Contents lists available at ScienceDirect

Harmful Algae

journal homepage: www.elsevier.com/locate/hal

Sub-monthly prediction of harmful algal blooms based on automated cell imaging

Vitul Agarwal^{a,*}, Jonathan Chávez-Casillas^b, Colleen B. Mouw^a

^a Graduate School of Oceanography, University of Rhode Island, Narragansett, United States of America
^b Department of Mathematics and Applied Mathematical Sciences, University of Rhode Island, Kingston, United States of America

ARTICLE INFO

Edited by Dr. C. Gobler

Keywords: Imaging FlowCytobot (IFCB) Empirical dynamic modeling (EDM) Narragansett Bay Phytoplankton population dynamics Ecological forecasting Pseudo-nitzschia Dinophysis

ABSTRACT

Harmful algal blooms (HABs) are an increasing threat to global fisheries and human health. The mitigation of HABs requires management strategies to successfully forecast the abundance and distribution of harmful algal taxa. In this study, we attempt to characterize the dynamics of 2 phytoplankton genera (Pseudo-nitzschia spp. and Dinophysis spp.) in Narragansett Bay, Rhode Island, using empirical dynamic modeling. We utilize a highresolution Imaging FlowCytobot dataset to generate a daily-resolution time series of phytoplankton images and then characterize the sub-monthly (1-30 days) timescales of univariate and multivariate prediction skill for each taxon. Our results suggest that univariate predictability is low overall, different for each taxon and does not significantly vary over sub-monthly timescales. For all univariate predictions, models can rely on the inherent autocorrelation within each time series. When we incorporated multivariate data based on quantifiable image features, we found that predictability increased for both taxa and that this increase was apparent on timescales >7 days. Pseudo-nitzschia spp. has distinctive predictive dynamics that occur on timescales of around 16 and 25 days. Similarly, Dinophysis spp. is most predictable on timescales of 25 days. The timescales of prediction for Pseudo-nitzschia spp. and Dinophysis spp. could be tied to environmental drivers such as tidal cycles, water temperature, wind speed, community biomass, salinity, and pH in Narragansett Bay. For most drivers, there were consistent effects between the environmental variables and the phytoplankton taxon. Our analysis displays the potential of utilizing data from automated cell imagers to forecast and monitor harmful algal blooms.

1. Introduction

Harmful algal blooms (HABs) are anomalous increases in phytoplankton abundance, biomass, or distribution that can negatively affect marine ecosystems and public health (Fleming et al., 2011; Berdalet et al., 2016; Karlson et al., 2021). The rising frequency of such events in the past few decades is of increasing global concern (Xiao et al., 2019; Gobler 2020). Some estimates of economic damage due to HABs exceed hundreds of millions of dollars (Anderson et al., 2000), often due to fisheries closures (Brown et al., 2020; Sakamoto et al., 2021), disruption to tourism (Smith et al., 2019; Béchard 2020) and damage to human health (Grattan et al., 2016; Kouakou and Poder 2019). Consequently, the successful prediction and mitigation of HABs is a research priority for state and national governments worldwide (Park et al., 2013; Brooks et al., 2016).

HAB predictions often require large amounts of data from various sources and sophisticated modeling techniques (Franks 2018; Ralston

and Moore 2020), as well as detailed information on local and regional oceanographic features (Anderson et al., 2010; Dippner et al., 2011; Lapucci et al., 2022). Due to the requirement of high resolution and consistent data, monitoring programs are implementing automated systems (Babin et al., 2005; Jochens et al., 2010) with extensively trained algorithms (Sosik and Olson 2007; Ellen et al., 2019; Orenstein et al., 2020) that can identify and alert local officials of the presence, abundance and risk of HAB development. The rapid deployment of such systems has greatly expanded the ability to detect HABs; however, less is known about the utility of imaging data for HAB prediction models.

In this study, we explored the use of phytoplankton imaging data for HAB predictions in Narragansett Bay, Rhode Island (NBay). Narragansett Bay is a shallow coastal marine estuary of great cultural, historical, and economic importance to local communities (Herndon and Sekatau 1997; Dalton et al., 2010; Nixon and Fulweiler 2012). Coastal marine estuaries are highly dynamic environments that are subject to seasonality (Carstensen et al., 2015), the influence of both freshwater and

https://doi.org/10.1016/j.hal.2023.102386

Received 1 November 2022; Received in revised form 4 January 2023; Accepted 13 January 2023 Available online 1 February 2023 1568-9883/© 2023 Elsevier B.V. All rights reserved.







^{*} Corresponding author. E-mail address: vitulagarwal@uri.edu (V. Agarwal).

Table 1

Sensitivity and precision of the automatic classifier for each of the phytoplankton classes.

	Sensitivity	Precision
Pseudo-nitzschia spp. ($N = 626$)	0.85	0.94
Dinophysis spp. $(N = 412)$	0.95	0.96

Table 2

List of all image features used in this study and their units.

Feature	Units
area	pixels ²
biovolume	pixels ³
major axis	pixels
minor axis	pixels
perimeter	pixels
orientation	degrees
eccentricity	-
solidity	-
texture uniformity	-
texture smoothness	-
texture gray	-
texture entropy	-
texture contrast	-
h90	pixels
h180	pixels
hflip	pixels
extent	pixels
equivalent diameter	pixels
convex area	pixels ²
convex perimeter	pixels

seawater sources (Pilson 1985), large-scale climate phenomena (Scavia et al., 2002), and anthropogenic inputs of nutrients (Cundell 1973; Goldberg et al., 1977). Within the past decade, toxic harmful algal blooms have led to fisheries closures in Narragansett Bay (Bates et al.,

2018) and motivated extensive research into the potential environmental drivers and species composition (Sterling et al., 2022) of the relevant bloom-causing phytoplankton genera. Of particular importance in this area are *Pseudo-nitzschia spp*. and *Dinophysis spp*. due to their potential toxicity and relevance for local fishery disruption.

Natural phytoplankton populations are variable from daily, seasonal to decadal timescales (Chavez et al., 2003; Barton et al., 2016; Blauw et al., 2018). High variability in natural phytoplankton populations is characteristic of non-linear and chaotic dynamics (Ascioti et al., 1993; Smayda 1998). In this paper, we used empirical dynamic modeling (EDM) to predict the abundance of *Pseudo-nitzschia spp.* and *Dinophysis spp.* in Narragansett Bay, Rhode Island. These genera were selected for a couple of reasons: their role in local harmful algal blooms and the availability of a dynamic, regular time series that would allow for the use of data-driven analyses. EDM is a non-parametric framework that can avoid the pitfalls of typical statistical modeling by relying on data-driven attractor reconstruction (Perretti et al., 2013; Chang et al., 2017).

Our goal was to characterize the sub-monthly univariate and multivariate prediction timescales of Pseudo-nitzschia spp. and Dinophysis spp. utilizing a high temporal resolution dataset generated with an Imaging FlowCytobot (Olson and Sosik 2007). Automated instruments such as the Imaging FlowCytobot generate datasets of phytoplankton images and many associated features (i.e. image texture, contrast, object size etc.). Specifically, we aim to answer (1) How predictable are the harmful algal target species? (2) How does this predictability vary with time? and (3) Which image features best describe the dynamics of the taxa? Once we identified the important timescales, we also linked specific environmental drivers to the dynamics of the phytoplankton populations. Our study did not attempt to offer detailed mechanistic explanations of observed phenomena, nor develop tools that might model the growth and termination of harmful algal blooms, but instead, it focused on identifying the potential of imaging data in prediction models. By identifying the relevant dynamical timescales of harmful algal blooms, we also hoped to provide local and regional management



Fig. 1. Time series of 2 harmful algal bloom-forming taxa in Narragansett Bay, Rhode Island (left column) and their associated autocorrelation functions (ACF; right column). Relative abundance is estimated from the number of unique images taken by the IFCB and classified as (a) *Pseudo-nitzschia spp.* and (c) *Dinophysis spp.* Autocorrelation decreases with time and varies depending on the dynamics of each specific taxon.



Fig. 2. Univariate prediction skill ($\Delta\rho$) of the time series of (a) *Pseudo-nitzschia spp.* and (b) *Dinophysis spp.* over a prediction horizon of 1–30 days. Model predictions (ρ_{model}) were calculated from 200 random libraries of 250 days each and the results were reported as an arithmetic mean with 95% confidence intervals ($\pm 1.96 \times S.E.$). $\Delta\rho$ was calculated by subtracting the autocorrelation coefficient at each prediction horizon.

with a critical timeframe of action for the development of environmental policy. Our underlying assumption was that the predictability of *Pseudo-nitzschia spp.* and *Dinophysis spp.* in Narragansett Bay had distinct timescales that varied in response to environmental drivers and intrinsic population dynamics.

2. Materials and methods

2.1. Automated cell imaging

All the time series used in this study were collected by deploying an Imaging FlowCytobot (IFCB) in Narragansett Bay, Rhode Island. The IFCB is an automated, flow-through imaging system that captures images of the extant phytoplankton community in seawater. The system has a maximum size limit of 150 μ m and works by drawing water at approximately 1 m under the surface at low tide. As our IFCB was deployed at the end of a pier (41.492°N, 71.419°W), the actual sampling depth varied with the tidal cycle. Images can be observed in real-time using the IFCB dashboard (http://ifcb-dashboard.gso.uri.edu/).

The IFCB samples approximately every 20 min depending on the number of cells within a given sample. We used daily aggregated IFCB data from 14th June 2017 to 20th October 2021, barring gaps in the time series due to equipment malfunction or maintenance. Our data span

1590 days of observation with 518 days of missing data. For prediction tasks, all missing data points were approximated using 30-day exponential moving averages (EMA) computed by the R package "imputeTS" (Moritz and Bartz-Beielstein 2017). We used daily aggregated data, instead of other shorter timescales (such as 1-hr or 12-hr) for three broad reasons: the influence of high time series autocorrelation, irregular gaps in data collection, and, to strike a balance between computational costs and expected analytical benefit.

A machine-learning approach was used to identify and classify the phytoplankton taxa from a subset of annotated images (Sosik and Olson 2007). All obtained images classified as *Pseudo-nitzschia spp*. and *Dinophysis spp*. were counted and reported as a concentration based on the average sampling volume for each day (*images* mL^{-1}). Higher concentrations of images act as a proxy of higher abundance in the natural environment and lower concentrations of images show that the taxon is rare/absent. To test the general ability to use image concentration as a proxy for phytoplankton abundance, we visually compared our image concentration time series to a long-term weekly monitoring site located approximately 12 km north of our IFCB location (https://web.uri.edu/gso/research/plankton/). Figure S1 highlights that our IFCB image concentration agreed with the general pattern of *Pseudo-nitzschia spp*. abundance in Narragansett Bay (as determined by microscopy counts) over the duration of our time series.

We evaluated the classifier's performance for sensitivity and precision with a manually annotated library of images. Table 1 reports the performance of the classifier for *Pseudo-nitzschia spp*. and *Dinophysis spp*.

$$Sensitivity = \frac{TP}{TP + FN}$$
$$Precision = \frac{TP}{TP + FP}$$

where TP, FP and FN were the number of true positive, false positive, and false negative images.

236 image features are automatically estimated for each IFCB image (Sosik and Olson 2007, https://github.com/hsosik/ifcb-analysis/wiki). We selected 20 features for further analysis based on their relevance to phytoplankton morphology and ecology (Sonnet et al., 2022). The average daily values for the image features, scaled by the average sampling volume for each day, formed a multi-dimensional time series for each taxon. Table 2 lists all the features and their units.

2.2. Environmental data

We compiled data from various monitoring programs located in and around Narragansett Bay. Daily averages of water temperature (°*C*), salinity (*ppt*), chlorophyll (μ g L⁻¹), and pH were requested from the Narragansett Bay Fixed Site Monitoring Network (NBFSMN, personal communication: Heather Stoffel). These measurements were co-located with the Imaging FlowCytobot. Daily averages of wind speed ($m s^{-1}$) were drawn from the Kingston weather station (41.49°N 71.54°W; U.S. Climate Reference Network; https://www1.ncdc.noaa.gov/pub/dat a/uscrn/products/subhourly01/). Daily averages of tidal height (Mean Sea Level; *m*) were calculated from measurements at the NOAA Quonset Point Buoy (41° 35.2 N, 71° 24.6 W; #8454049; https://tidesandcurren ts.noaa.gov/). Additional environmental data, such as ambient nutrient concentrations, were not available at the same temporal scale as the imaging data used in the study.

2.3. Univariate predictions

We used empirical dynamic modeling (EDM) to make univariate predictions for the time series of each taxon. Every time series was normalized (i.e. subtracting the mean value of the time series and dividing by the standard deviation of the time series) before the appli-



Fig. 3. Multivariate prediction skill of the time series of Pseudo-nitzschia spp. (left column) and Dinophysis spp. (right column) over a prediction horizon of 1-30 days. (a) and (c) report prediction skill ($\Delta \rho$) calculated by subtracting the autocorrelation coefficient at each prediction horizon, (b) and (d) report prediction error as the rootmean-squared-error (RMSE), (e) and (f) report prediction error as the meanabsolute-error (MAE). Model results were calculated from 500 embeddings of phytoplankton abundance and 3 unique image features. The results were reported as an arithmetic mean with 95% confidence intervals ($\pm 1.96 \times S.E.$).

Fig. 4. Multivariate prediction skill of the time series of *Pseudo-nitzschia spp.* (ρ ; left) over a prediction horizon of 1–30 days. Prediction skill (ρ) refers to the Pearson correlation coefficient between model predictions and actual observations. Model results were calculated from 1000 embeddings of phytoplankton abundance and 3 unique image features. Each point is the outcome of a single model run. Frequency of image features (right) summarizes the top 5% of model outcomes and the image features included in these models.



Fig. 5. Multivariate prediction skill of the time series of *Dinophysis spp.* (ρ ; left) over a prediction horizon of 1–30 days. Prediction skill (ρ) refers to the Pearson correlation coefficient between model predictions and actual observations. Model results were calculated from 1000 embeddings of phytoplankton abundance and 3 unique image features. Each point is the outcome of a single model run. Frequency of image features (right) summarizes the top 5% of model outcomes and the image features included in these models.

cation of EDM. Specifically, we relied on simplex projection (Sugihara and May 1990) with a consistent embedding dimension of 4 and leave-one-out cross validation. This means that the univariate attractor of a time series was embedded in a 4-dimensional space using the original times series and successive lags of the same time series. Each point is described by $(x_t, x_{t-1}, x_{t-2}, x_{t-3})$ where x_t is the value of x at time point t, x_{t-1} is its value at time t-1, and so on. The embedding dimension was set at 4 to prevent overfitting and maximize the utilization of our daily-scale time series data. Figure S2 shows how varying the embedding dimension affects the predictability of each taxon with fixed prediction intervals of 1, 7, 14 and 28 days. An embedding dimension of 4 allowed for reasonable descriptions across all timescales for both taxa, with a lower risk of overfitting our models to potentially noisy dynamics. The model creation and prediction sets were randomly selected from the entire time series in intervals of 250 days. After bootstrapping 200 samples for each taxon, we generated a mean prediction and 95% confidence intervals (1.96 \times SE). By randomizing the selection of model and prediction libraries, we tried to account for the effects of imputed data into the time series, as well as avoid the effects of possible non-stationarity over the 1590 days of data.

We tested the predictability of each taxon for timescales of 1 to 30 days. Predictability was described by ρ_{model} , the Pearson correlation coefficient, between the observed and the predicted values after attractor reconstruction. To account for inherent autocorrelation within each time series, we subtracted the absolute value of the autocorrelation coefficient at each timescale of prediction. The effective value of predictability was reported as $\Delta \rho$, which is the arithmetic difference of the univariate predictability ρ_{model} and the autocorrelation coefficient ρ_{auto} . Therefore, $\Delta \rho$ quantifies the ability of our model to predict dynamics beyond autocorrelation across a range of sub-monthly timescales. Due to the short timescales of prediction in this study (<30 days), our dataset of 1590 days provided reasonable coverage of all possible sub-monthly dynamics for these harmful algal taxa.

2.4. Multiview embeddings (MVE)

Multiview embeddings are an effective technique for increasing predictability and drawing out information from multiple related time series (Ye and Sugihara 2016). We used MVE to utilize the associated dataset of image features collected by the IFCB. Once more, the embedding dimension was set to 4 for all taxa and the entire time series was used for model and prediction libraries. We relied on leave-one-out cross-validation instead of separate model and prediction libraries.

Each multivariate attractor was created by randomly selecting 3 normalized time series of features and the original time series of image concentration (*images mL*⁻¹). Our goal was to predict the proxy abundance of each taxon by leveraging information stored in the image features. Predictability was evaluated for timescales of 1–30 days and reported as $\Delta \rho$ (model predictability beyond autocorrelation), RMSE (root-mean-square error) and MAE (mean absolute error). We considered 500 trials of image feature combinations and reported predictability as the arithmetic mean with 95% confidence intervals (1.96 × *SE*).

For the best multivariate models (top 5% in terms of ρ_{model}), we reported the frequency of appearance for each image feature as a proportion. A proportion of 0 implies that the feature did not show up in the best multivariate models while a proportion of 1 implies that it was always present. Based on the frequency of appearance, we could deduce the contribution of each feature in improving the overall predictability of the phytoplankton species.

2.5. Convergent cross mapping (CCM)

Once we identified any relevant timescales of prediction, we wanted to understand whether there was a link between the abundance of harmful algal taxa and relevant environmental drivers. We used convergent cross mapping (CCM; Sugihara et al., 2012) to infer causation between the environmental dataset and image concentration (*images mL*⁻¹). Embedding dimensions were optimized (i.e. selecting the embedding dimension that provides the highest prediction skill ρ) to each environmental variable (up to a maximum of 7 to prevent overfitting) and library sizes ranged from 100 - 1400 in intervals of 100 days. There were 20 samples each for every library size and the time to prediction ranged from 1–30 days. We tested whether we could infer causation by predicting the values of past environmental variables from the abundance of the harmful algal taxa. Predictability was quantified



Pseudo-nitzschia spp.

Fig. 6. Influence of environmental drivers on *Pseudo-nitzschia spp*. in Narragansett Bay quantified by the cross-map prediction skill (ρ based on convergent cross mapping; see Methods). The influence was measured over a prediction horizon of 1–30 days (black line). Red points indicate which models showed convergence. The dashed line refers to the Pearson correlation coefficient between the time series of *Pseudo-nitzschia spp*. abundance and the environmental variable.

by the cross-map prediction skill (ρ), where higher values indicate better predictions. Convergence was estimated using three tests – (1) Mann-Kendall trend test for ρ with increasing library size, (2) a Student's *t*-test for the ρ distributions at the maximum and minimum library size and (3) by validating that the prediction skill ρ at the maximum library size was greater than the Pearson correlation coefficient between image concentration and the environmental time series. Only the predictions which satisfied all conditions, and were significant for both the Mann-Kendall and the Student's *t*-test (p-value < 0.05) were deemed convergent. If any of the tests failed, then the causal effect of the environmental variable on the phytoplankton taxa was deemed to be unresolved at those specific timescales. Cross-map prediction skill (ρ) was normalized to the embedding dimension by averaging ρ across prediction horizons (Saberski et al., 2021).

2.6. Software

All the analyses were conducted in R (R Core Team 2021). For plotting and data visualization, we used the R packages "ggplot2" (Wickham 2016) and "cowplot" (Wilke 2020). EDM was applied using pre-built functions in the R package "rEDM" (Park et al., 2022). Additionally, the R package "Kendall" (McLeod 2022) was used to conduct some statistical tests.



Fig. 7. Influence of environmental drivers on *Dinophysis spp*. in Narragansett Bay quantified by the cross-map prediction skill (ρ based on convergent cross mapping; see Methods). The influence was measured over a prediction horizon of 1–30 days (black line). Red points indicate which models showed convergence. The dashed line refers to the Pearson correlation coefficient between the time series of *Dinophysis spp*. abundance and the environmental variable.

3. Results

Both *Pseudo-nitzschia spp.* and *Dinophysis spp.* in Narragansett Bay showed intermittent periods of high and low abundance in Narragansett Bay. The IFCB captured such bloom dynamics using the concentration of identified images of both taxa (Fig. 1; left column). When evaluated for the autocorrelation inherent within each time series, both *Pseudo-nitzschia spp.* and *Dinophysis spp.* had decreasing autocorrelation with time (Fig. 1; right column). The decrease was more rapid for *Pseudo-nitzschia spp.* (ACF < 0.25 within 7 days), whereas *Dinophysis spp.* showed a more gradual decrease over the entire 30 days.

The univariate predictability of both time series ($\Delta \rho$) was low overall and did not greatly change over a prediction horizon of 30 days (Fig. 2).

The univariate predictability of *Pseudo-nitzschia spp*. indicated some promise of the model over autocorrelation on horizons of 10 days, whereas the results for *Dinophysis spp*. indicated that there is little to no predictability inherent within the time series beyond autocorrelation across all sub-monthly timescales.

Multivariate prediction skill ($\Delta \rho$), calculated using the time series of image abundance and 3 associated image features, was much higher than the univariate prediction skill for both *Pseudo-nitzschia spp*. and *Dinophysis spp*. (Fig. 3). Using the original time series with only 3 image features at a time (multivariate embedding dimension = 4) allowed for direct comparisons to the univariate prediction skill. The predictability of *Pseudo-nitzschia spp*. had distinctive cycles with peaks every 16 and 25 days. An increase in model predictability over autocorrelation was most

prominent after a 3-day prediction horizon. The predictability of *Dinophysis spp*. was also higher than inherent autocorrelation after a 6-day prediction horizon. *Dinophysis spp*. had multiple peaks in $\Delta \rho$ at prediction horizons of 10–20 days, with a distinct peak at 25 days. The RMSE of the predictions also showed a distinct drop around the 25-day mark.

Without accounting for autocorrelation, some multivariate models for both *Pseudo-nitzschia spp.* and *Dinophysis spp.* reached prediction skills of 0.70 and greater (Fig. 4 & 5; left). The top 5% of these multivariate models had a fairly uniform distribution of image features, with some clear exceptions. The time series of biovolume and orientation prominently appeared in the top multivariate models for *Pseudo-nitzschia spp.*, whereas the time series of solidity and hflip were present but in a distinctly lower frequency compared to the other features. For *Dinophysis spp.*, the time series of biovolume, texture gray and orientation were prominently present in the top multivariate models.

Environmental influence on the populations of *Pseudo-nitzschia spp.* and *Dinophysis spp.*, as measured by the cross-map prediction skill (ρ), showed variable effects across different prediction horizons (Figure 6 and 7). For *Pseudo-nitzschia spp.*, the prediction skill of all the environmental parameters converged with increasing library size. The strength and timescales of inferred causal influence differed across the variables. The influence of chlorophyll on *Pseudo-nitzschia spp.* had a peak at timescales around 28 days, whereas the influence of tidal height was strongest around 2 weeks. The time series of water temperature, pH, wind speed, and salinity showed significant and consistent effects on the *Pseudo-nitzschia spp.* time series across most sub-monthly timescales.

For *Dinophysis spp.*, there were more models that showed convergence across all prediction horizons. The time series of *Dinophysis spp.* was consistently affected by tide height, water temperature, and pH across all timescales. Total biomass (chlorophyll) appeared as a significant driver of *Dinophysis spp.* with peaks around 5, 14 and 27 days. The effects of salinity were consistent and stronger in the short-term (1–3 days). Predictability was higher than the Pearson correlation coefficient for most environmental variables and showed consistency, which might suggest specific mechanisms of causal influence.

4. Discussion

4.1. Predictability of Pseudo-nitzschia spp. and Dinophysis spp.

Perturbations in phytoplankton population dynamics typically decorrelate within timescales of a month (Kuhn et al., 2019). When we tested for the inherent autocorrelation within the time series of both Pseudo-nitzschia spp. and Dinophysis spp., we found that the autocorrelation decreased significantly within the first 10 days for Pseudo-nitzschia spp., but Dinophysis spp. had higher autocorrelation for up to 30 days. After accounting for autocorrelation, the univariate predictability of both Pseudo-nitzschia spp. and Dinophysis spp. was low overall; however, the univariate predictability of Pseudo-nitzschia spp. showed some cyclical behavior. Our univariate models likely picked up on repetitive population-level mechanisms that increased or decreased abundance on sub-monthly timescales. Some examples of such mechanisms could include regular switching between periods of growth and sexual reproduction (D'Alelio et al., 2009; Annunziata et al., 2022), density-dependent interactions with parasitic protists (Berdjeb et al., 2018), or the tidal transport of productive populations from nearby sites (Shanks et al., 2014). Part of the lack of univariate predictability could be due to the presence of measurement error and stochasticity in the time series of both taxa, as well as a general lack of natural predictability for larger diatoms and dinoflagellates (Agarwal et al., 2021).

In the multivariate case, we found the predictability of both *Pseudo-nitzschia spp.* and *Dinophysis spp.* improved on timescales of greater than 1 week. Multiview embeddings have been previously shown to improve the univariate predictability of short time series (Ye and Sugihara 2016). By leveraging information stored across multiple related image features,

our approach of randomly creating non-lagged embeddings could have allowed us to create better and more reliable estimates of predictive dynamics (Ma et al., 2018). The cyclical predictability of *Pseudo-nitzschia spp.* was more prominent in the multivariate models, implicating predictable behavior on 16-day and 25-day timescales. *Dinophysis spp.* was most predictable on timescales of 25 days. Due to the presence of distinct timescales of predictability for both taxa, our results suggest that future development of HAB models would benefit by resolving dynamics on daily and weekly timescales. The identification of relevant ecological and environmental drivers of population dynamics on these timescales might also aid in the development of automated monitoring and early-warning systems.

4.2. Relative contribution of IFCB image features

When we evaluated the relative proportions of image features among the top multivariate models, the time series of biovolume was prominently present for *Pseudo-nitzschia spp.* and *Dinophysis spp.* This implies that the time series of biovolume adds considerable information to the future predictability of harmful algal taxa. Biovolume estimates from IFCB images (Moberg and Sosik 2012) are often used as an important marker of phytoplankton community structure and function (Brosnahan et al., 2015; Oliver et al., 2021). Although image-derived biovolume estimates might differ from microscopy-derived estimates (Kraft et al., 2021), cell biovolume typically varies linearly with other phytoplankton functional traits (Edwards et al., 2012). Our results suggest that including biovolume estimates and other high-performing image descriptors into models for harmful algal taxa improves predictability beyond autocorrelation.

Image descriptors derived from flow cytometers have found utility in studies of phytoplankton morphology (Sonnet et al., 2022), as well as for the training of different image classifiers (Mosleh et al., 2012; Zheng et al., 2017). In general, "features" from an IFCB image are all calculated from the pixels of the image and the relationships between them (see Table 2). As most features share the fundamental quantity underlying their calculations (i.e. the image itself), we expect all time series to be nonlinear approximations of one another. The relatively consistent proportions of most features in the top multivariate models indicate that the use of features themselves, and not necessarily their "character", increases the predictability of harmful algal taxa. Unless there is a particular reason to prefer one feature for another (such as biovolume for its relationship to other traits), prediction models relying on automated imaging systems would benefit from using any associated image data. Detailed information on the causal relationships between image features for Pseudo-nitzschia spp., as well as the partial correlations between each image feature and our time series of image concentration, can be found in the Supplemental material.

4.3. Potential environmental drivers

To further investigate the timescales of prediction for both Pseudonitzschia spp. and Dinophysis spp., we evaluated any inferred causal relationships between environmental drivers and the proxy abundance of each taxon. Consistent causal influence on either taxon would show variable but significant, effects across sub-monthly timescales. We found that both Pseudo-nitzschia spp. and Dinophysis spp. are affected by water temperatures, wind speed, tidal height, salinity, pH, and total biomass (chlorophyll). Previous studies across various regions, have hypothesized correlative relationships between harmful algal blooms and these environmental drivers (Almandoz et al., 2007; Sildever et al., 2019; Zhang et al., 2020; Lima et al., 2022). In our study, convergent model predictions with increasing library size, which greatly exceed the Pearson correlation coefficients between the individual time series, indicate that there are causal relationships beyond simple covariance. None of the environmental drivers we tested overlapped with the previously quantified multivariate timescales. This implies that the

dynamics of both taxa in Narragansett Bay are subject to multiple context-dependent forces that interact with each other. Successful prediction models for *Pseudo-nitzschia spp.* and *Dinophysis spp.* would need to incorporate the specific local conditions under which the harmful algal blooms develop. An ideal prediction model would attempt to combine data from relevant image properties and environmental drivers for particular timescales of prediction. Different model combinations could be optimized for forecasting at certain points during the 30-day prediction horizon. Table S3 explores the outcomes of some illustrative combined models for both *Pseudo-nitzschia spp.* and *Dinophysis spp.* with a prediction horizon of 5, 10 and 15 days. Our results also indicate that there can be lags between an environmental driver and the driven harmful algal taxa. Future studies that attempt to predict the dynamics of *Pseudo-nitzschia spp.* and *Dinophysis spp.* might need to characterize the causal timescales of their predictors.

As there has been rapid deployment of automated imaging systems for the early detection of harmful algal bloom events (Campbell et al., 2010, 2013), our results suggest that there is potential to use such data sources in advanced prediction models. Monitoring programs that concurrently deploy other environmental and biogeochemical sensors might be able characterize the relevant timescales of dynamics, and consequently, predict the magnitude and spatial distribution of harmful algal events across broader regions. Although this study focuses on the population dynamics of the harmful algal bloom-causing taxa, our prediction models could also be coupled with other broad-scale ecosystem models to potentially include the impacts on higher trophic levels and human health.

4.4. Study limitations and future directions

Although we have demonstrated the potential of using automated cell imaging data in prediction models, there are several considerations involved that merit further discussion. First, as our sampling location is fixed, the influence of different water masses and a lack of spatial information can limit real-time projections of HAB abundance across entire regions. Future studies should consider the concurrent deployment of multiple different systems to accurately map and forecast spatial population patterns. Second, as Pseudo-nitzschia spp. is a chain-forming diatom, the use of image concentration is not a measure of the actual abundance of the taxon within the water column - there can be a variable number of cells within an image. Instead, image concentration is a measure of our ability to detect and identify the taxa. Although detection numbers are high when abundance is typically high (Figure S1), future studies might need to accurately quantify the relationship between the in-situ abundance of chain-forming organisms and their image detection. Third, the deployment and maintenance of IFCB systems may lead to some irregularities and gaps within a long-term time series. Despite multiple years of data collection, a large proportion of our daily-scale time series had to be approximated from existing observations. Our approach requires sufficient long-term coverage for the development of prediction models and future studies could evaluate alternative methods of data processing and interpolation of missing observations. Fourth, the development of harmful algal blooms likely depends on a suite of unknown environmental triggers (such as the nutrient regime, ambient light levels, etc.). The identification of specific causal mechanisms would depend on careful experimentation in laboratory studies, where confounding factors can be controlled, and additive influence can be disentangled.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This study was supported by the Rhode Island Sea Grant (Grant number: NA22OAR4170123) and NASA (Grant Supplement: NNX14AB80G). We would like to thank Heather Stoffel for the provision of environmental data at the sampling location and Audrey Ciochetto for managing the Imaging FlowCytobot data streams, instrument maintenance, image processing and databasing. We would also like to acknowledge the efforts of Jessica Carney and Virginie Sonnet for IFCB image annotations. The following students have helped maintain the IFCB over the years of operation: Kyle Turner, Ian Lew, Cassandra Alexander, Christopher Jenkins, Jessica Carney, Virginie Sonnet, and Somang Song.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.hal.2023.102386.

References

- Agarwal, V., James, C.C., Widdicombe, C.E., Barton, A.D., 2021. Intraseasonal predictability of natural phytoplankton population dynamics. Ecol. Evol. 11, 15720–15739. https://doi.org/10.1002/ece3.8234.
- Almandoz, G.O., Ferrario, M.E., Ferreyra, G.A., Schloss, I.R., Esteves, J.L., Paparazzo, F. E., 2007. The genus Pseudo-nitzschia (Bacillariophyceae) in continental shelf waters of Argentina (Southwestern Atlantic Ocean, 38-55°S). Harmful Algae 6, 93–103. https://doi.org/10.1016/j.hal.2006.07.003.
- Anderson, C.R., Sapiano, M.R.P., Prasad, M.B.K., Long, W., Tango, P.J., Brown, C.W., Murtugudde, R., 2010. Predicting potentially toxigenic Pseudo-nitzschia blooms in the Chesapeake Bay. J. Mar. Syst. 83, 127–140. https://doi.org/10.1016/j. jmarsys.2010.04.003.
- Anderson, D.M., P. Hoagland, Y. Kaoru, and A.W. White. 2000. Estimated annual economic impacts from harmful algal blooms (HABs) in the United States.
- Annunziata, R., Mele, B.H., Marotta, P., others, 2022. Trade-off between sex and growth in diatoms: molecular mechanisms and demographic implications. Sci. Adv. 8, 1–17. https://doi.org/10.1126/sciadv.abj9466.
- Ascioti, F.A., Beltrami, E., Carroll, T.O., Wirick, C., 1993. Is there chaos in plankton dynamics? J. Plankton Res. 15, 603–617. https://doi.org/10.1093/plankt/15.6.603.
- Babin, M., Cullen, J.J., Roesler, C.S., others, 2005. New approaches and technologies for observing harmful algal blooms. Oceanography 18, 210–227. https://doi.org/ 10.5670/oceanog.2005.55.
- Barton, A.D., Irwin, A.J., Finkel, Z.V., Stock, C.A., 2016. Anthropogenic climate change drives shift and shuffle in North Atlantic phytoplankton communities. Proc. Natl. Acad. Sci. U. S. A. 113, 2964–2969. https://doi.org/10.1073/pnas.1519080113.
- Bates, S.S., Hubbard, K.A., Lundholm, N., Montresor, M., Leaw, C.P., 2018. Pseudonitzschia, Nitzschia, and domoic acid: new research since 2011. Harmful Algae 79, 3–43. https://doi.org/10.1016/j.hal.2018.06.001.
- Béchard, A., 2020. Economics losses to fishery and seafood related businesses during harmful algal blooms. Fish. Res. 230, 105678 https://doi.org/10.1016/j. fishres.2020.105678.
- Berdalet, E., Fleming, L.E., Gowen, R., others, 2016. Marine harmful algal blooms, human health and wellbeing: challenges and opportunities in the 21st century. J. Mar. Biol. Assoc. United Kingdom 96, 61–91. https://doi.org/10.1017/ S0025315415001733.
- Berdjeb, L., Parada, A., Needham, D.M., Fuhrman, J.A., 2018. Short-term dynamics and interactions of marine protist communities during the spring-summer transition. ISME J. 12, 1907–1917. https://doi.org/10.1038/s41396-018-0097-x.
- Blauw, A.N., Benincà, E., Laane, R.W.P.M., Greenwood, N., Huisman, J., 2018. Predictability and environmental drivers of chlorophyll fluctuations vary across different time scales and regions of the North Sea. Prog. Oceanogr. 161, 1–18. https://doi.org/10.1016/j.pocean.2018.01.005.
- Brooks, B.W., Lazorchak, J.M., Howard, M.D.A., others, 2016. Are harmful algal blooms becoming the greatest inland water quality threat to public health and aquatic ecosystems? Environ. Toxicol. Chem. 35, 6–13. https://doi.org/10.1002/etc.3220.
- Brosnahan, M.L., Velo-Suárez, L., Ralston, D.K., others, 2015. Rapid growth and concerted sexual transitions by a bloom of the harmful dinoflagellate Alexandrium fundyense (Dinophyceae). Limnol. Oceanogr. 60, 2059–2078. https://doi.org/ 10.1002/lno.10155.
- Brown, A.R., Lilley, M., Shutler, J., Lowe, C., Artioli, Y., Torres, R., Berdalet, E., Tyler, C. R., 2020. Assessing risks and mitigating impacts of harmful algal blooms on mariculture and marine fisheries. Rev. Aquac. 12, 1663–1688. https://doi.org/ 10.1111/raq.12403.

Campbell, L., Henrichs, D.W., Olson, R.J., Sosik, H.M., 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. https://doi.org/ 10.1007/s11356-012-1437-4.

Campbell, L., Olson, R.J., Sosik, H.M., Abraham, A., Henrichs, D.W., Hyatt, C.J., Buskey, E.J., 2010. First harmful dinophysis (dinophyceae, dinophysiales) bloom in the U.S. is revealeD by automated imaging flow cytometry. J. Phycol. 46, 66–75. https://doi.org/10.1111/j.1529-8817.2009.00791.x.

Carstensen, J., Klais, R., Cloern, J.E., 2015. Phytoplankton blooms in estuarine and coastal waters: seasonal patterns and key species. Estuar. Coast. Shelf Sci. 162, 98–109. https://doi.org/10.1016/j.ecss.2015.05.005.

Chang, C.W., Ushio, M., hao Hsieh, C., 2017. Empirical dynamic modeling for beginners. Ecol. Res. 32, 785–796. https://doi.org/10.1007/s11284-017-1469-9.

Chavez, F.P., Ryan, J., Lluch-Cota, S.E., Niquen, C.M., 2003. Climate: from anchovies to sardines and back: multidecadal change in the Pacific Ocean. Science (80-.) 299, 217–221. https://doi.org/10.1126/science.1075880.

Cundell, A.M., 1973. Plastic materials accumulating in Narragansett Bay. Mar. Pollut. Bull. 4, 187–188. https://doi.org/10.1016/0025-326X(73)90226-9.

D'Alelio, D., Amato, A., Luedeking, A., Montresor, M., 2009. Sexual and vegetative phases in the planktonic diatom Pseudo-nitzschia multistriata. Harmful Algae 8, 225–232. https://doi.org/10.1016/j.hal.2008.05.004.

Dalton, T., Thompson, R., Jin, D., 2010. Mapping human dimensions in marine spatial planning and management: an example from Narragansett Bay, Rhode Island. Mar. Policy 34, 309–319. https://doi.org/10.1016/j.marpol.2009.08.001.

Dippner, J.W., Nguyen-Ngoc, L., Doan-Nhu, H., Subramaniam, A., 2011. A model for the prediction of harmful algae blooms in the Vietnamese upwelling area. Harmful Algae 10, 606–611. https://doi.org/10.1016/j.hal.2011.04.012.

Edwards, K.F., Thomas, M.K., Klausmeier, C.A., Litchman, E., 2012. Allometric scaling and taxonomic variation in nutrient utilization traits and maximum growth rate of phytoplankton. Limnol. Oceanogr. 57, 554–566. https://doi.org/10.4319/ lo.2012.57.2.0554.

Ellen, J.S., Graff, C.A., Ohman, M.D., 2019. Improving plankton image classification using context metadata. Limnol. Oceanogr. Methods 17, 439–461. https://doi.org/ 10.1002/lom3.10324.

Fleming, L.E., Kirkpatrick, B., Backer, L.C., others, 2011. Review of Florida red tide and human health effects. Harmful Algae 10, 224–233. https://doi.org/10.1016/j. hal.2010.08.006.

Franks, P.J.S. 2018. Recent advances in modelling of harmful algal blooms, p. 359–377. *In* global ecology and oceanography of harmful algal blooms.

Gobler, C.J., 2020. Climate Change and Harmful Algal Blooms: insights and perspective. Harmful Algae 91, 101731. https://doi.org/10.1016/j.hal.2019.101731.

Goldberg, E.D., Gamble, E., Griffin, J.J., Koide, M., 1977. Pollution history of Narragansett Bay as recorded in its sediments. Estuar.Coastal Mar.Sci 5, 549–558. https://doi.org/10.1016/0302-3524(77)90101-3.

Grattan, L.M., Holobaugh, S., Morris, J.G., 2016. Harmful algal blooms and public health. Harmful Algae 57, 2–8. https://doi.org/10.1016/j.hal.2016.05.003.

Herndon, R.W., Sekatau, E.W., 1997. The right to a name: the Narragansett people and Rhode island officials in the revolutionary era. Ethnohistory 44, 433–462.

Jochens, A.E., Malone, T.C., Stumpf, R.P., others, 2010. Integrated ocean observing system in support of forecasting harmful algal blooms. Mar. Technol. Soc. J. 44, 99–121. https://doi.org/10.4031/MTSJ.44.6.16.

Karlson, B., Andersen, P., Arneborg, L., others, 2021. Harmful algal blooms and their effects in coastal seas of Northern Europe. Harmful Algae 102, 101989. https://doi. org/10.1016/j.hal.2021.101989.

Kouakou, C.R.C., Poder, T.G., 2019. Economic impact of harmful algal blooms on human health: a systematic review. J. Water Health 17, 499–516. https://doi.org/10.2166/ wh.2019.064.

Kraft, K., Seppälä, J., Hällfors, H., others, 2021. First application of IFCB high-frequency imaging-in-flow cytometry to investigate bloom-forming filamentous cyanobacteria in the Baltic Sea. Front. Mar. Sci. 8, 1–17. https://doi.org/10.3389/ fmars/2021_594144

Kuhn, A.M., Dutkiewicz, S., Jahn, O., Clayton, S., Rynearson, T.A., Mazloff, M.R., Barton, A.D., 2019. Temporal and spatial scales of correlation in marine phytoplankton communities. J. Geophys. Res. Ocean. 124, 9417–9438. https://doi. org/10.1029/2019JC015331.

Lapucci, C., Maselli, F., Chini Zittelli, G., others, 2022. Towards the prediction of favourable conditions for the harmful algal bloom onset of ostreopsis ovata in the Ligurian sea based on satellite and model data. J. Mar. Sci. Eng. 10, 461. https://doi. org/10.3390/jmse10040461.

Lima, M.J., Relvas, P., Barbosa, A.B., 2022. Variability patterns and phenology of harmful phytoplankton blooms off southern Portugal: looking for region-specific environmental drivers and predictors. Harmful Algae 116, 102254. https://doi.org/ 10.1016/j.hal.2022.102254.

Ma, H., Leng, S., Aihara, K., Lin, W., Cheni, L., 2018. Randomly distributed embedding making short-term high-dimensional data predictable. Proc. Natl. Acad. Sci. U. S. A. 115, E9994–E10002. https://doi.org/10.1073/pnas.1802987115.

McLeod, A.I. 2022. Kendall: Kendall rank correlation and Mann-Kendall trend test. R package version 2.2.1. https://CRAN.R-project.org/package=Kendall.

Moberg, E.A., Sosik, H.M., 2012. Distance maps to estimate cell volume from twodimensional plankton images. Limnol. Oceanogr. Methods 10, 278–288. https://doi. org/10.4319/lom.2012.10.278.

Moritz, S., Bartz-Beielstein, T., 2017. imputeTS: time series missing value imputation in R. R J. 9 (1), 207–218. https://doi.org/10.32614/RJ-2017-009.

Mosleh, M.A.A., Manssor, H., Malek, S., Milow, P., Salleh, A., 2012. A preliminary study on automated freshwater algae recognition and classification system. BMC Bioinformatics 13 (Suppl 1). https://doi.org/10.1186/1471-2105-13-s17-s25. Nixon, S.W., Fulweiler, R.W., 2012. Ecological footprints and shadows in an urban estuary, Narragansett Bay, RI (USA). Reg. Environ. Chang. 12, 381–394. https://doi. org/10.1007/s10113-011-0221-1.

Oliver, H., Zhang, W.G., Smith, W.O., others, 2021. Diatom hotspots driven by western boundary current instability. Geophys. Res. Lett. 48, 1–10. https://doi.org/10.1029/ 2020GL091943.

Olson, R.J., Sosik, H.M., 2007. A submersible imaging-in-flow instrument to analyze nano-and microplankton: imaging FlowCytobot. Limnol. Oceanogr. Methods 5, 195–203. https://doi.org/10.4319/lom.2007.5.195.

Orenstein, E.C., Kenitz, K.M., Roberts, P.L.D., Franks, P.J.S., Jaffe, J.S., Barton, A.D., 2020. Semi- and fully supervised quantification techniques to improve population estimates from machine classifiers. Limnol. Oceanogr. Methods 18, 739–753. https://doi.org/10.1002/lom3.10399.

Park, T.G., Lim, W.A., Park, Y.T., Lee, C.K., Jeong, H.J., 2013. Economic impact, management and mitigation of red tides in Korea. Harmful Algae 30, S131–S143. https://doi.org/10.1016/j.hal.2013.10.012.

Park, Joseph, C. Smith, G. Sugihara and E. Deyle. 2022. rEDM: empirical dynamic modeling ('EDM'). R package version 1.13.0. https://CRAN.R-project.org/ package=rEDM.

Perretti, C.T., Munch, S.B., Sugihara, G., 2013. Model-free forecasting outperforms the correct mechanistic model for simulated and experimental data. Proc. Natl. Acad. Sci. U. S. A. 110, 5253–5257. https://doi.org/10.1073/pnas.1216076110.

Pilson, M.E.Q., 1985. On the residence time of water in Narragansett Bay. Estuaries 8, 2–14. https://doi.org/10.2307/1352116.

R Core Team, 2021. R: A language and Environment For Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL. https://www.R-project. org/.

Ralston, D.K., Moore, S.K., 2020. Modeling harmful algal blooms in a changing climate. Harmful Algae 91, 1–29. https://doi.org/10.1016/j.hal.2019.101729.

Saberski, E., Bock, A.K., Goodridge, R., Agarwal, V., Lorimer, T., Rifkin, S.A., Sugihara, G., 2021. Networks of causal linkage between eigenmodes characterize behavioral dynamics of caenorhabditis elegans. PLoS Comput. Biol. 17, 1–15. https://doi.org/10.1371/journal.pcbi.1009329.

Sakamoto, S., Lim, W.A., Lu, D., Dai, X., Orlova, T., Iwataki, M., 2021. Harmful algal blooms and associated fisheries damage in East Asia: current status and trends in China, Japan, Korea and Russia. Harmful Algae 102, 101787. https://doi.org/ 10.1016/j.hal.2020.101787.

Scavia, D., Field, J.C., Boesch, D.F., others, 2002. Climate change impacts on U.S. coastal and marine ecosystems. Estuaries 25, 149–164. https://doi.org/10.1007/ BF02691304.

Shanks, A.L., Morgan, S.G., MacMahan, J., Reniers, A.J.H.M., Reniers, M., Brown, J., Fujimura, A., Griesemer, C., 2014. Onshore transport of plankton by internal tides and upwelling-relaxation events. Mar. Ecol. Prog. Ser. 502, 39–51. https://doi.org/ 10.3354/meps10717.

Sildever, S., Kawakami, Y., Kanno, N., Kasai, H., Shiomoto, A., Katakura, S., Nagai, S., 2019. Toxic HAB species from the Sea of Okhotsk detected by a metagenetic approach, seasonality and environmental drivers. Harmful Algae 87, 101631. https://doi.org/10.1016/j.hal.2019.101631.

Smayda, T.J., 1998. Patterns of variability characterizing marine phytoplankton, with examples from Narragansett Bay. ICES J. Mar. Sci. 55, 562–573. https://doi.org/ 10.1006/jmsc.1998.0385.

Smith, R.B., Bass, B., Sawyer, D., Depew, D., Watson, S.B., 2019. Estimating the economic costs of algal blooms in the Canadian Lake Erie Basin. Harmful Algae 87, 101624. https://doi.org/10.1016/j.hal.2019.101624.

Sonnet, V., Guidi, L., Mouw, C.B., Puggioni, G., Ayata, S.D., 2022. Length, width, shape regularity, and chain structure: time series analysis of phytoplankton morphology from imagery. Limnol. Oceanogr. 67, 1850–1864. https://doi.org/10.1002/ lno.12171

Sosik, H.M., Olson, R.J., 2007. Automated taxonomic classification of phytoplankton sampled with imaging-in-flow cytometry. Limnol. Oceanogr. Methods 5, 204–216. https://doi.org/10.4319/lom.2007.5.204.

Sterling, A.R., Kirk, R.D., Bertin, M.J., others, 2022. Emerging harmful algal blooms caused by distinct seasonal assemblages of a toxic diatom. Limnol. Oceanogr. 0–1. https://doi.org/10.1002/lno.12189.

Sugihara, G., May, R.M., 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature 344, 24–26.

Sugihara, G., May, R., Ye, H., Hsieh, C.H., Deyle, E., Fogarty, M., Munch, S., 2012. Detecting causality in complex ecosystems. Science (80-.) 338, 496–500. https://doi. org/10.1126/science.1227079.

Wickham, H., 2016. ggplot2: Elegant Graphics For Data Analysis. Springer-Verlag, New York.

Wilke, Claus O. 2020. Cowplot: streamlined plot theme and plot annotations for 'ggplot2'.

Xiao, X., Agustí, S., Pan, Y., Yu, Y., Li, K., Wu, J., Duarte, C.M., 2019. Warming amplifies the frequency of harmful algal blooms with eutrophication in Chinese coastal waters. Environ. Sci. Technol. 53, 13031–13041. https://doi.org/10.1021/acs.est.9b03726.

Ye, H., Sugihara, G., 2016. Information leverage in interconnected ecosystems: overcoming the curse of dimensionality. Science (80-.) 353, 922–925. https://doi. org/10.1126/science.aag0863.

Zhang, Y., Su, J.Z., Su, Y.P., Lin, H., Xu, Y.C., Barathan, B.P., Zheng, W.N., Schulz, K.G., 2020. Spatial distribution of phytoplankton community composition and their correlations with environmental drivers in taiwan strait of Southeast China. Diversity (Basel) 12, 1–15. https://doi.org/10.3390/d12110433.

Zheng, H., Wang, R., Yu, Z., Wang, N., Gu, Z., Zheng, B., 2017. Automatic plankton image classification combining multiple view features via multiple kernel learning. BMC Bioinformatics 18, 1–18. https://doi.org/10.1186/s12859-017-1954-8.