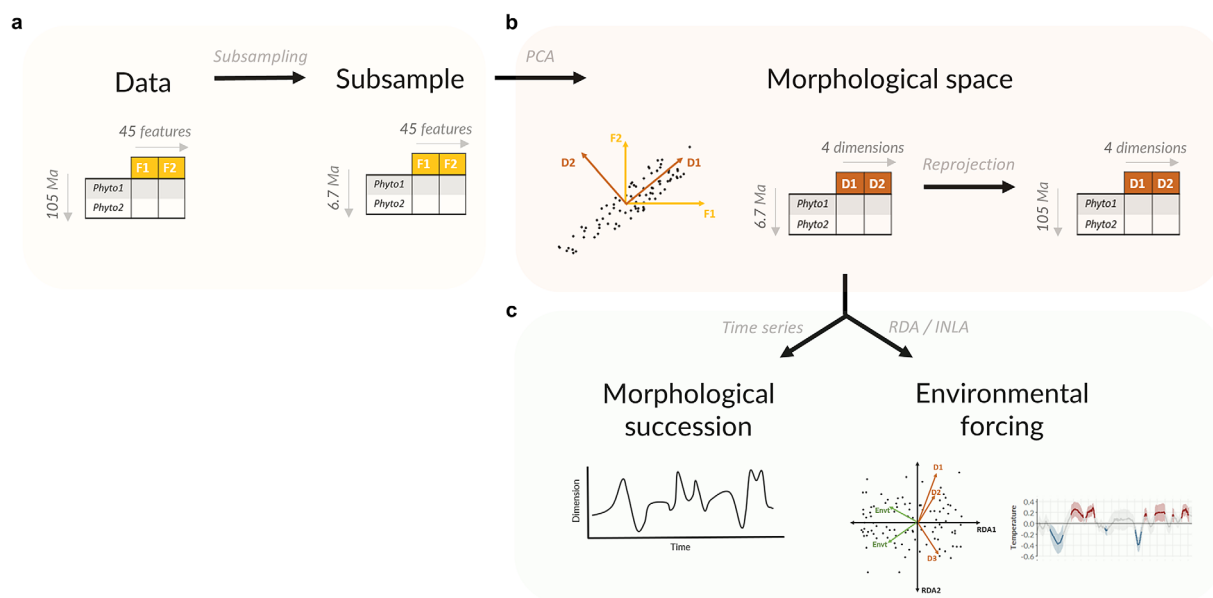
Environmental parameters were compiled from two locations (Fig. 2). Temperature and salinity were continuously measured at 15-min intervals at the GSO Pier by the Narragansett Bay Fixed-Site Monitoring Network (NBFSMN, contact: Heather Stoffel) and nutrients (silica, phosphate and nitrogen) were measured weekly at the same location by the Marine

Ecosystems Research Lab (Reed and Oviatt, 1976-2019, Marine Ecosystems Research Laboratory) Solar radiation was measured every 5 min in Kingston, west of the Pier, by the U.S. Climate Reference Network (<https://www1.ncdc.noaa.gov/pub/data/uscrn/products/subhourly01/>).

**Morphological analysis**

The morphological analysis of the phytoplankton images involved three steps (Fig. 4): data preprocessing, dimension reduction to define morphological traits, and time series analysis.





**Fig. 4.** Workflow of our phytoplankton morphological analysis. **(a)** The data processing consists of selecting the features relevant for morphology, removing the nonphytoplankton images, and subsampling the dataset with 250 images per sample, **(b)** the PCA was used for dimension reduction to create a morphological space and, using the SVM matrix, we reprojected every image in this morphological space, **(c)** the time series analysis included looking at the morphological succession and the environmental forcing through pairwise mean comparison statistical test, RDA, and Bayesian DLM (West and Harrison 1997).

### Data preprocessing

The IFCB is deployed in a flowthrough system; if the pump is not operating optimally, bubbles can lower the cell counts and eventually stop the flow. We used manual quality control to remove samples that had artifacts or bubbles. We also removed all files with less than 250 cells  $\text{mL}^{-1}$  and less than 250 images to exclude the files that would not accurately represent the community. Manually annotated images via the tool IFCB annotate (Woods Hole Oceanographic Institution, <https://ifcb-annotate.whoi.edu/>) were used to create a classifier based on a Random Forest algorithm (Breiman 2001). This led to the removal of images classified as “bubbles”, “detritus”, “zooplankton”, “ciliates”, and images within the category “unclassified” with an area greater than 90,000 pixels to avoid nematodes or macroalgae. The removal of these classes allowed for keeping all images classified as phytoplankton, that is, 68,539,385 images of the 104,613,040 initial images set for 26,720 samples. A simple classifier with very few broad classes can also be used when no previous classification is available. The classification mentioned above was later used to extract species composing the greatest percentage of the community every week.

Due to the redundancy of our high-frequency data and the computation cost of working with more than 68 million images, we chose to adopt a subsampling strategy to create the morphological space. We randomly chose 5% of the total samples and extracted 3, 10, 50, 100, 250, 500, and 1000 random images per sample and repeated this process 10 times to

estimate how a principal component analysis (PCA; Legendre and Legendre 1998) would be affected by the number of images extracted. Considering the extent of our dataset and the desire to also capture rare events, we chose to subsample 250 images per sample since the variation between replicates converged with this sample size and no greater value was gained by sampling 500 or 1000 images (see Supporting Information Figs. S1, S2). This corresponds to 6,680,000 images for 26,720 samples.

### Dimension reduction: Defining a morphological space from individual phytoplankton images

We used 45 morphological features (see Supporting Information Table S1) of the 6,680,000 subsampled phytoplankton individual images to define a morphological space using multivariate analyses in an approach similar to the one used for copepods by Vilgrain et al. (2021). Among the 236 image descriptors that were automatically calculated on each image by the MATLAB<sup>®</sup> routine, the 45 features selected were important in describing the whole cells and not only part of the image, hence having ecological importance. They correspond to the size, volume, symmetry, texture, shape, colony, and chain measures. This includes, among others, area, convex area, major axis, minor axis, circularity, the ratio of perimeter over major axis, levels of gray, number of blobs detected on the image (see Supporting Information Table S1 for a full list and definitions). After a Box Cox transformation (Box and Cox 1964) on skewed variables (all but eccentricity, extent, number of blobs, circularity, ratio of area over perimeter

squared, and the features related to distance or levels of gray) and normalization, we generated an observation-weighted PCA to define a morphological space. The interest of multivariate analysis such as PCA resides in the reduction of complex data into a small number of composite axes correlated with the original variables but orthogonal, thus independent from each other. However, the PCA gives more weight to the most extreme values even if they are extremely rare. To avoid giving an out-of-proportion weight to very extreme but very rare morphologies and to better represent the most abundant morphologies, we chose to implement an observation-weighted PCA using the concentration of the samples as weight. The weight was calculated by Eq. (1) with  $N_s$  the number of images in the sample,  $0.25 \text{ mL min}^{-1}$  the flow speed,  $\text{runTime}$  the time the sample was run, and  $\text{inhibitTime}$  the time the IFCB was writing images to the disk and not recording.

$$C = \frac{N_s}{0.25 \times \frac{(\text{runTime} - \text{inhibitTime})}{60}} \quad (1)$$

The significant axes of the observation-weighted PCA were determined using Cattell's scree test, also called the visual elbow method (Cattell 1966, see Supporting Information Fig. S3), and used as new synthetic morphological traits. To obtain the full picture of the community for every sample, we used the singular value decomposition matrix from the PCA to reproject every one of the 68,539,385 images into our morphological space. The details of the data preprocessing, subsampling, and dimension reduction are included in a GitHub repository, along with a 2-d subset of our dataset and accompanying data files for illustration (<https://github.com/VirginieSonnet/spidr>).

### Statistical and environmental analyses

We used a Wilcoxon test of comparison of means to compare the inter-annual seasonal variability of each of the morphological traits. To explore the potential relationships between morphological traits and environment variables on a weekly time scale we ran a redundancy analysis (RDA; Legendre and Legendre 1998) with light, salinity, temperature, nitrogen, phosphate, and silicate as explanatory variables. The RDA, similarly to the PCA, creates composite axes that are orthogonal to each other and represent the variance explained by the environmental variables. Going down to a finer time scale, at the hourly level, we used Bayesian DLM (West and Harrison 1997) to explore how these relationships between morphology and environment may change over time. In a DLM, environmental parameters can be either static (as in a regular linear regression model) or dynamic, with regression coefficients changing over time. DLMs are especially suitable for complex time series as they can handle non-stationary processes and missing observations (Laine 2020). DLMs are represented as a system of two equations: the observational equation, modeling the time series with static and dynamic

parameters, and the evolution equation, controlling how the dynamic parameters change over time. The observations  $Y_t$  were modeled as dependent of  $S_t$ , a seasonal component,  $Z_t$ , a set of covariates whose vector of regression coefficients  $\beta$  does not change over time,  $F_t$ , a set of covariates (including a dynamic intercept) with dynamic regression coefficients  $x_t$ , and  $\nu_t$  the observation error:

$$Y_t = S_t + Z_t\beta + F_t x_t + \nu_t. \quad (2)$$

The dynamic intercept and the observation error bring additional information on top of the covariates: the former is the value of the response when all the other predictors are equal to zero while the latter measures the component not captured by the model. Different specifications with several combinations of covariates and static/dynamic coefficients were considered. Model selection was conducted to maximize the marginal log-likelihood. After testing whether each covariate was better modeled with either fixed or dynamic parameters, the best fit was obtained with a model including temperature, salinity, and light with dynamic parameters, no fixed parameters, a 24-h seasonal component, and a dynamic intercept. All dynamic components were normalized and modeled with random walks as we expect the regression coefficients to be similar to their value at the previous time. The dynamic regression coefficients  $x_t$  were estimated from the evolution equation below:

$$x_t = G_t x_{t-1} + \omega_t, \quad (3)$$

where  $G_t$  is the matrix of evolutions coefficients and  $\omega_t$  is a Gaussian evolution error.

Although DLM parameter estimation is often carried out using Markov Chain Monte Carlo algorithms, here we used a more recent and faster approach based on the Integrated Nested Laplace Approximation (INLA) (Rue et al. 2009; Ruiz-Cárdenas et al. 2012). The codes for the morphological analysis and figures can be found in a second Github repository (<https://github.com/VirginieSonnet/mapi>).

### Statistical software

Data preprocessing steps were run with bash scripts from the terminal for efficiency. We used MATLAB® 2019a (The MathWorks, Inc. 2019) to extract morphological features from the images and R 3.6.0. (R Core Team 2019) for statistical analyses. The main packages used include “tidyverse” (Wickham 2017) for data manipulations and graphics, “car” for Box-Cox transformation (Fox and Weisberg 2011), “FactoMineR” (Lê et al. 2008) for PCA, “factoextra” (Kassambara and Mundt 2017), and “INLA” (Martins et al. 2013, [www.r-inla.org](http://www.r-inla.org)) for the DLM.

## Results

### Morphological space

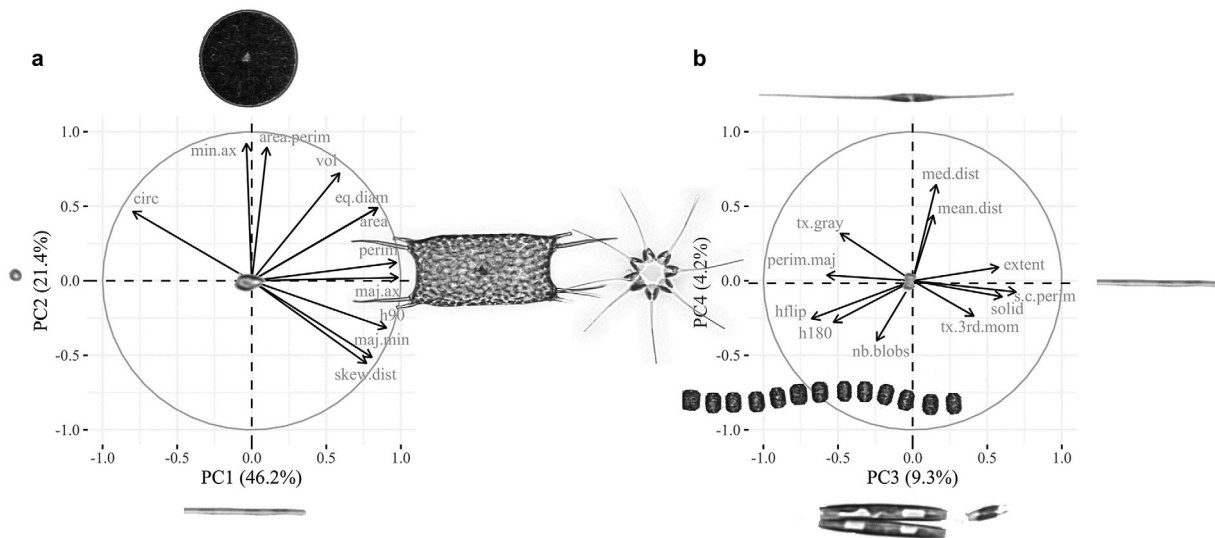
The three first axes of the PCA were significant and represented 76.9% of the morphological variations of the

phytoplankton cells (Fig. 5). We also considered the 4<sup>th</sup> axis for its ecological significance as it allows us to differentiate between chain structures. The 1<sup>st</sup> axis of the morphological space (PC1, 46.2% of the variance) corresponded to elongation. It was driven by the measures of major axis length, cell perimeter, and the different measures related to the surface area, hence separating longer and bigger cells (PC1 > 0) from smaller ones (PC1 < 0). The 2<sup>nd</sup> axis (PC2, 21.4% of the variance) represented the cell width. It was mainly driven by the minor axis length and the ratio of area over perimeter, both characteristics of the width of the cells, hence separating more circular and wider cells (PC2 > 0) from narrower cells (PC2 < 0). The biovolume and area contributed to both PC1 and PC2; while the area was more correlated with PC1, biovolume was more correlated with PC2. The 3<sup>rd</sup> axis (PC3, 9.3% of the variance) represented cell shape complexity or regularity. It separated cells with a simple shape (PC3 > 0), that is, regular cells having a high ratio of convex perimeter over perimeter, generally good symmetry (low Hausdorff distances after rotation), and a small difference between the area and the convex or bounding box area (high solidity and high extent), from cells with complex shape (PC3 < 0) with a high ratio of perimeter over major axis. Eventually, the 4<sup>th</sup> axis (PC4, 4.1% of variance) is more complex to interpret but, when interpreted in conjunction with PC1, can help differentiate chain structures. It is mainly driven by the median distance between points along the perimeter and the number of

blobs that can be detected on each image, which corresponds to the number of different enough entities for the processing software to consider them as different cells. Both of these measures, although in opposite directions along the PC4 axis, can be seen as chains structure measurements: very elongated and thin chains with little difference between the cells (e.g., *Leptocylindrus minimus*) will have a very high median distance between points along their perimeter but most likely will not be detected as several blobs while chains with very distinct chloroplasts (e.g., *Skeletonema* sp.) or connected by threads (e.g., *Thalassiosira* sp.) are likely to have more than one blob detected and thus have more negative values along PC4. When PC1 values are high, meaning organisms are elongated, PC4 can be used to differentiate the average chain structure. Table 1 illustrates the values of the main correlated variables for the average and extreme values of each principal component.

**Morphological succession**

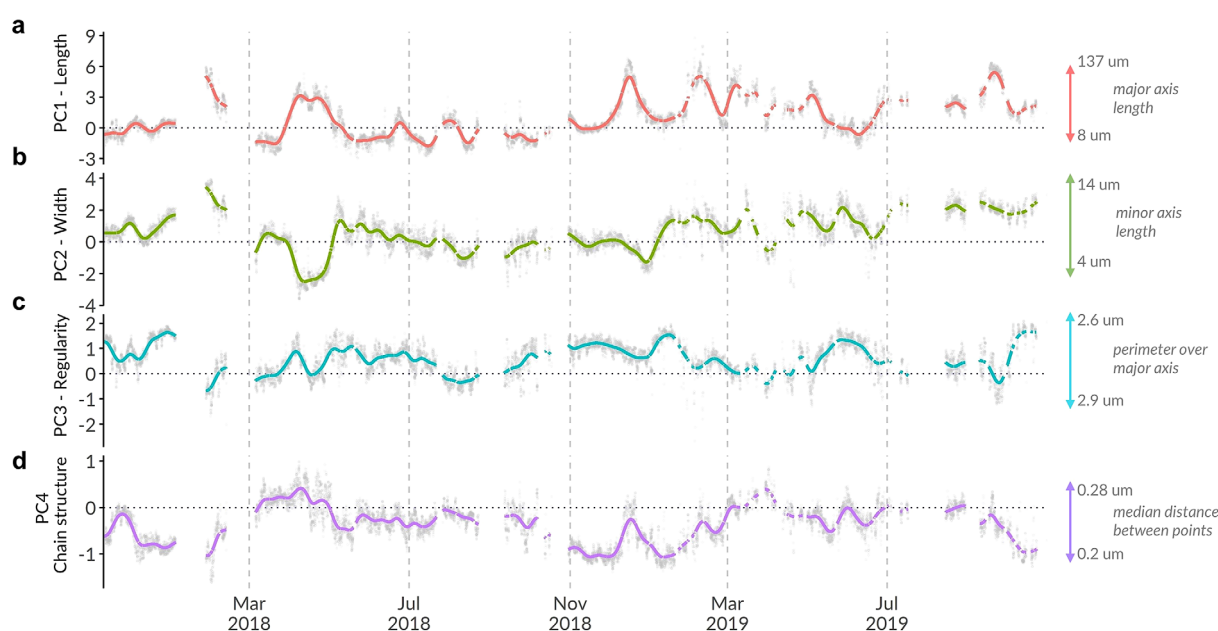
The temporal evolution of the four main axes of the morphological space, that is, the main morphological traits characterizing the phytoplankton community, revealed complex successions of morphologically different communities (Fig. 6) and contrasted seasonal and interannual patterns (Fig. 7). Contrary to our expectation, there was no clearly defined seasonal pattern for cell morphology. Instead of observing an increase in length in spring and a decrease in summer, the average phytoplankton community length (PC1) oscillated,



**Fig. 5.** Description of the morphological space of phytoplankton cells. (a) Main morphological features (i.e., with a  $\cos^2 > 0.7$  and only one variable for very correlated groups, for example, area for area, summed area, convex area) in the 1<sup>st</sup> two dimensions PC1 and PC2 of the morphological space. (b) Main morphological features (i.e., with a  $\cos^2 > 0.2$ ) in the 3<sup>rd</sup> and 4<sup>th</sup> dimensions PC3 and PC4. The size of the arrows materializes the importance of the correlation and, its direction, the sign of the correlation (positive or negative). The angle indicates the axis that is correlated with this variable: if the arrow is parallel to an axis, it is highly correlated, perpendicular means no correlation, and a 45° angle indicates a correlation with the two axes. Individual images with extreme values along each axis and in the center of the morphospace are shown to illustrate that PC1 represents cell length, PC2 cell width, PC3 cell shape regularity, and PC4 differences in the structure of chains detected. An extra image of *Thalassiosira* sp. is added to (b) to help illustrate PC4.

**Table 1.** Range of correspondences between principal components and main correlated variables. The general average and the average of the 5% having the maximum and minimum values along each principal component are indicated with the corresponding values for their most correlated variable(s). Values are rounded to the nearest integer for *Length* and *Width* and to the 1<sup>st</sup> decimal for *Shape regularity* and *Chain structure*. Biovolume was added to PC1 and PC2 because of its contribution to both axes.

	Length			Width			Shape regularity			Chain structure		
	PCI	Major axis ( $\mu\text{m}$ )	Biovolume ( $\mu\text{m}^2$ )	PC2	Minor axis ( $\mu\text{m}$ )	Biovolume ( $\mu\text{m}^2$ )	PC3	perim. maj	hflip ( $\mu\text{m}$ )	PC4	med. dist	nb. blobs
5% min	-6	6	62	-6	3	66	-4.4	2.4	1.4	-3.2	0.6	2.1
Mean	1	17	493	0	6	493	0.5	2.8	0.4	0	0.7	1.1
5% max	10	94	3744	7	16	3996	3.7	3.7	0.2	2.5	1.2	1



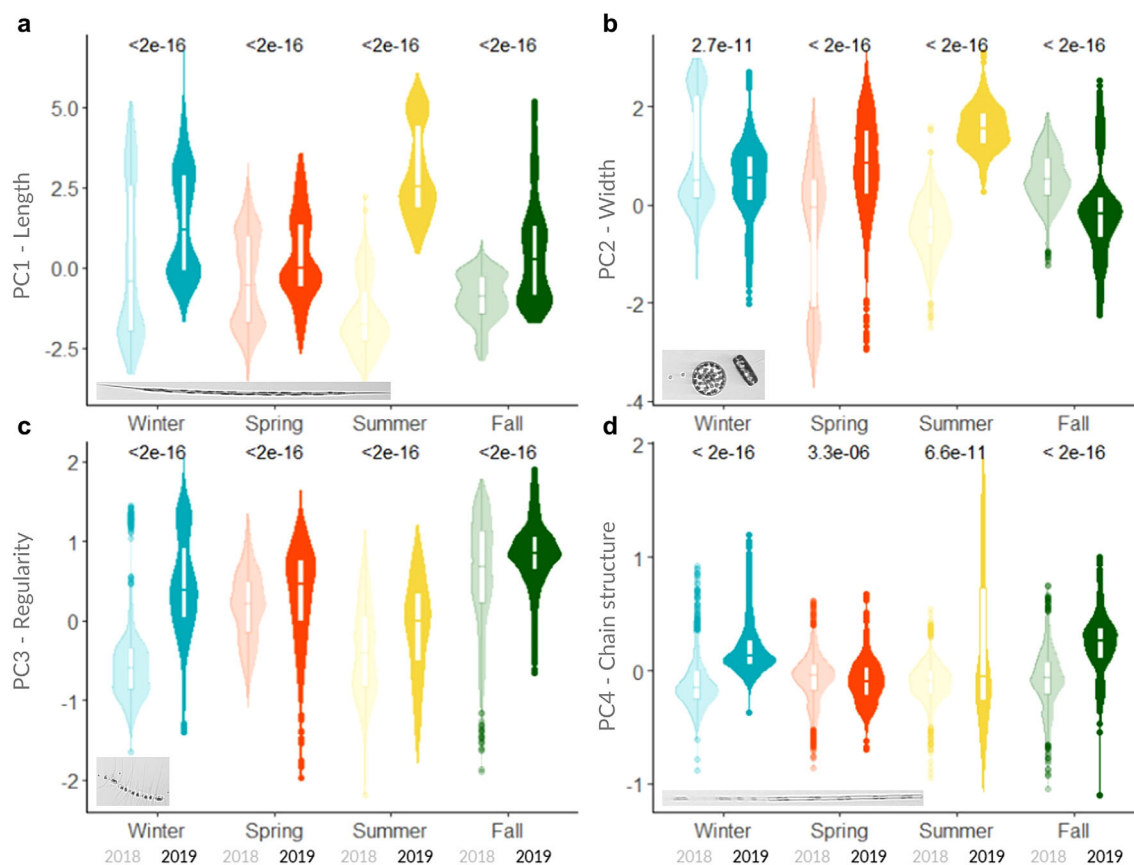
**Fig. 6.** Average value per hour of the main morphological traits (i.e., main axes of the morphological space) of the phytoplankton community in Narragansett Bay from November 2017 to October 2019. The 1<sup>st</sup> (a), 2<sup>nd</sup> (b), 3<sup>rd</sup> (c), and 4<sup>th</sup> (d) morphological traits, respectively, representing the length, the width, the shape regularity, and the structure of chains are shown. For a better visualization of the trends, the curves were smoothed 5 times with a weighted moving average over a window of 121 h and colored on top of the hourly averaged observations in gray. For reference, the range of the most correlated contributor for each component is indicated on the right.

suggesting a succession of communities characterized by smaller and longer cells (Fig. 6a). The length was generally lower in summer, but we saw several rises over the summer in 2018 and a peak—as important as the winter peaks—in September 2019. Length was also significantly higher in 2019 ( $p$ -value <  $2e-16$ , see Fig. 7a). Although missing values certainly increase the signal, it illustrates significant inter-annual variability and the importance of long-term time series. Cell width (PC2) significantly decreased in April 2018, coincidentally with a rise in length, as well as in late December 2018 (Fig. 6b). There might be a third major decrease in April 2019 but too many data points are missing to draw any robust conclusions. Cell shape regularity (PC3) also varied with time (Fig. 6c): phytoplankton communities generally had more regular shapes in autumn and late spring and more complex

shapes in February–March and summer (Fig. 7c). PC4 does not display a general seasonal pattern but when considered during the periods of high PC1, we can see that the February 2018 length increase was coupled with a low PC4 and thus included more chains with distant cells compared to the bloom of March 2018 (Fig. 6d). Similarly, the December 2018 and February 2019 length increases had lower PC4 values than March 2019, indicating more chains with differentiated cells.

There was high inter-annual variability between the years 2018 and 2019 (Fig. 7) for all the morphological traits but particularly in summer for the length and width, significantly higher in 2019 ( $p$ -value <  $2e-16$ ) and with little overlap in their distributions. Generally, for each season the length, width, shape regularity, and clearly connected chains were higher in 2019 than in 2018.





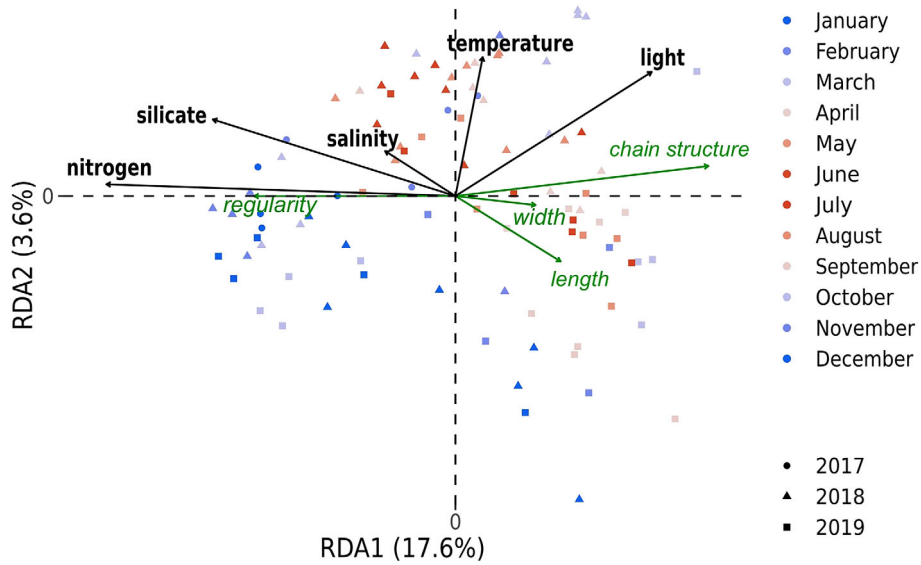
**Fig. 7.** Distribution of morphological traits per season and per year. The distributions are colored by season (winter in blue, spring in red, summer in yellow and fall in green) and contrasted per year (lighter in 2018 and darker in 2019) for (a) the length, (b) the width, (c) the shape regularity, and (d) the chain structure. The  $p$ -value of a Wilcoxon test comparing the 2018 and 2019 means for each season is displayed on top of the violin plots. The year 2018 refers to data between November 2017 and September 2018 and the year 2019 refers to data between October 2018 and October 2019.

### Morphological response to environmental forcing

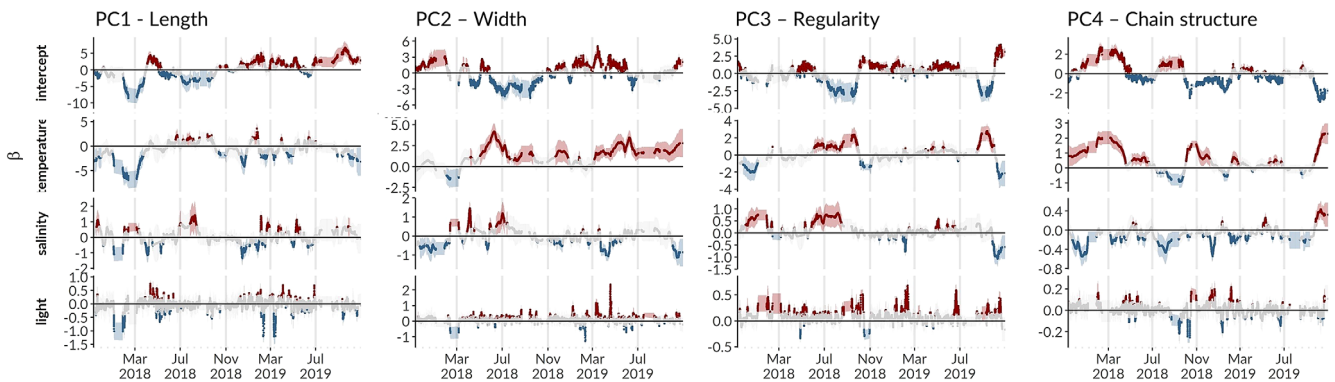
The RDA run on the weekly averages explains a significant proportion of the variance ( $p = 0.001$ ,  $R^2$ -adjusted = 0.18, Fig. 8) but only the 1<sup>st</sup> axis, constraining 17.6% of the variance, was significant ( $p = 0.001$ ). Light ( $p = 0.001$ ), nitrogen ( $p = 0.001$ ), silicate ( $p = 0.001$ ), and, in a lower measure, temperature ( $p = 0.02$ ) are the significant drivers of the RDA.

Cell length is strongly negatively correlated with salinity and silicate and negatively correlated with nitrogen. Cell width and clearly connected chains are negatively correlated with silicate, nitrogen, and salinity while shape regularity is positively correlated with them but negatively correlated with light. Along the 2<sup>nd</sup> axis, temperature is slightly negatively correlated with length and slightly positively correlated with uniform chains. Temporally, the RDA also reveals interannual variations in relation with environmental conditions, with larger and thicker cells in 2019 compared to 2018, with the addition of the coldest months of 2018 (December and January).

Bayesian DLM (Fig. 9) shows that at the hourly timescale, temperature is the factor with the highest magnitude coefficients, hence the one explaining most of the variability within the morphological signal. Similar to the RDA, temperature has a general positive influence on chain structure and a negative influence on length. However, for both morphological variables, this influence seems to decrease and sometimes even reverse in summer. The clear positive influence of temperature on width throughout the time series, contrary to the absence of a relationship in the RDA, could indicate that the daily cycle might influence this morphological variable. Salinity has a general negative influence on chain structure throughout the time series and a positive one on shape regularity but only until September 2018. Light seems to be generally positively influencing width and shape regularity, indicating that a daily cycle might be significant. The intercepts model the baseline characteristics of the signal when the environmental variables are equal to 0; they are generally negative in 2018 and positive in 2019 for length, width, and shape regularity, showing an increase in those measures and conversely for chain structure.



**Fig. 8.** Environmental influence on weekly changes in morphology. RDA on weekly environmental (black) and morphological (green) variables. The angle between morphological variables and environmental ones is proportional to their correlation (i.e., parallel is correlated, perpendicular uncorrelated) and the direction indicates the sign of this correlation (i.e., positive if in the same direction, negative if in the opposite direction). The months of sampling are differentiated by colors, on a bright red (summer months) to bright blue (winter months) gradient. The different years of sampling are materialized by shapes.



**Fig. 9.** Bayesian dynamic linear regression of the environmental factors. Posterior distribution around the median of the dynamic regression coefficients of temperature, salinity, and light on the main morphological traits estimated in the study: (a) length (PC1), (b) width (PC2), (c) shape regularity (PC3), and (d) chain structure (PC4). The median is in gray when the credible interval at 95% (shaded area) includes 0, in red when it is positive and blue when it is negative.

**Main phytoplankton groups**

Looking through the bank of images for the winter–spring 2018–2019, we can identify that the long and medium thick organisms in December 2018 are mostly driven by *Skeletonema* sp. and the drop in width at the end of December relates to an uptick in *Leptocylindrus* sp. in the community (Fig. 10). The thicker and less regular organisms observed in February 2019 are driven by *Thalassiosira* sp. The last peak in size in March 2019, which is slightly smaller, has similar values in width as the previous peak but less regular organisms and more uniform chains; being driven by an increase in chains of *Chaetoceros* spp. The nanophytoplankton are present year-round and make up a high percentage of the concentration in

the samples, including in the winter–spring season, which explains the consistent decrease in elongation observed in-between diatom blooms during this season.

**Discussion**

**Relationship between environment and phytoplankton morphology**

The strongest feature of our time series resides in the deepened oscillations in length during the winter and spring compared to the summer and fall. This is accompanied by a trend to go, between peaks in elongation, from regular chains with spaced cells to clearly connected chains. This implies not only



**Fig. 10.** The weekly average percentage of the community that are represented by the most dominant phytoplankton groups. The groups displayed are the ones representing the highest percentage of the community during at least 1 week of the time series (indicated in a darker color), excluding the unclassified, bubbles, and bead images and with the addition of *Cerataulina bergonii*. The names appear as they are defined by the classifier developed by the Mouw lab, grouping *Chaetoceros* spp. by single cells or colonies forming straight (or curly) chains rather than species. The pointed nanophytoplankton, square nanophytoplankton, and round nanophytoplankton all refer to different morphologies of nanophytoplankton.

a higher number of chains but also that these chains are more equally spaced. They are either linked by thin threads or with chloroplasts concentrated in one part of the cell such that the junctions between cells are not defined well enough for the algorithm to detect them. This results in accurately considering that there is more than one cell on the image (e.g., difference between *Skeletonema* sp. and *Cerataulina* sp.).

We observe that length correlates with lower salinity as winter mixing helps sustain large, non-motile cells (Stoyneva et al. 2007). Bigger organisms, whether as individual cells or chains also have an advantage against predation; data on copepods and zooplankton in Narragansett Bay show that increases in size occur right after increases in zooplankton concentration (Supporting Information Fig. S5). On the contrary, smaller phytoplankton in summer and fall are correlated with higher salinity and generally higher nutrient concentration (RDA, Fig. 8); before the blooms of long-chain diatoms deplete both silicate and nitrogen in winter–spring. Silicate shows the highest concentrations in summer and fall due to the low counts of diatoms. This occurs while nitrogen is depleted throughout the spring and summer but increases in fall before the chain diatoms blooms occur during the colder months. In summer, when turbulence is too weak to resuspend phytoplankton, smaller and buoyant cells have an advantage as they can regulate their vertical position, select

for the optimal light and nutrient conditions, and avoid sinking losses (Findlay et al. 2001; Huisman et al. 2004; Strecker et al. 2004). Zohary et al. (2017) also hypothesize that this adaptation enables species to adjust their sinking velocity in different seasons since sinking velocity naturally increases with cell size. When temperature increases, changing the density and absolute viscosity of the water thus increasing the sinking velocity, it pushes phytoplankton towards smaller cell size to compensate and keep a constant sinking velocity (Naselli-Flores et al. 2020). The width, independently from the length, tends to oscillate throughout the time series and generally increases with temperature, with higher temperature seemingly favoring more circular organisms than elongated ones. Throughout the winter, the width of the longest organisms also slightly increases.

Shape regularity oscillates throughout the time series and is high at the start of winter before slowly decreasing as nutrients also decrease. A decrease in regularity due to spines increases the form resistance factor (Walsby and Xypolyta 1977; Padišák et al. 2003), which lowers the sinking speed of the chain diatoms throughout winter. As nutrients are taken up and decrease, cells diverging from the regular spherical shape also increase their surface area to volume ratio and, de facto, their nutrient uptake efficiency (Karp-Boss and Boss 2016). However, a reduction in size is another possibility

to increase the nutrient uptake efficiency and such small cells are often more regular in shape. The nitrogen-depleted waters starting in March and especially the summer stratified waters offer thus opportunities for smaller cells as well as cells with more complex shapes, motile species, nitrogen fixers, and mixotrophic species (Litchman and Klausmeier 2008).

The measure brought forward by the 4<sup>th</sup> axis of the PCA relates to how many cells the algorithm detects on one image and how elongated they are. The same pattern occurs both in winter 2018 and 2019: as dominant chains succeed each other; the dominant chain structure also changes. Chains tend to increase as the water is cooler and less stratified (Kenitz et al. 2020), when the protection offered by a chain against predation by zooplankton might overcome the loss of efficiency in resource acquisition compared to single cells (Pahlow et al. 1997). However, the change in chain structure between different high elongation peaks is also an indication of how close the cells in the chains are to each other, either tightly linked together with many chloroplasts or, on the contrary, far apart with mainly chloroplasts in the middle of the cells or only linked by thin threads. As nutrients get depleted throughout the winter and spring, the uniformity in chains associated with size increases, showing that cells are more tightly linked with each other. A gain in surface area and thus efficiency in nutrient travel time for thin, elongated chains that are directly connected rather than linked by threads might be one of the drivers.

The effect of the spring-neap tide cycle on the level of mixing and the daily tidal cycle bringing organisms from upstream and downstream have been shown to influence the phytoplankton communities in estuaries (Cloern 1991; Sin 2000). The dynamic modeling approach applied to the hourly data shows that the relationship between environmental and morphological variables can change through time and can differ from the weekly time scale: temperature, for instance, incorporates both the tidal and daily signal and has a generally positive impact on width and chain structure and negative on elongation. The positive influence of light on width and shape regularity also at the hourly time scale confirms that there may be significant variations throughout the day. These 1<sup>st</sup> results highlight the role of these smaller time scales. In the future, the hourly resolution of the dataset will allow deeper analysis of the periodicities related to the seasonal, daily, and tidal cycles that might explain part of the morphological variability observed here.

### Phytoplankton morphology and taxonomy in Narragansett Bay

Phytoplankton display a wide range of sizes which is known to be a key parameter in phytoplankton ecology. Thus, we expected that a combination of size measures would be the main driver of morphological variation. We observed communities of longer and more regular cells in winter and early spring, in agreement with the diatom domination in

Narragansett Bay described by Pratt (1959) and Durbin et al. (1975). Indeed, diatoms are known to form very long chains and thus, would generally be the organisms with the longest elongation in our dataset. The number of blobs detected was the highest in winter but the lowest in spring, highlighting different diatoms dominate these seasons with winter diatoms without spines, more regularly shaped and more equally spaced than spring ones since the software was able to differentiate several cells on the image. The succession of peaks in length observed in winter and spring indicates that there are discontinuities in-between the different blooms: there is not just one diatom species dominating and being gradually replaced by a morphologically similar species, the diatom domination decreases before a bloom morphologically different from the previous one takes over. Adding taxonomic information from automatic classification confirms the succession of *Skeletonema* sp., *Leptocylindrus* sp., *Thalassiosira* sp., and *Chaetoceros* sp.-dominated communities. Since these communities are not gradually replacing each other, the dominance of diatoms decreases between blooms and the morphological signal comes back to a community dominated by smaller organisms with a regular shape; the taxonomic data corroborates the dominance of nanophytoplankton during those periods. As such, the morphological analysis allows us to distinguish which morphological traits are significantly changing throughout the blooming season: the general morphology decreases in length throughout the blooms while increasing in width and shape complexity. *Skeletonema* sp. have been extensively studied in Narragansett Bay and are considered the most abundant phytoplankton species of the Bay (Borkman and Smayda 2009; Windecker 2010; Canesi and Rynearson 2016). Although its highest cell counts have historically been recorded in late winter–early spring and mid-summer (Karentz and Smayda 1984), our analysis shows that, for our time period, the morphological signal of *Skeletonema* sp. most likely drives only the first winter–spring blooms and is not significantly influencing the summer communities. Morphological analysis provides an easy and fast way to investigate general and functional community changes without prior taxonomic identification and reduces the complexity of the data. Although species identity provides equally important information (Salmaso et al. 2015), the combination of both taxonomic and morphological information shall provide a more complete understanding of community, functional, and intra-specific changes.

### Deriving morphology from imagery

The use of continuous imaging devices offers new opportunities to monitor phytoplankton communities, especially to study the triggers and developments of harmful algae blooms (Campbell et al. 2013; Anderson et al. 2019). Indeed, those high-throughput devices allow investigations on a much finer time scale than discrete sampling. They also give the possibility to go back to the high-quality images at the origin of the



signal to perform further analysis retrospectively. Despite the cost of maintenance and computing (i.e., machine learning analysis of the images), their efficiency is high in comparison to the time-consuming task of identifying and measuring phytoplankton cells individually under a microscope. Similar to any net sampling, most of the imaging instruments cannot cover the whole size spectrum of the phytoplankton (Lombard et al. 2019), hence, in our analysis, picophytoplankton are not included, as the IFCB has a lower limit of 10  $\mu\text{m}$ . This limits the analysis of the full phytoplankton size range and the possibility to sample, with the same instrument, higher trophic levels such as zooplankton. However, on their own or coupled with other imagery devices with different range capabilities, they open new possibilities for learning and exploring changes in taxonomical, morphological, and functional community composition at high-frequency periodicities (light or tidal cycle) and on the morphological responses of phytoplankton to flow at smaller time scales as wished by Naselli-Flores et al. (2020). This work thus presents one of the many approaches to exploit such rich datasets beyond taxonomical labeling. Although our time coverage is limited (2 yr), the methodology we present here could be applied to longer time series; this one, in the future, to confirm those local trends or others, linked to instruments deployed in places that might have different environmental forcings and different targeted organisms.

## References

- Abonyi, A., Z. Horváth, and R. Ptacnik. 2018. Functional richness outperforms taxonomic richness in predicting ecosystem functioning in natural phytoplankton communities. *Freshw. Biol.* **63**: 178–186. doi:10.1111/fwb.13051
- Amato, A., L. Orsini, D. D'Alelio, and M. Montresor. 2005. Life cycle, size reduction patterns, and ultrastructure of the pennate planktonic diatom *Pseudo-nitzschia delicatissima* (Bacillariophyceae): Life cycle of *Pseudo-nitzschia delicatissima*. *J. Phycol.* **41**: 542–556. doi:10.1111/j.1529-8817.2005.00080.x
- Anderson, C. R., and others. 2019. Scaling up from regional case studies to a global harmful algal bloom observing system. *Front. Mar. Sci.* **6**: 250. doi:10.3389/fmars.2019.00250
- Borkman, D. G., and T. Smayda. 2009. Multidecadal (1959–1997) changes in *Skeletonema* abundance and seasonal bloom patterns in Narragansett Bay, Rhode Island, USA. *J. Sea Res.* **61**: 84–94. doi:10.1016/j.seares.2008.10.004
- Box, G. E. P., and D. R. Cox. 1964. An analysis of transformations. *J. R. Stat. Soc. Ser. B Methodol.* **26**: 211–252.
- Breiman, L. 2001. Random forests. *Mach. Learn.* **45**: 5–32.
- Campbell, L., D. W. Henrichs, R. J. Olson, and H. M. Sosik. 2013. Continuous automated imaging-in-flow cytometry for detection and early warning of *Karenia brevis* blooms in the Gulf of Mexico. *Environ. Sci. Pollut. Res.* **20**: 6896–6902. doi:10.1007/s11356-012-1437-4
- Canesi, K., and T. Ryneerson. 2016. Temporal variation of *Skeletonema* community composition from a long-term time series in Narragansett Bay identified using high-throughput DNA sequencing. *Mar. Ecol. Prog. Ser.* **556**: 1–16. doi:10.3354/meps11843
- Cattell, R. B. 1966. The scree test for the number of factors. *Multivar. Behav. Res.* **1**: 245–276. doi:10.1207/s15327906mbr0102\_10
- Cloern, J. E. 1991. Tidal stirring and phytoplankton bloom dynamics in an estuary. *J. Mar. Res.* **49**: 203–221. doi:10.1135/002224091784968611
- Dubelaar, G. B. J., P. L. Gerritzen, A. E. R. Beeker, R. R. Jonker, and K. Tangen. 2000. Design and first results of CytoBuoy: A wireless flow cytometer for in situ analysis of marine and fresh waters. *Cytometry* **8**: 247–254.
- Dunker, S. 2020. Imaging flow cytometry for phylogenetic and morphologically based functional group clustering of a natural phytoplankton community over 1 year in an urban pond. *Cytometry A* **97**: cyto.a.24044. doi:10.1002/cyto.a.24044
- Durante, G., A. Basset, E. Stanca, and L. Roselli. 2019. Allometric scaling and morphological variation in sinking rate of phytoplankton. *J. Phycol.* **55**: 1386–1393. doi:10.1111/jpy.12916
- Durbin, E. G., R. W. Krawiec, and T. J. Smayda. 1975. Seasonal studies on the relative importance of different size fractions of phytoplankton in Narragansett Bay (USA). *Mar. Biol.* **32**: 271–287. doi:10.1007/BF00399206
- Field, C. B., M. J. Behrenfeld, J. T. Randerson, and P. Falkowski. 1998. Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* **281**: 237–240.
- Findlay, D. L., S. E. M. Kasian, M. P. Stainton, K. Beaty, and M. Lyng. 2001. Climatic influences on algal populations of boreal forest lakes in the experimental lakes area. *Limnol. Oceanogr.* **46**: 1784–1793. doi:10.4319/lo.2001.46.7.1784
- Fox, J., and S. Weisberg. 2011. An R companion to applied regression. Sage.
- Fragoso, G. M., A. J. Poulton, N. J. Pratt, G. Johnsen, and D. A. Purdie. 2019. Trait-based analysis of subpolar North Atlantic phytoplankton and plastidic ciliate communities using automated flow cytometer. *Limnol. Oceanogr.* **64**: 1763–1778. doi:10.1002/lno.11189
- Gowen, R. J., P. Tett, and T. J. Smayda. 2012. Phytoplankton and the balance of nature: An opinion. *Estuar. Coast. Shelf Sci.* **113**: 317–323. doi:10.1016/j.ecss.2012.08.009
- Heffner, L. 2009. Nutrients in Mid-Narragansett Bay: A Spatial Comparison of Recent and Historical Data. M.S. Thesis. Univ. of Rhode Island.
- Hillebrand, H., C.-D. Dürselen, D. Kirschtel, U. Pollinger, and T. Zohary. 1999. Biovolume calculation for pelagic and benthic microalgae. *J. Phycol.* **35**: 403–424. doi:10.1046/j.1529-8817.1999.3520403.x

- Huisman, J., J. Sharples, J. M. Stroom, P. M. Visser, W. E. A. Kardinaal, J. M. H. Verspagen, and B. Sommeijer. 2004. Changes in turbulent mixing shift competition for light between phytoplankton species. *Ecology* **85**: 2960–2970. doi:10.1890/03-0763
- Irisson, J.-O., S.-D. Ayata, D. J. Lindsay, L. Karp-Boss, and L. Stemmann. 2022. Machine learning for the study of plankton and marine snow from images. *Ann. Rev. Mar. Sci.* **14**: 277–301. doi:10.1146/annurev-marine-041921-013023
- Karentz, D., and T. Smayda. 1984. Temperature and seasonal occurrence patterns of 30 dominant phytoplankton species in Narragansett Bay over a 22-year period (1959–1980). *Mar. Ecol. Prog. Ser.* **18**: 277–293. doi:10.3354/meps018277
- Karentz, D., and T. J. Smayda. 1998. Temporal patterns and variations in phytoplankton community organization and abundance in Narragansett Bay during 1959–1980. *J. Plankton Res.* **20**: 145–168. doi:10.1093/plankt/20.1.145
- Karp-Boss, L., and E. Boss. 2016. The elongated, the squat and the spherical: Selective pressures for phytoplankton shape, p. 25–34. *In* P. M. Glibert and T. M. Kana [eds.], *Aquatic microbial ecology and biogeochemistry: A dual perspective*. Springer International Publishing.
- Kassambara, A., and F. Mundt. 2017. factoextra: Extract and Visualize the Results of Multivariate Data Analysis. R package version 1.0.7. <https://CRAN.R-project.org/package=factoextra>
- Kenitz, K. M., E. C. Orenstein, P. L. D. Roberts, P. J. S. Franks, J. S. Jaffe, M. L. Carter, and A. D. Barton. 2020. Environmental drivers of population variability in colony-forming marine diatoms. *Limnol. Oceanogr.* **65**: 2515–2528. doi:10.1002/lno.11468
- Kincaid, C., D. Bergondo, and K. Rosenberger. 2008. The dynamics of water exchange between Narragansett Bay and Rhode Island sound, p. 301–324. *In* A. Desbonnet and B. A. Costa-Pierce [eds.], *Science for ecosystem-based management*. Springer.
- Kruk, C., V. L. M. Huszar, E. T. H. M. Peeters, S. Bonilla, L. Costa, M. Lürling, 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
- Kruk, C., E. T. H. M. Peeters, E. H. Van Nes, V. L. M. Huszar, L. S. Costa, and M. Scheffer. 2011. Phytoplankton community composition can be predicted best in terms of morphological groups. *Limnol. Oceanogr.* **56**: 110–118. doi:10.4319/lo.2011.56.1.0110
- Laine, M. 2020. Introduction to dynamic linear models for time series analysis. *In*: Montillet, J.P., Bos, M. (eds) *Geodetic Time Series Analysis un Earth Sciences*. Springer Geophysics. Springer, Cham. Stat: 139–156. doi:10.1007/978-3-030-21718-1\_4
- Lê, S., J. Josse, and F. Husson. 2008. FactoMineR: A package for multivariate analysis. *J. Stat. Softw.* **25**: 1–18. doi:10.18637/jss.v025.i01
- Legendre, P., and L. Legendre. 1998. *Numerical ecology*. Elsevier.
- Litchman, E., and C. A. Klausmeier. 2008. Trait-based community ecology of phytoplankton. *Annu. Rev. Ecol. Evol. Syst.* **39**: 615–639. doi:10.1146/annurev.ecolsys.39.110707.173549
- Lombard, F., and others. 2019. Globally consistent quantitative observations of planktonic ecosystems. *Front. Mar. Sci.* **6**: 196. doi:10.3389/fmars.2019.00196
- Marañón, E. 2015. Cell size as a key determinant of phytoplankton metabolism and community structure. *Ann. Rev. Mar. Sci.* **7**: 241–264. doi:10.1146/annurev-marine-010814-015955
- Martini, S., and others. 2021. Functional trait-based approaches as a common framework for aquatic ecologists. *Limnol. Oceanogr.* **66**: 965–994. doi:10.1002/lno.11655
- Martins, T. G., D. Simpson, F. Lindgren, and H. Rue. 2013. Bayesian computing with INLA: New features. *Comput. Stat. Data Anal.* **67**: 68–83. doi:10.1016/j.csda.2013.04.014
- Naselli-Flores, L., and R. Barone. 2011. Fight on plankton! Or, phytoplankton shape and size as adaptive tools to get ahead in the struggle for life. *Cryptogam. Algal.* **32**: 157–204. doi:10.7872/crya.v32.iss2.2011.157
- Naselli-Flores, L., T. Zohary, and J. Padišák. 2021. Life in suspension and its impact on phytoplankton morphology: An homage to Colin S. Reynolds. *Hydrobiologia.* **848**: 7–30 doi:10.1007/s10750-020-04217-x
- Olson, R. J., and H. M. Sosik. 2007. A submersible imaging-inflow instrument to analyze nano-and microplankton: Imaging FlowCytobot: In situ imaging of nano- and microplankton. *Limnol. Oceanogr. Methods* **5**: 195–203. doi:10.4319/lom.2007.5.195
- Padišák, 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, p. 243–257. *In* *Aquatic biodiversity*. Springer.
- Pahlow, M., U. Riebesell, and D. A. Wolf-Gladrow. 1997. Impact of cell shape and chain formation on nutrient acquisition by marine diatoms. *Limnol. Oceanogr.* **42**: 1660–1672. doi:10.4319/lo.1997.42.8.1660
- Pratt, D. M. 1959. The phytoplankton of Narragansett Bay. *Limnol. Oceanogr.* **4**: 425–440. doi:10.4319/lo.1959.4.4.0425
- R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing.
- Narragansett Bay National Estuarine Research Reserve. 2009. An ecological profile of the Narragansett Bay National Estuarine Research Reserve. K.B. Raposa and M.L. Schwartz (eds.). Rhode Island Sea Grant, Narragansett, RI, 176pp.
- Reed, L., and C. Oviatt. 1976–2019. Marine Ecosystem Research Laboratory, Graduate School of Oceanography, URI, Narragansett, RI. <https://web.uri.edu/gso/research/marine-ecosystems-research-laboratory/datasets/>

- Reynolds, C. S., V. Huszar, C. Kruk, L. Naselli-Flores, and S. Melo. 2002. Towards a functional classification of the freshwater phytoplankton. *J. Plankton Res.* **24**: 417–428. doi:10.1093/plankt/24.5.417
- Roman, C. T., N. Jaworski, F. T. Short, S. Findlay, and S. Warren. 2000. Estuaries of the northeastern United States: Habitat and land use signatures. *Estuaries* **23**: 743. doi:10.2307/1352997
- Rue, H., S. Martino, and N. Chopin. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71**: 319–392. doi:10.1111/j.1467-9868.2008.00700.x
- Ruiz-Cárdenas, R., E. T. Krainski, and H. Rue. 2012. Direct fitting of dynamic models using integrated nested Laplace approximations—INLA. *Comput. Stat. Data Anal.* **56**: 1808–1828. doi:10.1016/j.csda.2011.10.024
- Salmaso, N., L. Naselli-Flores, and J. Padisák. 2015. Functional classifications and their application in phytoplankton ecology. *Freshw. Biol.* **60**: 603–619. doi:10.1111/fwb.12520
- Sin, Y. 2000. Seasonal variations of size-fractionated phytoplankton along the salinity gradient in the York River estuary, Virginia (USA). *J. Plankton Res.* **22**: 1945–1960. doi:10.1093/plankt/22.10.1945
- Sosik, H. M., and R. J. Olson. 2007. Automated taxonomic classification of phytoplankton sampled with imaging-in-flow cytometry: Phytoplankton image classification. *Limnol. Oceanogr. Methods* **5**: 204–216. doi:10.4319/lom.2007.5.204
- Stoyneva, M. P., J.-P. Descy, and W. Vyverman. 2007. Green algae in Lake Tanganyika: Is morphological variation a response to seasonal changes? *Hydrobiologia* **578**: 7–16. doi:10.1007/s10750-006-0428-1
- Strecker, A. L., T. P. Cobb, and R. D. Vinebrooke. 2004. Effects of experimental greenhouse warming on phytoplankton and zooplankton communities in fishless alpine ponds. *Limnol. Oceanogr.* **49**: 1182–1190. doi:10.4319/lo.2004.49.4.1182
- The MathWorks, Inc. 2019. MATLAB Release 2019a. Natick, Massachusetts, United States.
- Vilgrain, L., F. Maps, M. Picheral, M. Babin, C. Aubry, J. Irsson, and S. Ayata. 2021. Trait-based approach using in situ copepod images reveals contrasting ecological patterns across an Arctic ice melt zone. *Limnol. Oceanogr.* **66**: 1155–1167. doi:10.1002/lno.11672
- Walsby, A. E., and A. Xypolyta. 1977. The form resistance of chitan fibres attached to the cells of *Thalassiosira fluviatilis* Hustedt. *Br. Phycol. J.* **12**: 215–223. doi:10.1080/00071617700650231
- Weithoff, G., and B. E. Beisner. 2019. Measures and approaches in trait-based phytoplankton community ecology—From freshwater to marine ecosystems. *Front. Mar. Sci.* **6**: 40. doi:10.3389/fmars.2019.00040
- West, M., and J. Harrison. 1997. Bayesian forecasting and dynamic models. Springer-Verlag.
- Wickham, H., M. Averick, J. Bryan, and others. 2019. Welcome to the Tidyverse. *JOSS* **4**: 1686. doi: 10.21105/joss.01686
- Windecker, L. A. 2010. Ten Years of Phytoplankton Species Abundance Patterns in Mid-Narragansett Bay, Rhode Island: 1999–2008. M.S. Thesis. Univ. of Rhode Island.
- Zohary, T., T. Fishbein, M. Shlichter, and L. Naselli-Flores. 2017. Larger cell or colony size in winter, smaller in summer—A pattern shared by many species of Lake Kinneret phytoplankton. *Inland Waters* **7**: 200–209. doi:10.1080/20442041.2017.1320505

#### Acknowledgments

Additional Pier data were provided by Heather Stoffel (temperature and salinity, GSO, URI), Laura Weber (nutrients, GSO, URI), and zooplankton data by Nicole Flecchia. The authors thank the numerous people that maintained the IFCB data record over the years—Audrey Ciochetto, Kyle Turner, Jessica Carney, Chris Jenkins, Ian Lew, and Cassandra Alexander—as well as Jean-Olivier Irsson (Sorbonne University, France) for his input on aspects of the methodology and the computational analysis. We also thank the two anonymous reviewers and the associate editors for their detailed and helpful comments on improving the manuscript. This material is based upon work supported in part by the National Science Foundation under EPSCoR Cooperative Agreement #OIA-1655221 and the NASA FINESST award # 80NSSC20K1638. This work was initiated during VS's Master's degree internship supported by the LOV (Sorbonne University/CNRS). French co-authors wish to thank public taxpayers who fund their salaries. SDA acknowledges funding from CNRS for her sabbatical in 2018-2020. Additional support was provided by the Institut des Sciences du Calcul et des Données (ISCD) of Sorbonne Université through the support of the sponsored junior team FORMAL (From Observing to Modeling ocean Life).

#### Conflict of Interest

None declared.

Submitted 07 July 2021

Revised 13 December 2021

Accepted 22 May 2022

Associate editor: Michael R. Stukel