Geophysical Research Letters

Supporting Information for

Plankton Imagery Data Inform Satellite-Based Estimates of Diatom Carbon

A. P. Chase[1], E. S. Boss[2], N. Haëntjens[2], E. Culhane[3], C. Roesler[4], and L. Karp-Boss[2]

[1]Applied Physics Laboratory, University of Washington, Seattle, WA, USA
[2]School of Marine Sciences, University of Maine, Orono, ME, USA
[3]Woods Hole Oceanographic Institution, Woods Hole, MA, USA
[4]Department of Earth and Oceanographic Science, Bowdoin College, Brunswick, ME, USA

Contents of this file

Introduction

This supporting information document provides expanding details and information regarding several areas of the main manuscript. These include: details of in situ data collection and the plankton classification network used to process and analyze the plankton imagery collected with the Imaging FlowCytobot; details of the CHEMTAX method and results as it was applied in this study; details of the neural network model for estimating diatom carbon and the associated model uncertainties; and information on the satellite data and remote sensing-based results and their comparison to in situ data. There are 10 figures and 2 tables contained within this document.

Text S1. In situ temperature, salinity, and pigments

Temperature and salinity were measured continuously by an SBE45 thermosalinograph connected to the ship's flow-through seawater system (intake at ~5 m depth). Concentrations of phytoplankton pigments were determined from surface water samples using HPLC (205 samples in total; Figure 1a). Water samples (1-2 L) were collected either from the flow-through seawater system, or from the 5 m Niskin bottle deployed on a

CTD rosette. Samples were filtered immediately onto 25 mm diameter GF/F filters (0.7 µm nominal pore size) and processed by the Ocean Ecology Laboratory at NASA Goddard Space Flight Center following Hooker et al. (2009) and Van Heukelem and Thomas (2001). Particulate absorption spectra ($a_p(\lambda)$) were measured continuously with a hyperspectral absorption and attenuation meter (ac-s, Sea-Bird Scientific) from the flow-through seawater system using automated inline filtration as a calibration-independent method (Boss et al., 2019; Slade et al., 2010). The $a_p(\lambda)$ spectra were binned to 1-minute temporal resolution and Chl *a* concentrations (mg m^{-3}) were estimated using the magnitude of a Gaussian function centered on the red peak at 675 nm ($a_{gaus}675$, units of m^{-1}) following Chase et al. (2013). The equations were tuned for the NAAMES dataset using HPLC pigment concentrations resulting in the following relationship: Chl a = $(a_{gaus}675/0.02)^{1.04}$.

Text S2. IFCB imagery and the plankton classification network

The IFCB automatically drew water samples (5 ml, nominally) approximately every 25 minutes from the flowing seawater system. Particles within the sample were channeled single-file and imaged with a camera if a laser-induced chlorophyll fluorescence threshold just upstream of the camera was reached. Imaging based on a chlorophyll fluorescence threshold in this way minimizes the number of non-phytoplankton particles recorded. From the four NAAMES cruises there were 4,328 IFCB samples that passed quality control standards and are used in the present study. Details on data preparation and quality flags can be found in the public code repository at: https://github.com/OceanOptics/ifcb-tools. All images used in this study are available at the EcoTaxa web platform (Picheral et al. 2017; https://ecotaxa.obs-vlfr.fr/; search "NAAMES").

A multi-step approach was used to classify the plankton and particle images collected with the IFCB from the four NAAMES cruises. A deep learning classification network was developed, which first requires a set of classified images for network training. To curate the images used in development of the deep learning network, several steps were taken: 1) a learning set was built by manually sorting 14,700 images into 58 categories, and 2) was applied within the EcoTaxa platform (https://ecotaxa.obs-vlfr.fr/explore/; search "NAAMES") using a random forest, machine learning approach to categorize all 4.8 million IFCB images from across the four NAAMES cruises. 3) Of these images, 2.2 million were manually confirmed or corrected in their categorization and labeled at two taxonomic resolutions: the highest level possible based on morphological features (usually genus level), and a broader categorization of 18 groups, where "diatoms" is one group of all diatom genera combined. The use of the 18 groups increases the number of cells per group, which subsequently increases the accuracy of the deep learning network, compared to training the network at the genus level where many groups do not have a sufficient number of cells for the training to be effective.

The 18 categories were used in a deep learning classification model that predicts the category of all classes with 86% accuracy, and diatoms as a group with 90% accuracy, across all 4.8 million images collected across the four NAAMES cruise. Details of the network testing and associated figures can be found at https://github.com/emmettFC/selected-projects/tree/master/plankton_vision, and code at https://github.com/ifcb-utopia/model-development-archive. In addition to traditional accuracy, precision, and recall metrics, we also considered model accuracy in terms of mislabeled data. Upon manual review of thousands of model predictions, we discovered that roughly a third (35%) of errors across all 18 groups are in fact original human error than had been corrected by the deep learning classification assignments. Thus, our model accuracy is likely higher than reported based on our analysis of mislabeled images corrected by the model. We incorporated the 90% accuracy value into our later analyses to estimate diatoms using satellite data. Given that manual review shows model prediction is likely substantially more accurate than reported, due to the model correcting incorrectly labelled data, our overall model error is likely overestimated.

Because larger phytoplankton are less abundant for a given volume compared to smaller plankton, we used the method of Chase et al. (2020) to combine IFCB samples and thus increase cell counts and reduce uncertainties. If a given IFCB sample contained fewer than 2,500 images, it was combined with subsequent samples until either 2,500 total images, or a 6-hour time span was reached (whichever came first). To avoid combining IFCB samples from different water masses, we then removed any combined samples that deviated from the originally defined sample by more than 20%, in terms of cell counts. The 20% was determined by assessing the distribution of all values of deviation in cell counts among all combined IFCB samples. We found that 20% was the threshold value for the first standard deviation above the mean (84^{th} percentile), and in this way we aim to account both for the natural variability in cell counts in situ, while also avoiding the combination of IFCB data from distinct water masses in terms of phytoplankton populations. This approach was found to also work well with regards to physical fronts, such that no strong temperature gradients were crossed when combined IFCB data were evaluated using the cell counts-based threshold (Chase et al., 2020). Combining IFCB samples reduced the number of samples from 4,328 to 1,755. Following image classification and sample combination, combined samples with fewer than 20 diatoms were removed, because their associated relative uncertainty in counts is high enough to render the quantification of diatom carbon not significantly different from zero. This resulted in the removal of 306 data points, for a final sample count of 1,499 used in the analysis of the present study. Diatom carbon concentrations are log-normally distributed (Fig. S1), and as a result the log-transformed carbon concentration values are used during neural network development (main text section 2.4).

Text S3. Application of CHEMTAX to estimate diatom carbon

S3.1 CHEMTAX Implementation

The CHEMTAX program predicts the relative contribution of different phytoplankton groups to total Chl *a* via analysis of the relative proportions of accessory pigments (Mackey et al., 1996). The program requires a set of phytoplankton groups and associated initial ratios of accessory pigments to Chl *a* to be defined. We used the initial pigment ratios and phytoplankton groups from two different published sources: Swan et al. (2016) and van de Poll et al. (2013) (hereafter denoted as S16 and vdP13, respectively). This results in the contribution of either seven or nine phytoplankton groups (depending on the initiation values used) to Chl *a* for each surface water HPLC data point during the four NAAMES cruises (n = 205).

The first set of initial ratios and groups we applied based on the reported results of S16, in which the authors performed a global clustering analysis of accessory pigment ratios and determined appropriate pigment ratios and phytoplankton groups for different ocean regions and during each of the four seasons. The geographic range of the North Atlantic Ocean is encompassed by two of the clusters defined by S16: cluster 5, which is denoted as "polar", and cluster 6, which is denoted as "global mesotrophic". For the inputs to our CHEMTAX runs, we used the mean values of pigment ratios for clusters 5 and 6 (noting that the taxonomic composition of the two cluster was the same with the addition of one more group in the global mesotrophic cluster (cyano-2). This group is included but not as a mean. As seasonal values are reported in S16, we use these means for the fall values of S16 for NAAMES cruises 1 (November) and 3 (September), and the spring values for NAAMES cruises 2 (May) and 4 (March).

The second set of pigment ratio values is from vdP13 and their associated supplementary information. Their study takes place in the Northeast Atlantic Ocean, and they applied two sets of initial pigment ratios: one for high-light and one for low-light acclimated phytoplankton. As the majority of our data is from underway ship measurements, we cannot easily diagnose the light history of phytoplankton in the NAAMES dataset. Therefore, we use an average of the high and low light values from vdP13, and the nine phytoplankton groups they report: prasinophytes, dinophytes, cryptophytes, haptophytes_1, haptophytes_2, pelagophytes, *Synechococcus*, *Prochlorococcus*, and diatoms. We use their initial ratio values, as they do not report final values.

For all CHEMTAX runs the settings were as follows:

Maximum number of iterations = 500;
Stop calculation when residual is below = 0.005;*
Initial step size (inverse) = 10;
Increase step size by a factor of = 1.3;
Halt calculation when step is larger than = 1000;

Verbosity = 3;
Vary how many ratios at each step = 5;
Interval between redetermination of what ratios to vary = 5;

* we also tested a value of 0.0001 and the results were unchanged

S3.2 CHEMTAX Results

Results of the two initializations are noticeably different in multiple phytoplankton groups, but particularly diatoms (Figs. S2). Diatoms differ most noticeably in their contribution to Chl *a* during the first part of the May campaign (sample numbers 95-170). Interestingly, although in general the results based on the vdP13 CHEMTAX algorithm more closely match the imagery-derived diatom contribution (Figs. S3, S4), they failed to 'detect' diatoms during an extended station occupation in September (sample numbers 187-200), while both imagery and the S16 CHEMTAX algorithm show approximately 3-10% diatom contribution to Chl *a*. Overall, the vdP13 CHEMTAX implementation shows reduced overestimation of diatoms relative to imagery, compared to S16 CHEMTAX algorithm (Fig. S3).

Text S4. Neural network for estimating diatom carbon

The shallow network was trained using the MATLAB *nftool* function using a process of Bayesian regularization, where updates to weight and bias values are made according to Levenberg-Marquardt optimization (MATLAB Deep Learning ToolboxTM Reference, 2020). The network is a two-layer feedforward network, with a hidden layer that uses a sigmoid transfer function, and an output layer that uses a linear transfer function. The inputs and target were randomly separated into independent training (n = 536) and testing (n = 95) datasets. Network training continued until a minimum performance gradient was reached, i.e., validation errors no longer decreased. This occurred with a run time of several seconds and therefore thresholds for training time were not a limiting factor. Through iterative testing, we found that a network with 15 neurons in the hidden layer performed the best in terms of median error between the modeled and target (i.e., imagery-based) diatom carbon. Following the network training with in-situ input and target data, a function for the network was saved and run using satellite data for the input values of Chl *a,* temperature, and salinity.

In situ concentration of Chl *a* determined via Gaussian decomposition of $a_p(\lambda)$ has a relative root-mean-square error (RMSE) value of 29% when compared to HPLC Chl *a* concentrations (not shown). In our network model, input parameter uncertainties are assumed to be negligible for situ-measured surface seawater temperature and salinity. Uncertainty in diatom carbon from imagery is determined by combining the error in estimating a cell biovolume from its image, which is reported in Moberg and Sosik (2012) as 17% (the same uncertainty applies to updated versions of the biovolume

method available at https://github.com/hsosik/ifcb-analysis, H. Sosik *pers. comm.*), with the uncertainty resulting from statistical counting errors, which has a mean value of 18%. Combined with these uncertainties is the 10% error associated with the accuracy of identifying all diatoms using the image classification network (Text S2). All errors are relative (percentage) values, allowing for propagation of error following:

$$\mathbf{Unc_{data}} = \sqrt{\mathbf{0.17^2 + 0.18^2 + 0.1^2 + 0.29^2}} = \mathbf{0.39},$$

This results in estimated uncertainty in the data (Unc_{data}) of 39%, assuming the different uncertainties are uncorrelated. The neural network used to estimate diatom carbon also has uncertainty associated with the accuracy of the model itself. Following the model training, errors are calculated by the difference between the model target ($C_{diat\_Imagery}$) and the model output. The median value of the error in estimated diatom carbon is 1.05 mg m^{-3} and the relative error is 52% (Fig. S6). This relative error is combined with the $\boldsymbol{Unc_{data}}$ (39%) shown above:

$$\mathbf{Unc_{NN}} = \sqrt{\mathbf{0.39^2 + 0.52^2}} = \mathbf{0.65},$$

resulting in an uncertainty value of the neural network model (Unc_{NN}) of 65%. Note that at low estimated diatom carbon values, the absolute error dominates over the relative error, and thus Unc_{NN} = max(1.05 mg m^{-3}, 65%).

For model results using satellite data input (Text S5), additional uncertainty were imposed by both the accuracy to which the satellite data estimates in situ parameters, as well as the environmental variability within a given pixel. The extent to which the value of a satellite pixel parameter (e.g., a given Chl *a* concentration or salinity value) represents the waters within the pixel size (ranging from 0.01 to 0.25 degrees square for the data used in the present study) was varied based on heterogeneity of and distribution of water masses.

Text S5. Satellite data used for neural network model input

Level 3 Chl *a* data, calculated using band ratio algorithms from multispectral reflectance data (Hu et al., 2012), were downloaded from the NASA OC.DAAC for NASA's MODIS Aqua instrument (https://oceancolor.gsfc.nasa.gov/l3/). The daily Chl *a* values for May 12, 2016 were used to compare to in situ data from May 11-13, 2016 (Scene 1), and the 8-day Chl *a* composite from May 16-23, 2016 was used to compare to in situ data from the matching dates of May 16-23, 2016 (Scene 2). Multi-scale Ultra-high Resolution sea surface temperature (SST) data were downloaded from the PO.DAAC (https://podaac.jpl.nasa.gov/), for May 12, 2016 (Scene 1) and May 20, 2016 (Scene 2). Third, monthly Remote Sensing Systems Soil Moisture Active Passive sea surface salinity (SSS) were downloaded for May 2016 (Figures S6-S7). The satellite datasets have resolutions of 4.6 km, 1.1 km, and 27.75 km for Chl *a,* SST, and SSS, respectively. All satellite data were linearly interpolated (without extrapolation) onto the 0.01 degree

spatial resolution of the SST data. Uncertainty in satellite values was calculated using the mean of a 3 x 3 grid of satellite data points surrounding the data point nearest to the in situ data**.** To compare in situ point measurements to the satellite pixel spatial resolution, the standard deviation of the values for a 20-minute duration of flow-through sample collection were used to account for environmental variability in situ data. The mean and median distances covered by the ship during the 20-minute duration of combined flow-through data were 9.67 km and 4.98 km, respectively.

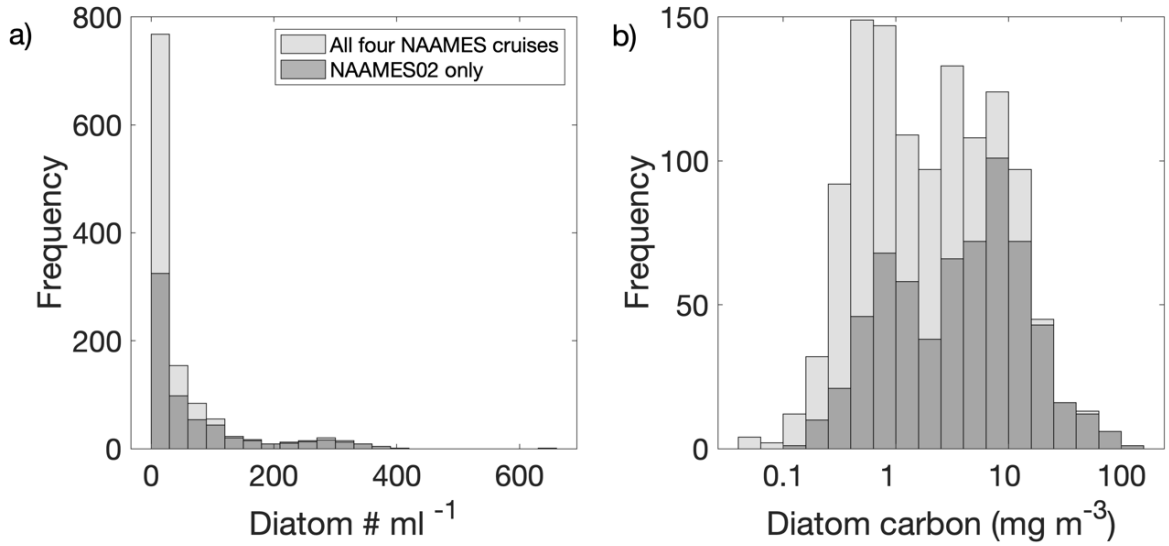

Figure S1. a) Number of diatoms per ml for samples from all four NAAMES cruises (light gray) and for NAAMES02 only (dark gray). **b)** Frequency distribution of diatom carbon per sample from all four NAAMES cruises (light gray) and for NAAMES02 only (dark gray).

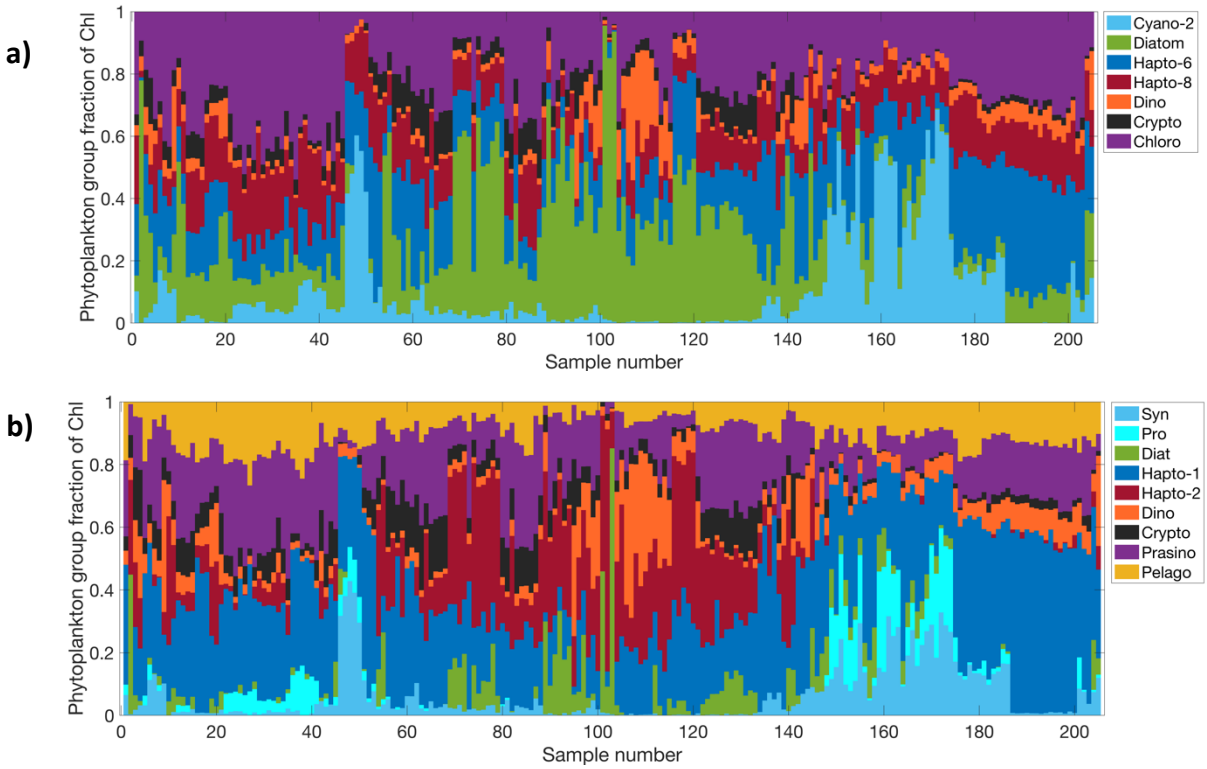

Figure S2. a) Phytoplankton group contribution to Chl *a* from CHEMTAX analysis with inputs from Swan et al. 2016. **b)** Phytoplankton group contribution to Chl *a* from CHEMTAX analysis with Inputs from van de Poll et al. 2013.

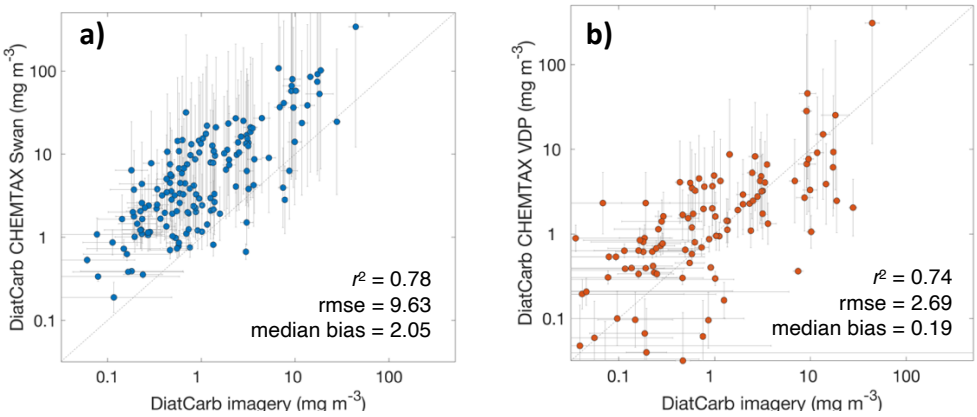

Figure S3. a) Diatom carbon from IFCB imagery vs. diatom carbon from CHEMTAX. Left: results calculated using initial ratios and phytoplankton groups from Swan et al. (2016). Error bars in the x direction are from statistical counting errors in imagery data, and error bars in the y direction are calculated from the uncertainty in converting chlorophyll to carbon. **b)** same as in the left panel, but with CHEMTAX inputs from van de Poll. (2013).

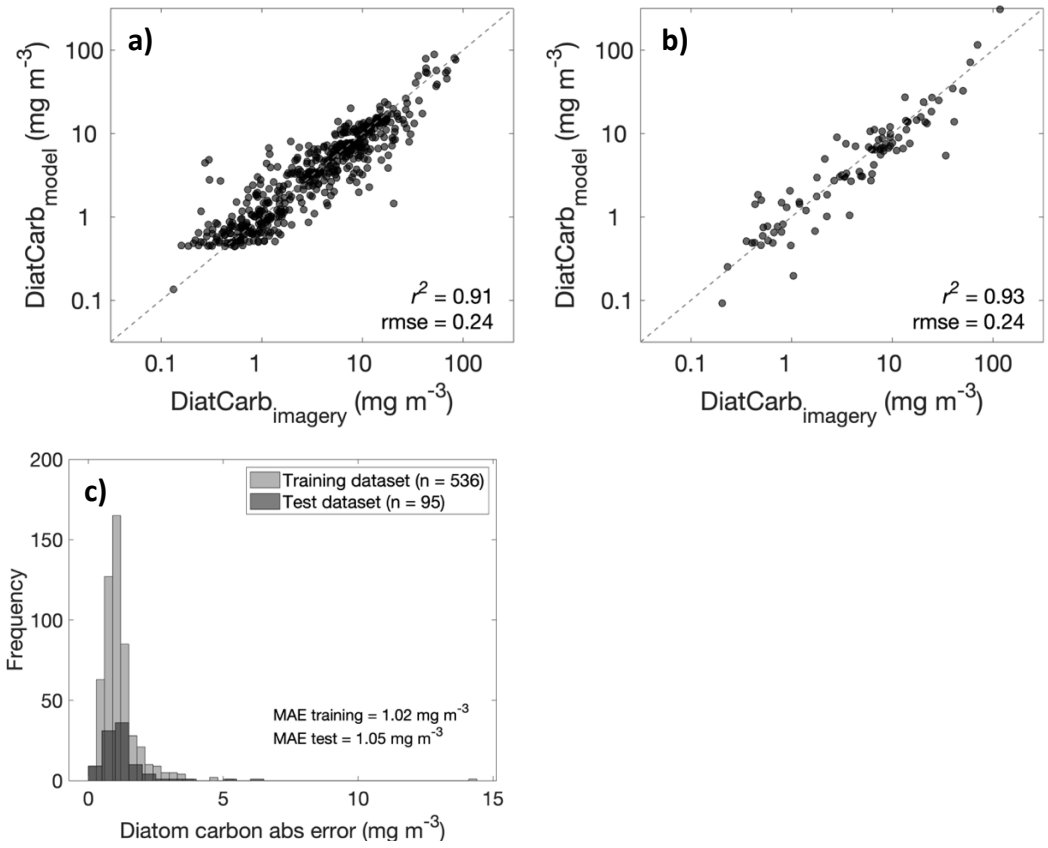

Figure S4. a) Network training results, n = 536. **b)** Network testing results, n = 95; data points are random and independent of the data used for training. r^2= Pearson correlation coefficient; RMSE = root mean squared error. **c)** Distributions of error in diatom carbon estimated from the neural network model. MAE = mean absolute error.

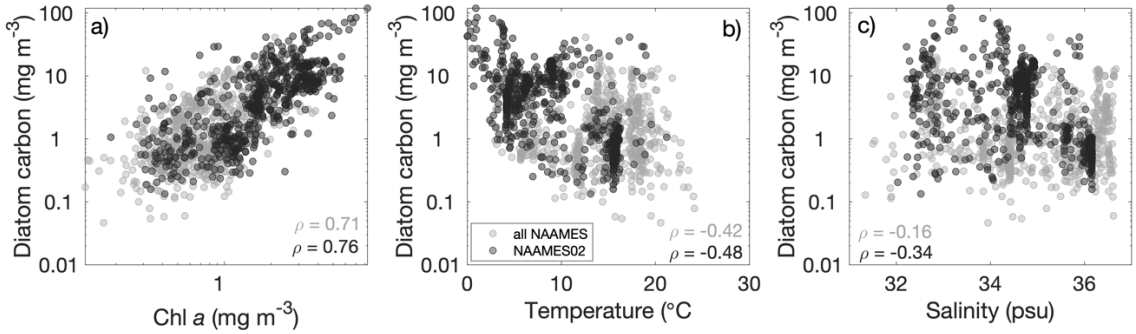

Figure S5. Diatom carbon from IFCB imagery as a function of **a)** Chl *a*, **b)** temperature, and **c)** salinity. 13 data points with values below 31 psu (values ranging from 28.2-30.9) are omitted from panel c) for ease of viewing. Data from all four NAAMES cruises are shown in gray, and from NAAMES02 alone shown in black. Spearman's rank correlation (ρ) values are shown on each panel and all correlations are significant with p-values < 0.001.

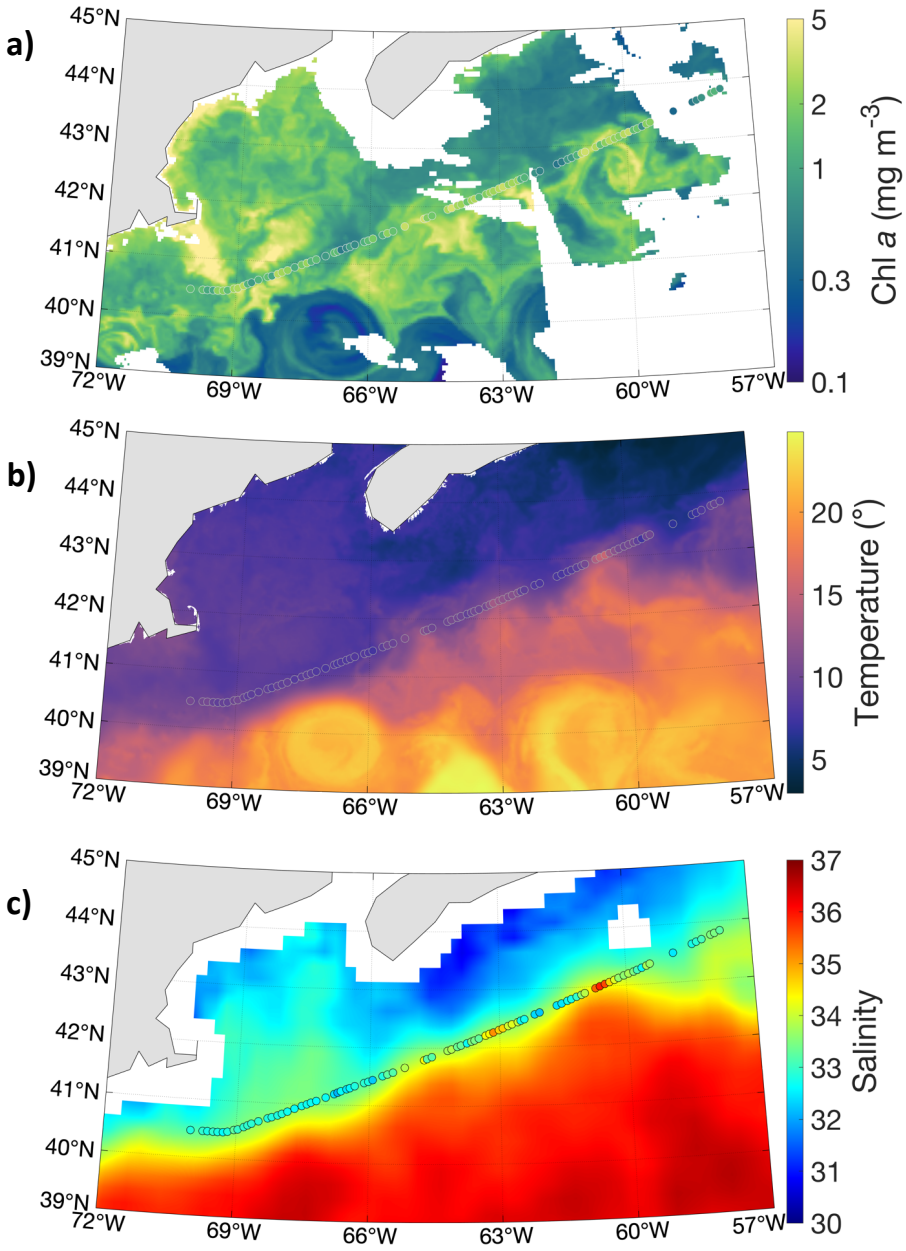

Figure S6. Satellite data used for neural network model input in the region of scene 1, with in situ data from May 11-13 shown in colored dots. **a)** MODIS Aqua L3 Chl *a* concentration daily product for May 12, 2016. **b)** SST from the MUR merged, multi-sensor L4 product for May 11, 2016. **c)** SMAP ocean surface salinity, monthly composite for May 2016. Data were linearly interpolated without extrapolation onto the 0.01 degree resolution SST latitude/longitude grid.

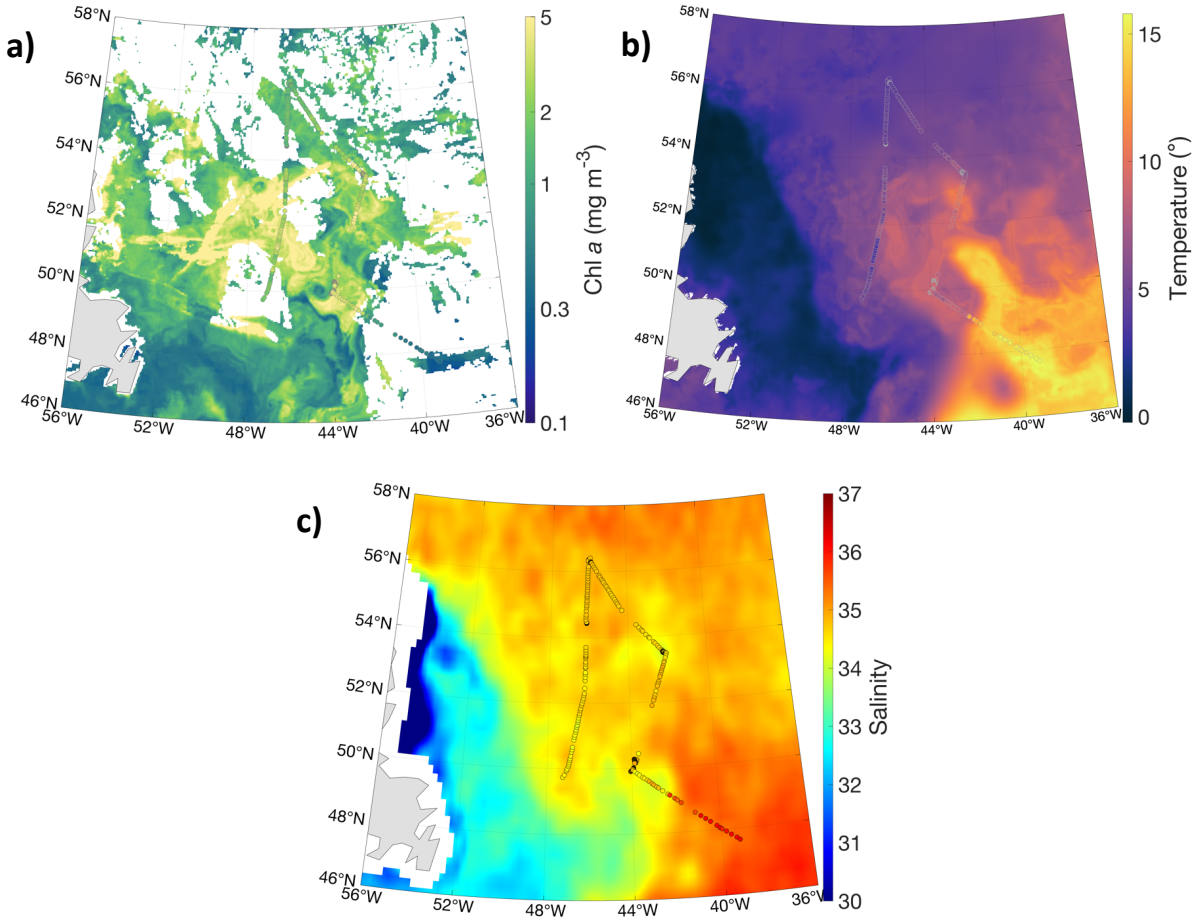

Figure S7. Satellite data used for neural network model input in the region of scene 2, with in situ data from May 16-23 shown in colored dots. **a)** MODIS Aqua L3 Chl *a* concentration weekly composite for May 16-23, 2016. **b)** SST from the MUR merged, multi-sensor L4 product for May 20, 2016. **c)** SMAP ocean surface salinity, monthly composite for May 2016. Data were linearly interpolated without extrapolation onto the 0.01 degree resolution SST latitude/longitude grid.

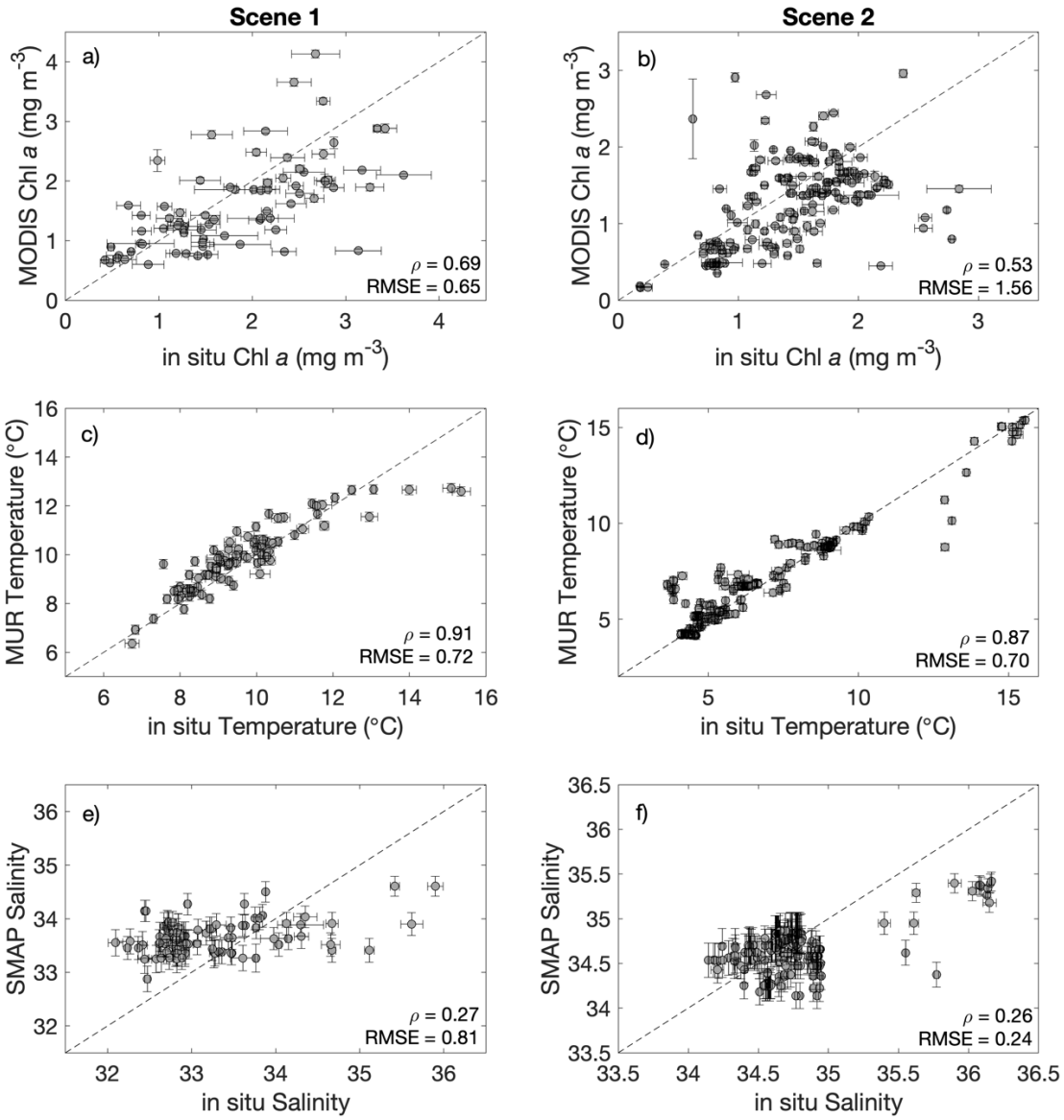

Figure S8. In situ vs. satellite-based estimates of Chl *a* (**a-b**), temperature (**c-d**), and salinity (**e-f**) for scene 1 (**a,c,e**) and scene 2 (**b,d,f**). Error bars in the x-axis show the standard deviation of the values for a 20 minute duration of flow-through sample collection, to account for environmental variability of a location. Error bars on the y-axis show the mean of a 3 x 3 grid of satellite data pixels surrounding the pixel nearest to the in situ data.

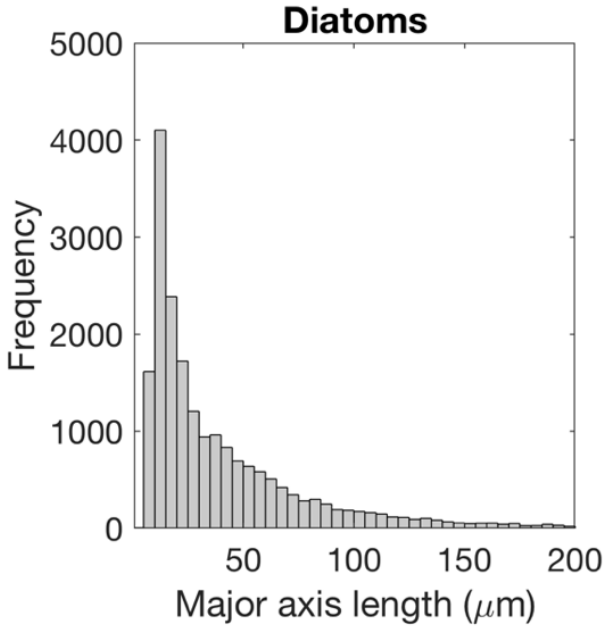

Figure S9. Major axis length of all diatom cells or cell chains imaged across all four NAAMES cruises.

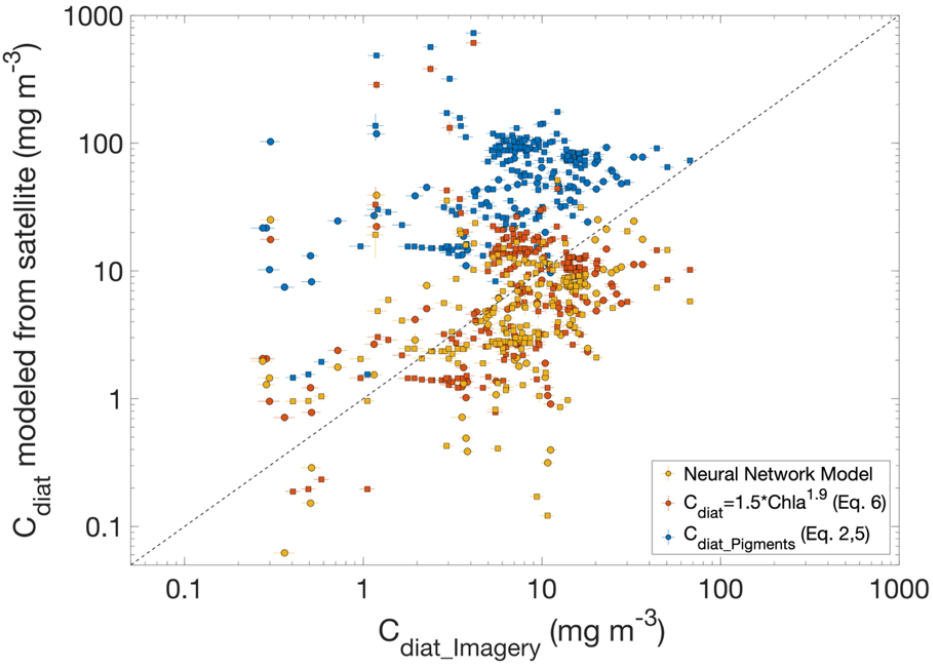

Figure S10. Diatom carbon estimated from in situ IFCB data (x-axis) vs. from satellite-based methods (y-axis). Square data points represent data from scene 1 (in situ data from May 11 to 13) and circle data points represent scene 2 (in situ data from May 16 to 23).

Table S1. Equations used to calculate the fraction of diatoms as a function of Chl *a*. Adapted from Table S2 in Losa et al. (2017).

Model	a_0	a_1	a_2	a_3	Reference
$(\boldsymbol{a_0} + \mathbf{exp}\,(\boldsymbol{a_1}\,\mathbf{log_{10}}\,\boldsymbol{x} + \boldsymbol{a_2})^{-\mathbf{1}}$	1.3272	-3.9828	0.1953		Hirata et al. (2011)
$(\boldsymbol{a_0} + \mathbf{exp}\,(\boldsymbol{a_1}\,\mathbf{log_{10}}\,\boldsymbol{x} + \boldsymbol{a_2})^{-\mathbf{1}}$	0.6544	-1.1796	0.8534		Losa et al. (2017) A
$\boldsymbol{a_0} + \boldsymbol{a_1}\mathbf{sin}(\boldsymbol{a_2}(\mathbf{log_{10}}\,\boldsymbol{x} + \boldsymbol{a_3}))$	0.4629	0.3921	1.2214	-0.0141	Soppa et al. (2014)
$\boldsymbol{a_0} + \boldsymbol{a_1}\mathbf{sin}(\boldsymbol{a_2}(\mathbf{log_{10}}\,\boldsymbol{x} + \boldsymbol{a_3}))$	0.6097	0.4948	0.8129	-0.7179	Losa et al. (2017) B

where $\boldsymbol{x}$ = $\log_{10}$(Chl *a*) and [Chl *a*]=mg m^{-3}.

Table S2. Statistics of the comparison between diatom carbon estimated from imagery data and from satellite-based methods (visualized in Figure S10).

	Spearman's rho (ρ)	root-mean squared error (mg m^{-3} carbon)
$C_{diat\_Pigments}$ (Eq. 2, 5)	0.32	62.3
C_{diat} = 1.5*$Chla^{1.9}$ (Eq. 6)	0.32	26.1
Neural Network model	0.42	7.4