https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

Phytoplankton Community Composition in the Eastern Subarctic Pacific Derived from Hyperspectral Optics

Sacchidanandan V. Pillai[1], M. Angelica Peña[2], Brandon J. McNabb[1], William J. Burt[3], and Philippe D. Tortell[1,4]

[1]Department of Earth, Oceans, and Atmospheric Science, University of British Columbia, Vancouver, British Columbia, Canada
[2]Institute of Ocean Sciences, Fisheries and Ocean Canada, Sidney, British Columbia, Canada
[3]Planetary Technologies, Halifax, NS, Canada
[4]Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada

Correspondence: Sacchidanandan Pillai (spillai@eoas.ubc.ca)

Abstract. We evaluate the utility of hyperspectral particulate absorption data to characterize phytoplankton community structure in the eastern Subarctic Pacific Ocean. Relative to existing algorithms based solely on Chlorophyll-a concentrations (Chla), improved taxonomic classification (validated with pigment-based data) was obtained by including Principal Components Analysis of hyperspectral absorption data. Multiple linear regression of hyperspectral absorption data yielded better taxonomic classification, particularly for estimates of haptophyte biomass. In addition, size-fractionated hyperspectral measurements were used to determine the dominant phytoplankton size of the phytoplankton community. Using high-frequency ship-board optical data, we examined the spatial patterns in phytoplankton taxonomic abundance in coastal and offshore waters around Vancouver Island, British Columbia. Results from this analysis were consistent with expectations based on previous low-resolution sampling, demonstrating expected seasonal succession of different phytoplankton groups, and significant variability in coastal phytoplankton taxonomy associated with dominant hydrographic features. In contrast, much less spatial and temporal variability was observed in offshore waters. Derived patterns in phytoplankton taxonomy were linked to observed patterns in surface water biogeochemical properties, notably the distribution of dimethyl sulfide (DMS) and dimethylsulfoniopropionate (DMSP) to Chla ratios. Our results highlight the potential for shipboard hyperspectral absorption data to describe phytoplankton community composition and ancillary biogeochemical variables.

1 Introduction

Phytoplankton taxonomy plays an important role in ocean biogeochemistry and carbon cycling (Falkowski et al., 1998), as different phytoplankton groups (or functional types) have distinct physiological and ecological traits that drive the emergent properties of marine ecosystems (Queré et al., 2005). Diatoms, for example, contribute disproportionately to carbon export production, by virtue of their large cell size (for many species) and dense siliceous frustules, which resulting in elevated sinking rates (Smetacek, 1985) and make them sensitive to silicic acid limitation (Krause et al., 2019). In contrast, some haptophytes produce calcium carbonate shells (Yoch, 2002) and are significant producers of the climate-active gas dimethyl sulfide (DMS)

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

(Archer et al., 2000), making them important contributors to ocean alkalinity balance and regional climate regulation. The results of anthropogenic forcing, such as ocean warming, increased stratification and changes to nutrient supply, are likely to cause shifts in the phytoplankton taxonomy (Henson et al., 2021). Knowledge of the distribution of the different phytoplankton functional types is thus important to observe ongoing changes in phytoplankton ecology and for the parameterisation and evaluation of ocean biogeochemical models (Anderson, 2005).

Phytoplankton taxonomic distributions in marine waters have typically been estimated from discrete shipboard measurements, which are labour-intensive and provide limited spatial and temporal resolution. Traditional methods include the microscopic examination of water samples and the analysis of diagnostic photosynthetic pigments by High-Performance Liquid Chromatography (HPLC). Microscopy requires expert knowledge and may require a scanning electron microscope for smaller phytoplankton. Pigment measurements can be used to estimate the contribution of different taxonomic groups to total chlorophyll concentration, but these are subject to some uncertainty due to overlapping pigment signatures of different groups, and variable pigment contents within groups (Roy et al., 2011). Nonetheless, the pigment-based classification remains the most commonly used and accessible method, with the potential to broadly distinguish different phytoplankton groups (Kramer and Siegel, 2019).

At present, there are several approaches for estimation of phytoplankton taxonomic composition . One approach focuses on characterizing the dominant group in a sample (e.g. diatoms vs. haptophytes), where the bulk properties of the assemblage are attributed to that group. We refer to this qualitative approach as phytoplankton community classification and note that it has been used in a variety of studies (Ciotti et al., 2002; Cetinic et al., 2015; Kramer et al., 2020). Another approach, referred to here as the assessment of phytoplankton composition, is to estimate the contributions of the different phytoplankton groups to total biomass (derived from chlorophyll concentration), using methods such as CHEMTAX (Mackey et al., 1996) or Diagnostic Pigment Analysis (DPA) (Uitz et al., 2006). This approach provides quantitative estimates of phytoplankton composition based on the relative abundance of different groups. While similar, qualitative phytoplankton community classification and quantitative estimates of phytoplankton composition, have different strengths, limitations and applications.

In recent years, optical measurements have become an increasingly common tool for monitoring changes in phytoplankton biomass and composition at high resolution across a range of spatial scales. For example, hyperspectral spectrophotometers have been used to measure chlorophyll concentrations from the continuous underway seawater supply of ships, exploiting the strong contribution of chlorophyll-a to the particulate absorption at 676 nm (Boss et al., 2013). Hyperspectral data have also been used to estimate photosynthetic pigment concentrations using a variety of inversion methods (Liu et al., 2019), and applied to validate satellite Chlorophyll-a data products (Brewin et al., 2016). Studies have also demonstrated the link between multi- and hyperspectral measurements and the phytoplankton composition (Alvain et al., 2005; Mao et al., 2010; Isada et al., 2015). These studies are particularly important in light of the upcoming launch of the PACE (Plankton, Aerosol, Cloud and

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

ocean Ecosystem) satellite, which has a spectral resolution of 3nm across the visible light spectrum (Werdell et al., 2019). Robust interpretation of the data from this new satellite will depend on high-quality validation with in situ observations.

The eastern Subarctic Pacific provides an ideal study area to investigate gradients in phytoplankton taxonomic composition across space and time. This region is characterized by a range of contrasting environments, such as high-productivity coastal regions, stratified nitrate-limited waters, and iron-limited high-nutrient-low chlorophyll waters (Whitney et al., 2005). Spatial, seasonal and inter-annual variability of nutrient concentrations and phytoplankton biomass in the subarctic NE Pacific is also linked to various physical processes, including coastal upwelling, wind-driven mixing and river inputs. (Peña and Varela, 2007; Peña et al., 2019a). Understanding ecosystem responses to this physical variability requires estimates of phytoplankton biomass and community composition with high spatial and temporal resolution.

Previously, Zeng et al. (2018) compared different methods for estimating phytoplankton composition and size classes in the eastern subarctic Pacific. Their results showed that the Chlorophyll-a (Chla) based algorithm of Hirata et al. (2011) was the most accurate relative to DPA-based estimates of relative contributions of different phytoplankton groups to total biomass. A potential limitation of this approach is the fact that Chla concentrations can also record changes in the physiological status of phytoplankton associated with environmental factors such as light and/or nutrient availability (Geider et al., 1997; Behrenfeld et al., 2016). Indeed, Burt et al. (2018) found that the C:Chl ratio varied significantly with light and nutrient conditions in the eastern subarctic Pacific, suggesting that Chla concentrations are impacted by both physiological and biomass / taxonomic changes in phytoplankton composition. Additional information from hyperspectral data may thus provide further constraints on phytoplankton community composition.

This study aims to estimate phytoplankton communities and composition using hyperspectral absorption data collected in the NE Pacific, bench-marked against phytoplankton pigments measurements. We demonstrate the applicability of hyperspectral data to derive taxonomic information about phytoplankton assemblages and examine oceanographic controls on the optically-derived phytoplankton community composition. Using our hyperspectral data, alongside other high-frequency oceanographic measurements, we also show how phytoplankton taxonomy can help explain the distribution of other biogeochemical variables in our survey region.

2 Methods

2.1 Field measurements

The optical and pigment data used in this study were derived from ship-based observations collected on 9 oceanographic cruises on the CCGS *John P. Tully* in coastal and offshore waters of the northeast subarctic Pacific between 2016 - 2018, and in August, 2022 (see Table 1 for specific cruises). These cruises are part of the Line P and La Perouse oceanographic monitoring programs led by Fisheries and Oceans Canada (DFO), with an overarching objective to observe long-term trends in ocean properties in

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

Table 1. Research cruises on the CCGS *John P. Tully* used for data collection. [a] pigment data are currently unavailable and thus only optical data from this cruise are utilized.

Program	Dates	Cruise ID
Line P	Jun 05 - 21, 2016	2016-06
	Feb 05 - 23, 2017	2017-01
	Jun 04 - 20, 2017	2017-06
	Aug 15 - 31, 2017	2017-08
	Sep 11 - 28, 2018	2018-40
La Perouse	May 24 - Jun 5, 2016	2016-47
	May 23 - Jun 4, 2017	2017-05
	Aug 31- Sep 12, 2017	2017-09
	Aug 25 - Sep 6, 2022	2022-22[a]

coastal and offshore waters of the northeast subarctic Pacific. The location of the sampling stations and cruise tracks are shown in Fig. 1. Some of the data used in this paper have been previously published in other contexts (e.g., Burt et al. (2018); Zeng et al. (2018); Peña et al. (2019b)). Pigment measurements are currently unavailable for the August 2022 cruise, and thus only the optical data from this cruise are presented here.

Figure 2 shows the overall methodological approach used to predict phytoplankton communities and estimate phytoplankton composition from the ship-board measurements. Individual steps in this method are described in detail in the following subsections.

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

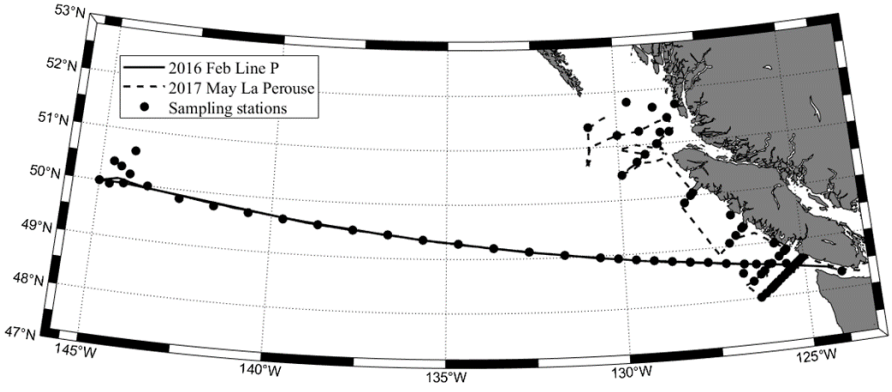

Figure 1. Location of HPLC sampling stations (black dots) and ship tracks during the 2016-01 Line P (solid line) and 2017-05 La Perouse (dotted line) cruises. Note that the actual cruise tracks change slightly by a few degrees between cruises.

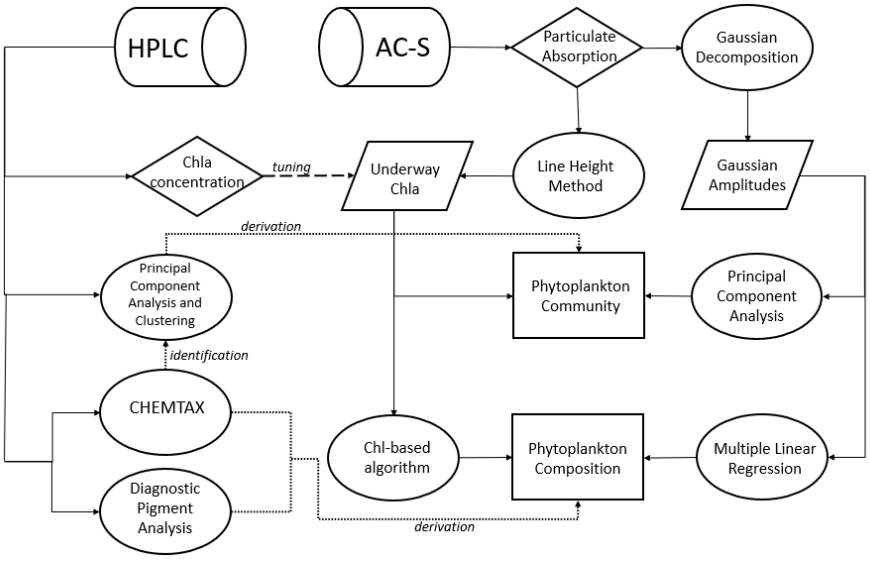

Figure 2. Flow-chart outlining the methods used to estimate phytoplankton community and composition based on HPLC-derived pigment concentrations and ship-board hyperspectral absorption data (WETLABS AC-S).

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

2.2 Pigment measurement

Phytoplankton pigment samples were collected from the upper 5 m at various stations (Fig. 1) using Niskin bottles attached to a CTD rosette package. Duplicate water samples ranging from 0.6 to 2 L (depending on phytoplankton biomass) were filtered onto Whatman GF/F filters and immediately stored at $-80°$C for subsequent determination of Chla and accessory pigment concentrations by HPLC analysis, following the method of Zapata et al. (2000). In total, data from 192 discrete HPLC samples were matched to the optical data using a time stamp alignment of $\pm$ 10 min for comparison of pigment and optically-derived phytoplankton composition estimates.

2.3 CHEMTAX and DPA analysis

To estimate the contribution of the main phytoplankton taxonomic groups to Chla, phytoplankton pigment data were analysed using CHEMTAX (v1.95; Mackey et al. (1996)), following the procedures detailed in Peña et al. (2019b). The data was separated into coastal (within the continental shelf, less than 200 m water depth) and offshore (beyond the continental shelf with water depth > 200 m) stations to account for differences in phytoplankton community composition between these regions (Peña et al., 2019b; Del Bel Belluz et al., 2021). The phytoplankton groups and initial pigment ratios used for the coastal and offshore stations are given in Table A1 and A2, respectively.

As an alternative to CHEMTAX, the Chla concentration attributable to the different phytoplankton groups was also calculated using diagnostic pigment analysis (DPA), which assigns certain pigments to different phytoplankton groups. This method was originally developed to determine the contribution of different phytoplankton size classes to Chla (Uitz et al., 2006), but has been subsequently used to estimate the contributions of the different phytoplankton taxonomic groups (Hirata et al., 2011). We use the methods and equations of Zeng et al. (2018), to allow a better comparison with their results from the NE Subarctic Pacific. We note that there was generally good agreement between DPA and CHEMTAX results in our data set, with a strong correlation between the estimates of diatoms ($R^2 = 0.81$) and haptophytes ($R^2 = 0.55$) produced by these two methods (Fig. B1), albeit with some offsets from a 1:1 line. Notably, CHEMTAX predicted somewhat lower diatom and higher haptophyte abundances than DPA in our data set, but the overall results are broadly consistent between these two methods. This difference between the two classification methods is likely due, in part, to the fact that DPA assigns each pigment to only one phytoplankton group, whereas CHEMTAX permits pigments to be shared by multiple groups.

Both DPA and CHEMTAX provide estimates of phytoplankton composition relative to the total phytoplankton biomass (estimated from total Chla concentration). Besides the errors associated with the chemotaxonomic derivation of phytoplankton composition, variability in the C:Chl ratio (Behrenfeld et al., 2016), introduces additional uncertainty in the relationship between Chla concentrations and carbon-based phytoplankton biomass. Indeed, Chase et al. (2022) found that the carbon biomass of diatoms predicted from pigment-based approaches differed significantly from that obtained from imaging flow cytometry,

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

suggesting a complex relationship between pigments and carbon biomass of phytoplankton groups. Nonetheless, our focus here is on comparing the performance of different Chla-based methods, which remain the most commonly used approach.

2.4 Pigment clustering

To identify the different phytoplankton communities found in our study region, principal component analysis (PCA) was applied to the HPLC-derived pigment concentrations. This approach clusters data into representative groups with co-varying pigment signatures (Borcard et al., 2018). Following the methods of Kramer and Siegel (2019), we excluded the concentration of degradation pigments from the analysis (chlorophyllide-a, pheophorbide-a, pheophythin-a, methyl chlorophyllide-a) and redundant pigments (Monovinyl Chlorophyll-a). The resulting 17 pigments used in the analysis are chlorophyll c1, c2, c3, peridinin, 19′- butanoyloxyfucoxanthin, fucoxanthin, neoxanthin, prasinoxanthin, violaxanthin, 19′- hexanoyloxyfucoxanthin, diadinoxanthin, alloxanthin, diatoxanthin, zeaxanthin, lutein, Chlorophyll-b, divinyl chlorophyll-a.

The pigment concentrations were normalized by the Chla concentration prior to performing PCA. Clusters were then classified using gaussian mixture model clustering. The number of clusters was chosen based on a visual inspection of the data as well as hierarchical clustering of pigment data (Kramer and Siegel, 2019). The composition of the different clusters was determined using the CHEMTAX results (see below), and the differences between the clusters were tested using ANOVA. A full description of the clustering procedure is described in the Appendix C.

2.5 Nutrient, Salinity and Sea Surface Temperature

Nutrient concentrations, salinity and sea surface temperature for each of the cruises were accessed from the Institute of Ocean Sciences Data Archives (DFO, 2023). Surface concentrations of nitrate, phosphate and silicate were determined using an Astoria Autoanalyser with the methods described in Barwell-Clarke and Whitney (1996). Sea Surface Temperature (SST) and Salinity were determined using a SEABIRD SBE-45 CTD connected to the ships underwater supply.

2.6 DMS and DMSP concentrations

During the 2022 August La Perouse Cruise (refer to Table 1), dimethyl sulfide (DMS) and dimethylsulphoniopropionate (DMSP) concentrations were measured approximately every 40 minutes throughout the cruise using the Organic Sulphur Sequential Chemical Analysis Robot (OSSCAR) using methods described in Asher et al. (2015). OSSCAR uses a purge and trap gas chromatograph, coupled with a capillary inlet mass spectrometer to measure DMS and DMSP (hydrolyzed to DMS). We examined the influence of phytoplankton taxonomy on DMS / DMSP concentrations using ANOVA across different phytoplankton clusters.

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

2.7 Continuous ship-board optical measurements

The wavelength (λ)-dependent absorption of light in seawater, $a(\lambda)$, can be separated into several individual constituents,

$$a(\lambda) = a_w(\lambda) + a_{CDOM}(\lambda) + a_{NAP}(\lambda) + a_{ph}(\lambda) \quad (1)$$

where $a_w(\lambda)$ is absorption by pure water, $a_{CDOM}(\lambda)$ is absorption by CDOM (Coloured Dissolved Organic Matter), $a_{NAP}(\lambda)$ is absorption by Non-Algal Particles (NAP) and $a_{ph}(\lambda)$ is absorption by phytoplankton. The particulate absorption, defined as absorption by particles larger than 0.2μm, $a_p(\lambda)$ consists of the combined absorption of NAP and phytoplankton.

$$a_p(\lambda) = a_{NAP}(\lambda) + a_{ph}(\lambda), \quad (2)$$

Hyperspectral particulate absorption, $a_p(\lambda)$, was continuously measured from the ship's seawater intake (at a depth of $\sim$ 7 m) using a Wetlabs AC-S spectrophotometer, following the method described in Burt et al. (2018). Seawater was passed through a Vortex debubbler before flowing through the measurement tubes of the AC-S. To measure the apparent optical properties of dissolved components, an upstream 3-way valve diverted water through a 0.2μm cartridge for 10 min out of every hour. Particulate absorption spectra were calculated by subtracting the filtered seawater absorption signal (representing $a_w(\lambda) + a_{CDOM}(\lambda)$), and corrected for both scattering and temperature changes between the hourly measurements, using methods described in Slade et al. (2010). Similarly, the hyperspectral particulate beam attenuation, $c_p(\lambda)$ was calculated from the AC-S.

2.8 Chla concentrations from the Absorption Line Height method

Following Boss et al. (2007), particulate absorption at 676 nm ($a_{ph}(676)$) was used to estimate Chla concentration after calibration against HPLC measured Chla (Chlorophyll-a + Divinyl Chlorophyll-a + Chlorophyllide-a). A linear equation

$$a_{ph}(676) = a_p(676) - \frac{39}{65}a_p(650) - \frac{26}{65}a_p(715), \quad (3)$$

was used to subtract a baseline absorption (between 650 and 715 nm). In our data set, $a_{ph}(676)$ showed a strong linear correlation with Chla (R^2 of 0.96; *n*=192),

$$\text{Chla} = 76.70 \pm 2.7 a_{ph}(676) \quad (4)$$

with a slope that was not significantly different (at the 95% confidence level) to that reported by Burt et al. (2018) in previous surveys of our sampling region.

2.9 Phytoplankton size index

To evaluate changes in phytoplankton size from the optical measurements, we compared the spectral slope of the particulate beam attenuation (γ_{cp}) throughout the cruises. This derived variable has been shown to correspond well with the mean size of

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

suspended particles (Boss et al., 2001). Although γ_{cp} may not be specific to phytoplankton (as it is inclusive of all types of particles), it has been shown to good proxy as a phytoplankton size index in the absence of significant terrestrial contributions (Haëntjens et al., 2022). Following (Boss et al., 2001) , γ_{cp} was calculated as:

$$c_p(\lambda) = A\lambda^{-\gamma_{cp}}. \tag{5}$$

To provide additional context for γ_{cp} values, we measured size fractionated values by diverting seawater for at least 10 minutes through a Purtrex $20\mu m$ filter when the ship was at the station during the August 2022 cruise. The nominal retention rating of the filter was chosen to separate micro-phytoplankton ($> 20\mu m$ in size) from nano-phytoplankton and pico-plankton ($< 20\mu m$). To obtain the spectra for particles greater than $20\mu m$, the attenuation spectra measured for particles smaller than $20\mu m$ were subtracted from that of the total particulate seawater.

2.10 Gaussian Decomposition

Since the majority of phytoplankton absorption, $a_{ph}(\lambda)$, is attributable to phytoplankton pigments, this variable can be described as a function of pigment concentrations as,

$$a_{ph}(\lambda) = Q_a^*(\lambda) \sum_{i=1}^{n} a_i^*(\lambda) C_i \tag{6}$$

where *n* is the total number of pigments considered, a_i^* is the specific absorption coefficient of the *i*th pigment at wavelength λ, and C_i is the concentration of the *i*th pigment. $Q_a^*(\lambda)$ is the pigment package effect index which describes actual absorption compared to if the pigments were dispersed into solution (Morel and Bricaud, 1981). The Gaussian Decomposition method assumes that the phytoplankton absorption spectra can be decomposed into a number of discrete Gaussian curves described by,

$$a_{ph}(\lambda) = \sum_{j=1}^{l} a_{mj}^*(\lambda) C_j \exp\left[\frac{(\lambda - \lambda_{mj})^2}{2\sigma_j^2}\right], \tag{7}$$

where C_j is the concentration of pigment in the *j*th absorption band, with a maximum absorption at λ_{mj} and Gaussian width of σ_j, a_{mj}^* is the specific absorption coefficient and *l* is the total number of Gaussian bands (Hoepffner and Sathyendranath, 1991).

Following Chase et al. (2013), the absorption of NAP, $a_{NAP}(\lambda)$, is modelled as an exponential decreasing function of wavelength:

$$a_{NAP}(\lambda) = a_{NAP}(400) e^{-0.01(\lambda - 400)}, \tag{8}$$

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

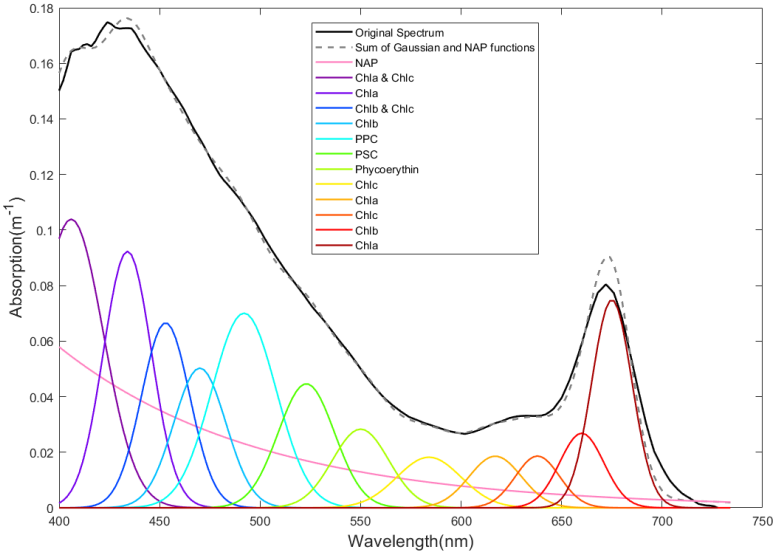

Figure 3. Example of Gaussian Decomposition of an absorption spectrum. Following Chase et al. (2013), Chlorophyll is abbreviated as Chl., Photosynthetic Carotenoids (PSC) are defined as the sum of 19′hexanoyloxyfucoxanthin, fucoxanthin, 19′butanoyloxyfucoxanthin and peridinin concentrations. In contrast, photoprotective pigment concentrations (PPC) are derived from the sum of α-carotene, β-carotene, zeaxanthin, alloxanthin and diadinoxanthin.

where $a_{NAP}(400)$ is the amplitude of NAP absorption at 400 nm (Chase et al., 2013).

We used the methods and code from Chase et al. (2013) to decompose the hyperspectral particulate absorption spectrum into 12 Gaussian curves and NAP absorption. Figure 3 shows an example of the Gaussian decomposition. The amplitudes of these curves have been shown to be strongly related to pigment concentrations (Liu et al., 2019), and the decomposition acts to describe the spectral shape using the minimum number of variables possible (Orkney et al., 2020)

2.11 Principal component analysis of spectra

Principal Component Analysis was used on the Gaussian amplitudes of the spectra to identify whether spectral changes correspond to differences in phytoplankton communities. This method differentiates changes in the absorption spectra that may occur due to changes in the pigment assemblages, concentration and/or phytoplankton size (Morel and Bricaud, 1981; Bricaud et al., 2004; Ciotti et al., 2002). Cael et al. (2020) found that the majority of the variation in the absorption spectra was primarily influenced by changes in spectral amplitude (due to biomass changes). In order to examine the residual variability, we normalized the spectra to the Gaussian peak centred at 470 nm. This peak was chosen as it provided the best separation of

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

Table 2. Best-fit equations for the estimation of the phytoplankton groups. $a_{gauss}(\lambda)$ refers to the gaussian peak centred at wavelength, λ.

Diatom		
CHEMTAX	Coastal	$323.113a_{gauss}(675)^{1.894}$
	All	$982.4a_{gauss}(675)^{2.325}$
DPA	Coastal	$85.9701a_{gauss}(675)^{1.493}$
	Al	$109.7275a_{gauss}(675)^{1.534}$

Offshore Haptophytes	
CHEMTAX	$59.13a_{gauss}(550) + 142.03a_{gauss}(584) - 86.66a_{gauss}(617)$ $-109.28a_{gauss}(638) + 55.48a_{gauss}(675)$
DPA	$-9.29a_{gauss}(470) - 35.61a_{gauss}(550) + 57.88a_{gauss}(584)$ $-40.41a_{gauss}(617) - 28.92a_{gauss}(660) + 14.49a_{gauss}(675)$

the different communities. This is likely due to the peak being tightly correlated with Chla (R=0.90) and also more strongly affected by changes in the pigment packaging effect compared to other parts of the spectrum. (Bricaud et al., 2004). Following normalization, linear discriminant analysis was used to classify the phytoplankton communities, based on a linear combination of input variables (Chla and the second principal component) to differentiate between discrete groups (Fisher, 1936). The accuracy of the analysis was determined using leave-one-out cross-validation (Guo et al., 2007).

2.12 Multiple linear regression

To quantitatively estimate phytoplankton composition from hyperspectral data, we employed multiple linear regression (MLR) of the Gaussian amplitudes. This method exploits the differences in pigment concentrations and ratios between phytoplankton groups, which leads to varying ratios of the Gaussian amplitudes. Different numbers and combinations of the 12 Gaussian amplitudes (a_{gauss}) were regressed against pigment-based derived phytoplankton composition, with an accuracy determined by computing the R^2 value. The best-fit equations shown in Table 2 were determined by sequentially increasing the number of gaussian peaks used until further additions resulted in negligible (<5%) improvements in the R^2 value. In this method, we used CHEMTAX and DPA-derived biomass of diatoms and haptophytes as the independent variables. Prior to the regression analysis, data were log-transformed to account for the wide range of the diatom biomass encountered across our sampling region (10^{-3} to 10^{1} mg m^{-3}). We also examined whether separating the data between offshore and coastal waters improved the accuracy of phytoplankton composition estimation, as reported below.

2.13 Chla algorithm

We compared our hyperspectral estimates of phytoplankton composition to the regionally-tuned Chla based algorithm of Zeng et al. (2018) based on the Chla-alogrithm of Hirata et al. (2011), which was tuned against DPA-derived taxonomic estimates.

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

This algorithm was found to be more accurate in deriving subarctic Pacific phytoplankton functional types and size classes compared to the gloabal Chla algorithm of Hirata et al. (2011), the spectral end-member algorithms of Zhang et al. (2013) and Zhang et al. (2015), and the back-scattering slope algorithm of Kostadinov et al. (2009),

3 Results

3.1 Phytoplankton community

3.1.1 Pigment-based clustering

Analysis of HPLC pigment data indicates that the first two principal components describe 51% of the variation in the dataset, and can be used to cluster pigment distributions into three groups with different dominant taxa. The derived clusters are shown in Fig. 4, and were named based on the phytoplankton group with the highest abundance based on CHEMTAX analysis shown in Table 3. The three dominant clusters have significantly different proportions of the different phytoplankton groups (F-statistics and *p*-values presented in Table 3, with the exception of dinoflagellates, which are present in low percentages across all three clusters). Whereas the diatom and haptophyte-dominated clusters are largely composed of those two groups, the mixed-assemblage group represents a diverse community with significant proportions of diatoms, haptophytes, prasinophytes, and cyanobacteria.

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

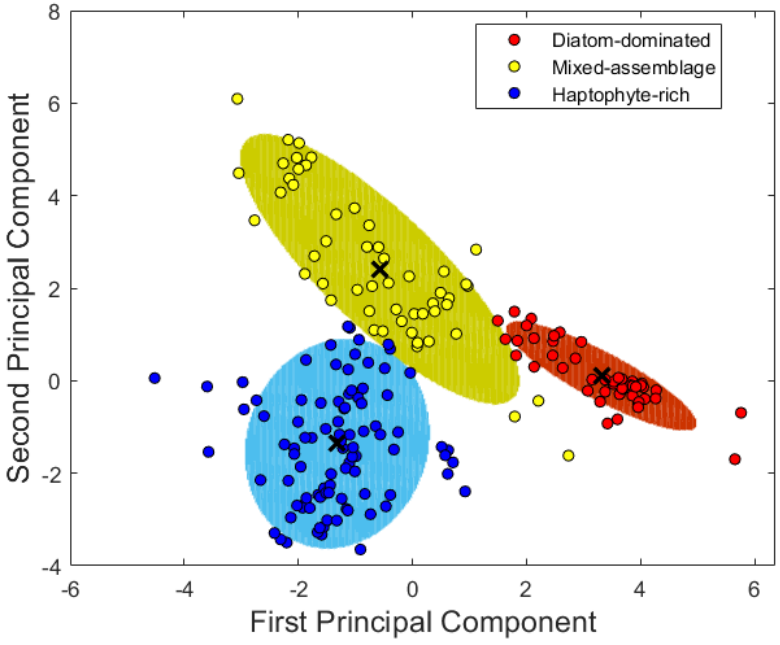

Figure 4. Clusters of phytoplankton community dominance identified from the principal components of the HPLC pigments. Shaded region represent the 99% confidence ellipsoids of each cluster.

Table 3. The mean proportions of phytoplankton groups in the clusters described above as determined from CHEMTAX analysis. Proportions of Pelagophytes were only determined in offshore stations and Dictyophytes in coastal stations.

	Cluster				
	Diatom	Mixed	Haptophyte		
Phytoplankton group	Mean Proportion in the cluster			F-statistic	*p*-value
Cyanobacteria	0.01	0.12	0.08	17.21	1.3e-7
Chlorophytes	0.01	0.05	0.16	56.72	4.6e-20
Prasinophytes	0.04	0.24	0.05	111.62	7.78e-33
Cryptophytes	0.06	0.15	0.05	30.10	4.34e-12
Diatoms	0.80	0.22	0.06	515.26	1.10e-77
Dinoflagellates	0.03	0.04	0.05	2.09	0.13
Haptophytes	0.04	0.16	0.47	282.10	9.64e-58
Pelagophytes*	-	0.02	0.07	19.8	2.55e-05
Dictyophytes*	0.01	0.03	0.14	126.2	1.53e-28

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

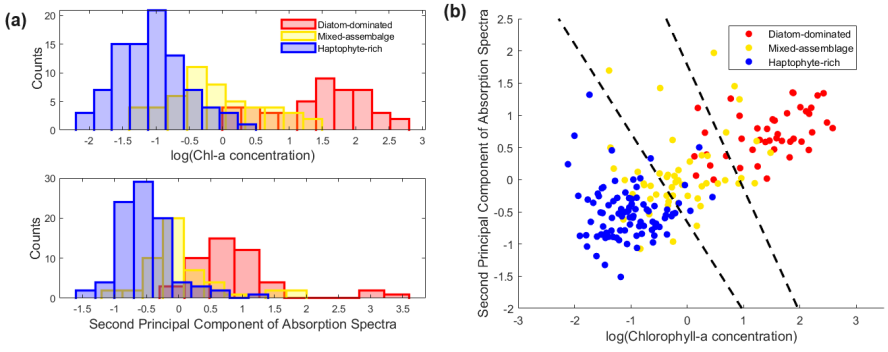

Figure 5. Separation of phytoplankton groups based on principle components analysis of optical data. Panel A shows the range of Chlorophyll-a (Chla) and the second principal component of the different phytoplankton communities. Panel B shows how combining Chla and the second principal produces a good separation between the different clusters. The dotted line shows the linear decision boundaries determined by linear discriminant analysis.

3.1.2 Assessment of phytoplankton communities using optical data

The distinct phytoplankton communities identified from the pigment-based clustering occupy an overlapping range of Chla values (Fig. 5a), making it difficult to discriminate the groups using this variable alone. In contrast, principal component analysis of the hyperspectral data can be used to separate the distinct phytoplankton communities. The first two principal components explain 87% of the total variation in the absorption spectra. The first principal component of absorption spectra was not strongly related to the phytoplankton community, and analysis of the eigenvectors suggests that it is more strongly related to the inversion of NAP as part of the Gaussian decomposition. In contrast, the second principal component of the absorption spectra is more closely related to the pigment-derived phytoplankton community dominance (Fig. 5a). When used in combination with Chla data, this second principle component provides good separation between three main phytoplankton communities (diatom, haptophyte and mixed-assemblage), as shown in (Fig. 5b).

Given that the three taxonomic clusters appear to separate along linear boundaries, we used Linear Discriminant analysis as a method to classify the phytoplankton communities based on Chl concentrations and the second PCA of hyperspectral data. The prediction accuracy was found to be 79% (N=192), based on leave-one-out cross-validation. The bulk of the error comes from the misclassification of the mixed-assemblage, with good separation between the diatom and haptophyte clusters.

3.1.3 Assessment of phytoplankton size of the different communities

Figure 6 and Table 4 show the size distributions of the different phytoplankton communities derived from the wavelength-dependent slope of particulate light attenuation. Based on comparisons with size-fractionated data, phytoplankton in the diatom-dominated communities can be inferred to be predominantly $> 20\mu m$, while both the mixed-assemblage and haptophyte-

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

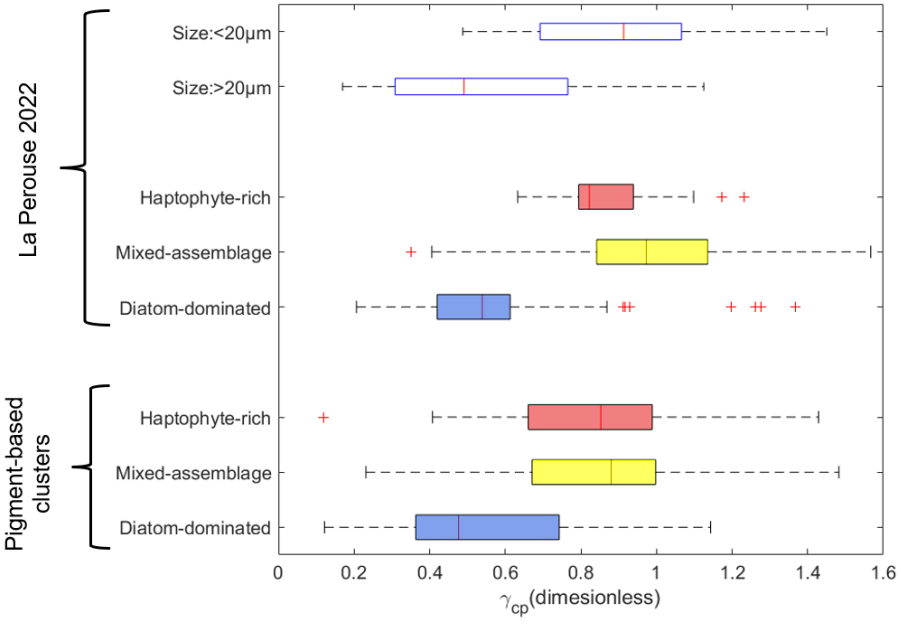

Figure 6. Boxplot shows the distribution of gamma values associated with the phytoplankton communities (pigment-based and optically based for La Perouse 2022) and size classes. Red pluses (+) indicate outliers which are determined as points whose values are greater than the 1.5 of the interquartile range from the 25th or 75th percentile.

rich communities are mainly composed of phytoplankton < $20\mu m$. The wide range of apparent sizes in the optically based mixed assemblage community is attributed to the diversity of phytoplankton comprising it.

3.2 Phytoplankton composition

In deriving quantitative estimates of phytoplankton composition, we focused specifically on diatoms and haptophytes, as both these groups are abundant in our study region and play distinct but important roles in ecology and biogeochemistry (Peña et al., 2019b). As shown in Table 5 and Fig. 7, the MLR approach was able to accurately reproduce observed diatom biomass based on both DPA and CHEMTAX analysis, explaining more than 70% of observed variability. The best-fit equations obtained from this analysis predict diatom biomass using only the gaussian peak at 675nm (Table 2), which is strongly associated with total Chla concentrations (Bricaud et al., 2004; Chase et al., 2013). The strong correlation between diatom biomass and Chla concentrations explains the accuracy of the Chla based approach of Zeng et al. (2018). This result, in turn, reflects the dominance of diatoms in the high productivity coastal waters of the eastern subarctic Pacific.

In contrast to diatoms, haptophytes are more abundant in offshore waters. As shown in Fig. 8 and Table 5, the MLR analysis was able to explain 63% of the observed variability in haptophyte abundance in oceanic stations as compared to CHEMTAX and DPA analysis. However, the accuracy of the method decreases significantly (R^2 = 0.3) when data from coastal and offshore

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

Table 4. The table shows γ_{cp} values of the different phytoplankton communities (pigment-based and optically based for La Perouse 2022) and size classes measured. Two-sample t-tests were used to compare the distribution of γ_{cp}, to that of the size-fractionated data from La Perouse 2022. Std Dev. refers to the standard deviation and T-stat refers to the t-static.

	Pigment based clusters					
	Diatom		Mixed		Haptophyte	
Mean γ_{cp}	0.541		0.830		0.842	
Std Dev. of γ_{cp}	0.236		0.260		0.228	
N	47		50		94	
Size Class	T-stat	p-value	T-stat	p-value	T-stat	p-value
$> 20\mu m$	-0.0624	0.950	3.43	0.001	4.25	4.70e-5
$< 20\mu m$	-5.66	3.98e-7	-1.36	0.179	-1.44	0.15

	La Perouse 2022								
	Diatom		Mixed		Haptophyte		$< 20\mu m$		$> 20\mu m$
Mean γ_{cp}	0.532		1.01		0.862		0.930		0.546
Std Dev. of γ_{cp}	0.161		0.265		0.126		0.277		0.291
N	2578		1916		1428		18		13
Size Class	T-stat	p-value	T-stat	p-value	T-stat	p-value	T-stat	p-value	
$> 20\mu m$	-0.4973	0.620	6.28	4.18e-10	8.85	2.58e-18	-3.7262	8.37e-4	
$< 20\mu m$	-10.6	8.22e-26	1.28	0.201	-2.2075	0.0274			

waters are combined. By comparison, the Chla based approach of Zeng et al. (2018) has lower accuracy for haptophyte prediction, with an R^2 of 0.43 and <0.05 in offshore and coastal waters, respectively. This result indicates that the MLR method performs better than the Chla based approach in predicting Haptophyte biomass across our sampling region. Notably, the

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

Table 5. Comparison of the Chla based approach and the multivariate linear regression (MLR) in estimating the relative biomass of diatoms (all and coastal stations) and Haptophytes biomass (all stations and offshore stations only). Values from these algorithms are compared against CHEMTAX and DPA analysis of discrete samples analysed via HPLC. The uncertainties given are the 95% confidence bounds. Slopes approaching 1 indicate no bias in the derived values compared to measurements.

		Chla based approach			MLR	
		Slope	Intercept	R^2	R^2	N
Diatoms	CHEMTAX Coastal Diatoms	-	-	-	0.81	104
	CHEMTAX All Diatoms	-	-	-	0.71	192
	DPA Coastal Diatoms	0.86±0.06	0.03±0.18	0.89	0.89	104
	DPA All Diatoms	0.86±0.04	0.03±0.09	0.91	0.91	192
Haptophytes	CHEMTAX Offshore Hapt	-	-	-	0.63	87
	CHEMTAX All Hapt	-	-	-	0.3	192
	DPA Offshore Hapt	0.88±0.20	-0.05±0.03	0.43	0.63	87
	DPA All Hapt	-0.01±0.07	0.08±0.2	0	0.28	192

Gaussian peaks (a_{gauss}) used in our analysis correlated with the absorption bands of key pigments found in haptophytes (Table 2). Gaussian peaks used to estimate the biomass of nanophytoplankton (470, 550, 584, 617, 638 and 675 nm) correspond to absorption by Chlorophyll-a, photosynthetic carotenoids (19'-Hex, 19'-But, Fucoxanthin) and Chlorophyll-c (Bricaud et al., 2004), pigments strongly associated with haptophytes.

3.3 Oceanographic controls on phytoplankton community composition

We used the methods described above to estimate the phytoplankton community and composition during August 2022 La Perouse cruise, where pigment data are currently unavailable. As shown in Fig. 9, the relative biomass of diatom increased closer to the shore, associated with lower sea surface temperature and salinity. In contrast, haptophyte-abundant communities were found further offshore, in regions with warmer and more saline oceanic waters. This predicted distribution agrees with late summer observations from pigment measurements from 2016-2018 and other studies in this region (Peña et al., 2019b), suggesting the utility of our approach to reconstruct meaningful phytoplankton taxonomic distributions without pigment data. The changes in community composition along the coastal-offshore gradient likely reflect the transition from a coastal macronutrient-rich water mass to more nutrient-depleted offshore waters, as described by (Boyd and Harrison, 1999). The high diatom biomass found in the south is likely sustained by the nitrate/silicate supplied by the Juan de Fuca Strait outflow (Crawford and Dewey, 1989).

In addition to examining spatial variability, we compared the phytoplankton communities across the different seasons sampled over a 3-year period between 2016 - 2018. We found that seasonal variability was most pronounced in coastal waters,

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

Table 6. Results from multiple logistic regressions which estimate the probability of the different phytoplankton communities (each performed with a comparison of two of the three communities), given the nutrient concentrations, Sea Surface Temperature (SST) and salinity. The estimates of the regression and the *p*-values for each variable are shown above. Bolded variables are statistically significant for an alpha value of 0.05.

	Diatom/Mixed		Mixed/Haptophyte		Diatom/Haptophyte	
	Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value
Intercept	-14.1290	0.3626	**-118.7865**	**0.0000**	**-136.8695**	**0.0000**
Nitrate	0.2834	0.3534	0.3157	0.3530	-0.5176	0.2542
Silicate	**-0.4535**	**0.0023**	0.0513	0.6140	**-0.4352**	**0.0024**
Phosphate	5.4871	0.1212	-6.8738	0.1160	**13.1026**	**0.0482**
SST	**0.7434**	**0.0027**	0.0817	0.4990	**0.5557**	**0.0070**
Salinity	0.1594	0.75818	**3.7815**	**0.0000**	**4.0650**	**0.0000**

with phytoplankton communities transitioning from diatom-rich during the early spring bloom to haptophyte-rich and mixed assemblage later in the summer (Fig. 10.) In contrast, the phytoplankton community in offshore waters remains haptophyte-rich throughout all seasons, in agreement with the results of Peña and Varela (2007) who found that phytoplankton biomass shows low seasonal variability offshore.

We used multiple logistic regressions to identify the potential oceanographic drivers of phytoplankton taxonomic variability across our sampling region, with a particular focus on nutrient concentrations, salinity and sea surface temperature. The results (Table 6) of comparisons between the drivers of each community suggest that the occurrence of diatom-dominated communities is strongly regulated by silicate concentrations and SST (which reflects seasonality). By comparison, the Haptophyte-rich community are found further offshore (indicated by the significance of salinity). The intercept of the logistic regression was also found to be statistically significant, suggesting that additional variables, such as light availability, micronutrients, or zooplankton grazing may also play an important role in controlling taxonomic distributions.

3.4 DMS and DMSP concentrations

Figure 11 shows the relationship between the phytoplankton community and the Chla normalized concentrations of DMS and DMSP. Results from one-way ANOVA demonstrated statistically-significant differences between the DMS/Chla ($p <$ .001, N=162) and DMSP/Chla ($p <$ .001, N=180) across the different phytoplankton communities ($p <$ .001, N=152). As expected from previous studies, haptophyte-rich communities showed the highest DMS and DMSP concentrations per unit Chla, reflecting the high DMS/P production by this group (Keller, 1989). The contribution of dinoflagellates is largely low and invariant between the different groups, and thus unlikely to explain DMS/P variability across the different communities.

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

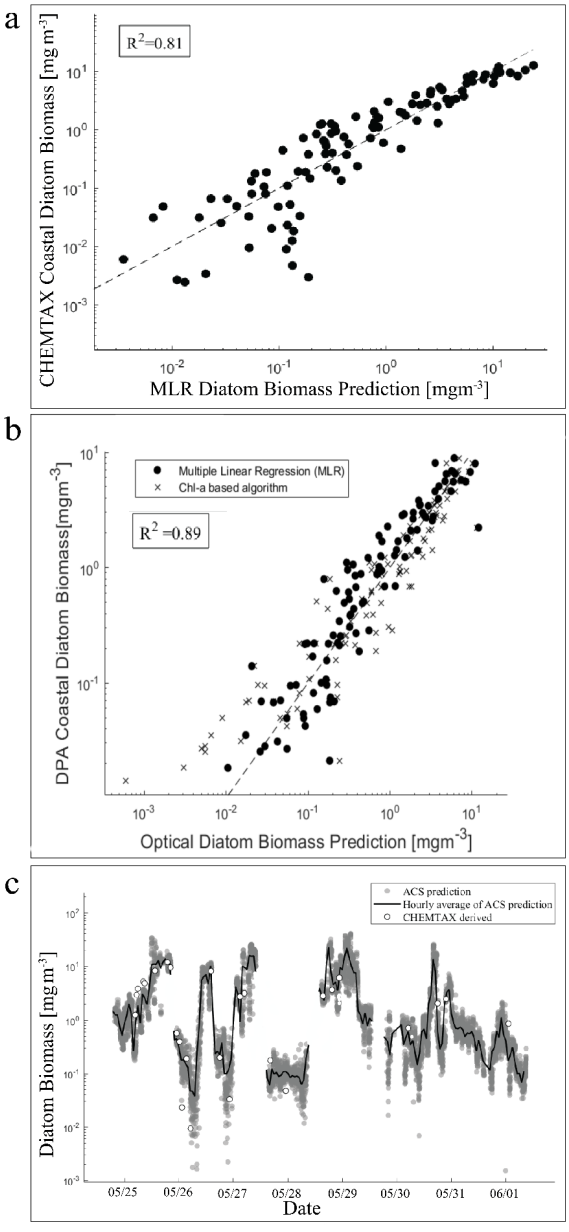

Figure 7. Relationship between multivariate linear regression (MLR)-predicted and pigment-derived diatom biomass in coastal waters of the northeast subarctic Pacific for (a) CHEMTAX-derived diatoms [R^2=0.81, N=103], and (b) DPA-derived Diatoms [R^2=0.89, N=103]. The dashed line represents the 1:1 relationship. The Chla based estimation are shown in (b). Panel (c) shows predicted diatom biomass during the 2017 August La Perouse cruise calculated using the power law relationship derived from MLR analysis (Table 2). The grey symbols and black line show values derived from underway hyperspectral measurements (raw data and hourly average, respectively), while the white dots show estimates derived from CHEMTAX analysis of pigment data at discrete stations.

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

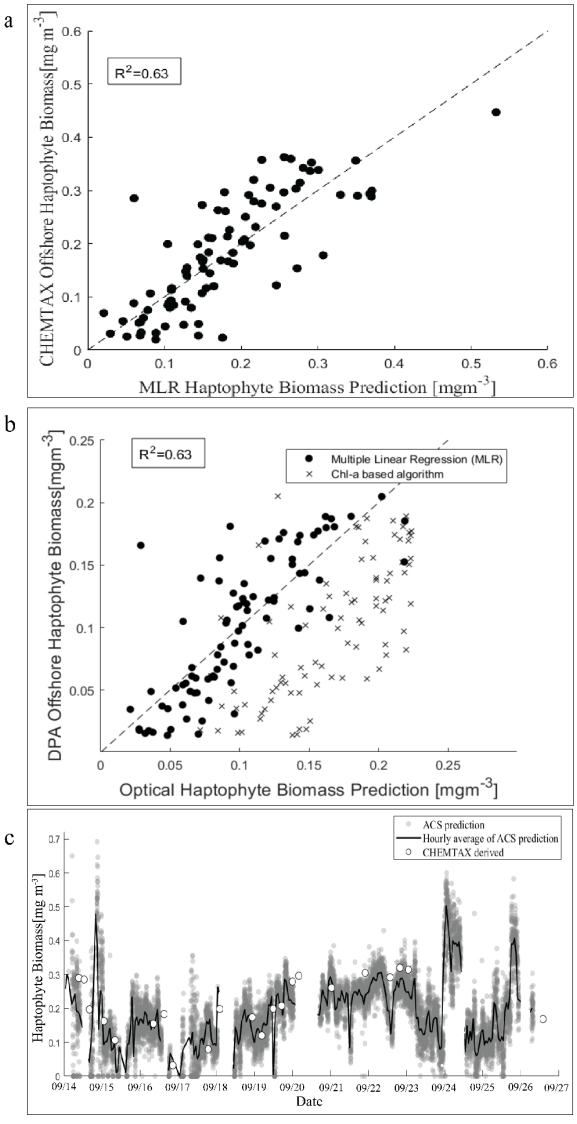

Figure 8. Relationship between multivariate linear regression (MLR)-predicted and pigment-derived haptophyte biomass in offshore waters of the northeast subarctic Pacific for (a) CHEMTAX-derived Haptophytes [R^2=0.63, N=87], and (b) DPA-derived Haptophytes [R^2=0.63, N=87]. The dotted line represents the 1:1 relationship. The Chla based estimation are shown in (b). Panel (c) shows predicted haptophyte biomass during the 2018 Line P cruise calculated using the power law relationship derived from MLR analysis (Table 4). The grey symbols and black line show values derived from underway hyperspectral measurements (raw data and hourly average, respectively), while the white dots show estimates derived from CHEMTAX analysis of pigment data at discrete stations.

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

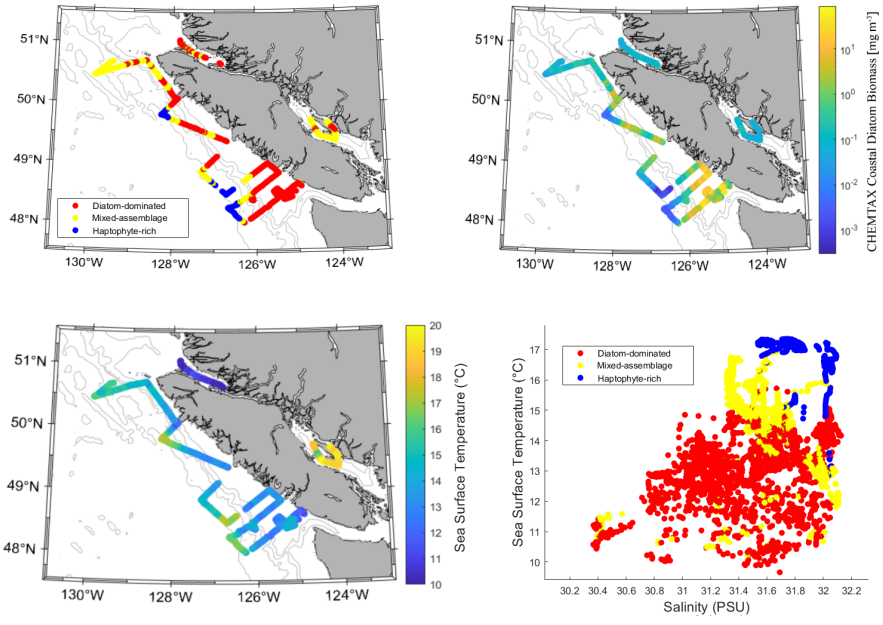

Figure 9. Top panels show Phytoplankton community and the CHEMTAX-derived diatom biomass using equation listed in Table 2 during La Perouse 2022 cruise as determined using the methods listed above. The bottom panels show hydrographic properties during the cruise. The bottom left is the sea surface temperature measured during the cruise, while the bottom right is a scatter plot of sea surface temperature against salinity. The points are coloured with the different communities as determined using the methods listed above.

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

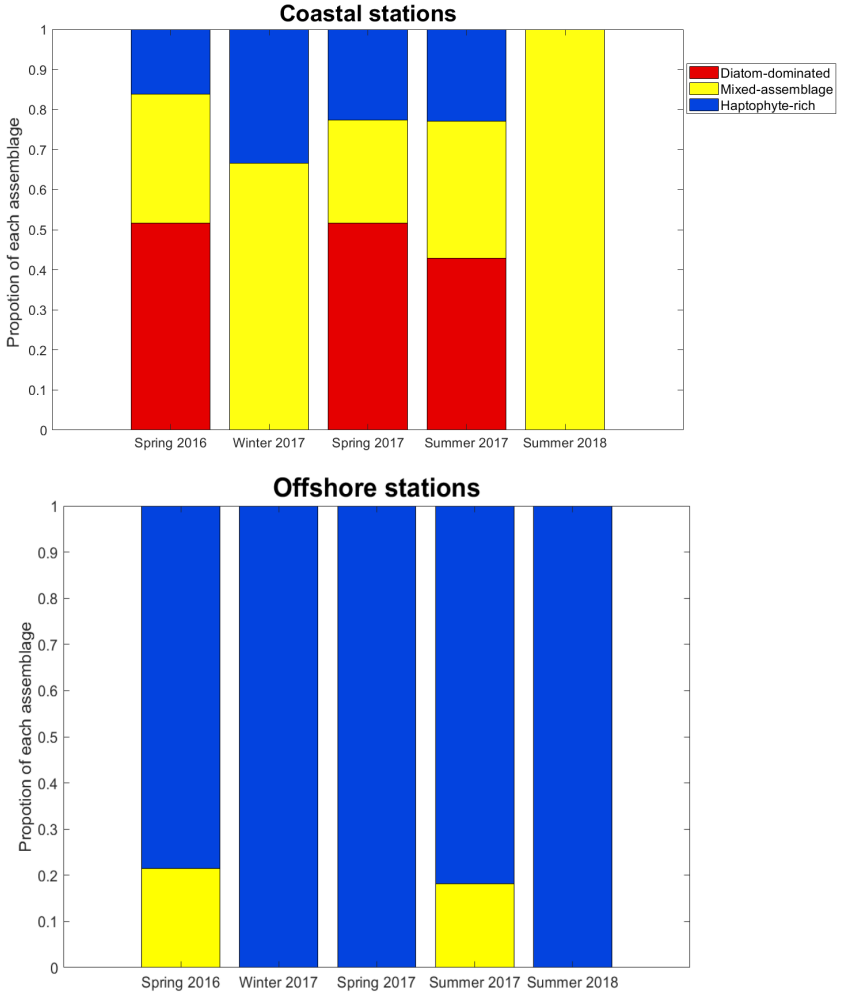

Figure 10. Proportions of the different phytoplankton communities sampled during each cruise season. Coastal stations are shown on the left and offshore stations on the right.

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

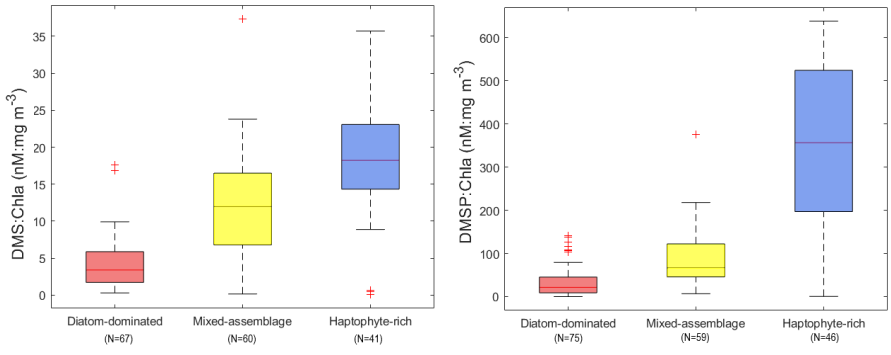

Figure 11. DMS/Chla(left; nM:mg m-3) and DMSP/Chla (right; nM:mg m-3) ratios of the different phytoplankton communities observed during the 2022 La Perouse cruise.

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

4 Discussion

Our results provide further evidence that ship-based hyperspectral absorption data can be used to derive phytoplankton community and composition with high spatial resolution and reasonable accuracy, particularly for diatoms and haptophytes. This work builds on previous studies, with a focus, for the first time, on the NE Subarctic Pacific. As shown in Fig. 8, and Table 5, the approach used here performs significantly better than Chla based algorithms (particularly for haptophytes), and can produce high spatial-resolution data capturing sharp gradients in diatom and haptophyte abundance across oceanographic frontal zones. This information may be utilized alongside other underway measurements, including C:Chl ratios (Fox et al., 2022), photosynthetic electron transport rates from fast repetition rate fluorometry, $\Delta O_2/Ar$ and pCO_2, to better understand the roles phytoplankton community and composition in ocean biogeochemistry.

Within our study region in the eastern Subarctic Pacific, the biomass of diatoms appears to vary most among the different phytoplankton groups, correlating well with the first principal component of the pigment data set. However, we find that the overall changes in the phytoplankton community cannot be completely described using only diatom concentrations, with important changes in groups such as haptophytes, prasinophytes, cryptophytes, and cyanobacteria occurring somewhat independently from diatoms. This underscores the need to characterize both the phytoplankton community in addition to the composition of the individual phytoplankton groups, to fully examine changes in the phytoplankton taxonomy in this region.

To validate our optically-derived taxonomic classification, we used pigment-based methods as a benchmark for phytoplankton community and composition. These pigment-based methods remain the most commonly used approach to assess phytoplankton taxonomy (Kramer et al., 2020) and validate optical algorithms to retrieve different phytoplankton taxonomic groups (Mouw et al., 2017). It is important to note, however, that variability in pigment concentrations may result not only from changes in phytoplankton taxonomy, but also from changing cell physiology in response to environmental conditions (van Oijen et al., 2004). For this reason, our approach to determining the phytoplankton community using the principal components of the pigment dataset does not make any explicit assumptions about the pigment content but is rather based on the observed statistical clustering among pigment groups. This approach follows that used by Catlett and Siegel (2018) and Kramer et al. (2020) to classify different phytoplankton communities, and has the advantage that no prior information or assumptions are needed. The strong coherence between the resulting pigment clusters and CHEMTAX-derived composition suggests that the chemotaxonomic methods within our study area reflect unique signatures of different phytoplankton groups.

Although CHEMTAX and DPA produced slightly different estimates of diatoms and haptophyte biomasses, we found that the MLR method worked equally well against both of these validation methods. We thus expect that our results should apply to a range of chemotaxonomic methods used to derive phytoplankton compositions. The inherent uncertainties stemming from assumptions and variations within the pigment-taxa relationship in chemotaxonomic methods are carried forward into any optical algorithms and are an important source of error. Future studies should therefore focus on comparing optical data with

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

the phytoplankton taxonomy derived from different (i.e. non-pigment-based) means, including automated quantitative cell imagery, and flow cytometry to capture smaller species. This will help validate the robustness of pigment-based classification (Chase et al., 2020).

Diatoms tend to dominate high Chla waters, and their biomass is strongly correlated with total Chla. This explains why Chla is a useful predictor of diatom abundance (Zeng et al., 2018). In contrast, other phytoplankton groups tend to be found in more mixed phytoplankton assemblages, where the relationship between composition and Chla is more complex. For these phytoplankton groups, the incorporation of hyperspectral data significantly improved taxonomic resolution as compared to Chla data alone, agreeing with previous studies (Catlett and Siegel, 2018). Further improvements in method performance could be obtained using more complex non-linear approaches (El Hourany et al., 2019).

The methods presented here add to a growing number of approaches aiming to link optical measurements to phytoplankton community composition (Alvain et al., 2005; Mao et al., 2010; Isada et al., 2015). These methods (including those presented here) are subject to errors and uncertainty from a number of sources. These include uncertainties in pigment-taxa relationships (elaborated further above), variations in optical properties due to physiological conditions and regional variations (Robinson et al., 2021), statistical noise associated processing and collection of optical data (Slade et al., 2010; Brewin et al., 2016). Nonetheless, the results presented here suggest that hyperspectral methods can estimate the phytoplankton community composition and sizes in our region with sufficient accuracy to investigate the oceanographic controls on broad-scale community composition and biogeochemical variability of DMS and DMSP. Indeed, the reproduced distribution of phytoplankton community, sizes and diatom concentration shown in Fig. 9 is broadly consistent with our current understanding of phytoplankton biogeography in this region.

5 Conclusions

We were able to relate the optically derived phytoplankton community composition to phytoplankton sizes and other oceanographic and biogeochemical variables, highlighting the utility of hyperspectral data to better understand oceanic processes. While the multiple logistic regression used in this study may not resolve all the complex processes that govern phytoplankton growth, it was able to highlight key factors influencing the distribution of different taxonomic groups. The inclusion of phytoplankton taxonomic information into biogeochemical models such as Peña et al. (2019a), may be useful for improving their accuracy. Similarly, given the significant differences between the DMS and DMSP concentrations for the different phytoplankton communities, hyperspectral data could be used to improve the predictive algorithms for DMS and DMSP (e.g. McNabb and Tortell (2022)). Optically derived phytoplankton taxonomic estimates could also help assess whether DMS "hotspots" arise from floristic shifts (i.e. high DMS associated with high biomass of strong DMS/P producers), as opposed to environmental factors (e.g. high PAR, strong nutrient stress, mixing) (McParland and Levine, 2019).

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

Data availability. Data used in the publication are available on request

Appendix A: CHEMTAX pigment ratios

Table A1. Initial CHEMTAX pigment ratio matrix used for the coastal stations.

Class/pigments	Chl c3	Chl c2	Chl c1	Peri	19-but	Fuco	Prasin	Viola	19-hex	Allo	Zea	Lutein	Chl b	Chl a
Cyanobacteria	0	0	0	0	0	0	0	0	0	0	0.64	0	0	1
Chlorophytes	0	0	0	0	0	0	0	0.05	0	0	0.03	0.17	0.32	1
Prasinophytes	0	0	0	0	0	0	0.25	0.05	0	0	0.06	0.02	0.73	1
Cryptophytes	0	0.2	0	0	0	0	0	0	0	0.38	0	0	0	1
Diatoms	0	0.15	0.03	0	0	0.85	0	0	0	0	0	0	0	1
Dinoflag	0	0.22	0	0.56	0	0	0	0	0	0	0	0	0	1
Haptophytes	0.18	0.21	0	0	0.04	0.29	0	0	0.47	0	0	0	0	1
Dictophyte	0.06	0.13	0	0	0.38	0.24	0	0	0	0	0	0	0	1

Table A2. Initial CHEMTAX pigment ratio matrix used in the offshore stations

Class/pigments	Chl c3	Peri	19-but	Fuco	19-hex	Allo	Zea	Chl b	Chl a
Cyanobacteria	0	0	0	0	0	0	0.64	0	1
Chlorophytes	0	0	0	0	0	0	0.03	0.32	1
Prasinophytes	0	0	0	0	0	0	0.06	0.73	1
Cryptophytes	0	0	0	0	0	0.38	0	0	1
Diatom-2	0.08	0	0	0.99	0	0	0	0	1
Dinoflag-1	0	0.56	0	0	0	0	0	0	1
Haptophytes	0.18	0	0.04	0.29	0.47	0	0	0	1
Pelagophytes	0.23	0	0.66	0.78	0	0	0	0	1

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

Appendix B: CHEMTAX and DPA comparisons

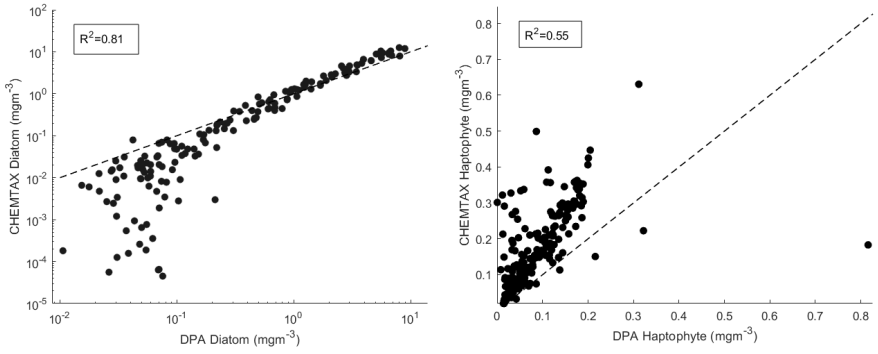

Figure B1. A comparison of derived diatom and haptophyte abundance from DPA and CHEMTAX. The dashed line indicates a 1:1 ratio.

Appendix C: Gaussian Mixture Model Clustering

The Gaussian Mixture Model clustering assumes that the data can be characterized by a mixture of Gaussian distributions. Each distribution is described by a mean and variance. An Expectation-Maximization algorithm is used to estimate the probabilities of data points to the clusters(McLachlan and Peel, 2000). To perform the clustering, we needed to describe the initial conditions, covariance matrix type, and number of clusters. The initial conditions were randomly generated, and we specified that the covariance matrix was full (both non-diagonal and not shared).

The number of clusters to be fit is the most subjective part of the process. We chose to fit 3 clusters based on a combination of visual inspection and examining the hierarchical relationship between pigment ratios to Chla. We followed the methods of Kramer et al. (2020) to examine hierarchical clusters within the pigment dataset. The resulting dendrogram is shown in Fig. C1. Using a linkage cutoff distance of 1.5, three clusters can be identified. Taxonomic designations of the hierarchical clusters are based on the pigment ratios associated with major phytoplankton taxa. The three clusters identified are similar to those classified using the Gaussian Mixture Modelling clusters on the principal components of the pigment dataset.

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

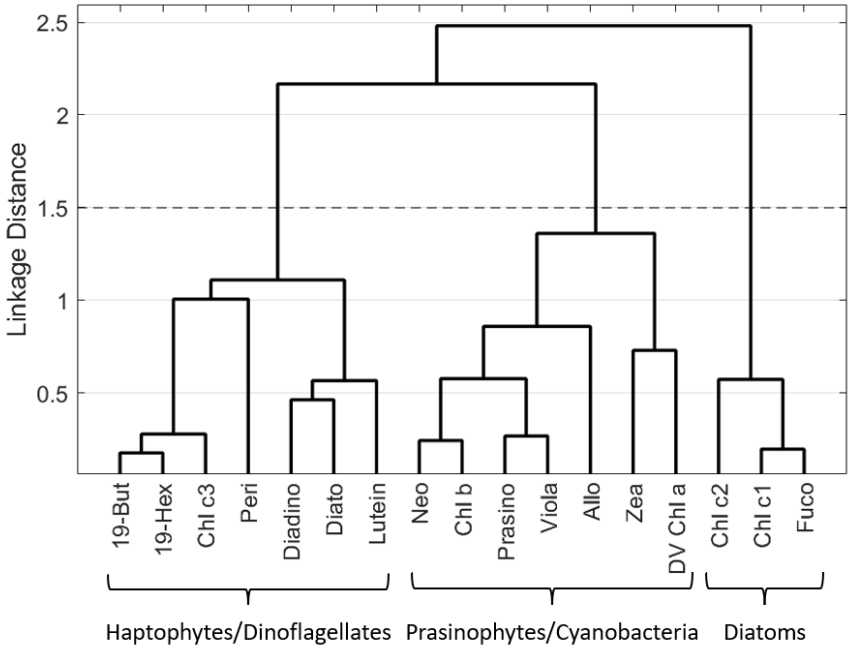

Figure C1. Dendrogram of phytoplankton pigment ratios to Chla, with different clusters defined by a linkage distance cutoff of 1.5.Abbreviations used: 19-But=19′butanoyloxyfucoxanthin, 19-Hex=19′hexanoyloxyfucoxanthin, Chl c3=Chlorophyll-c3, Peri=peridinin, Diadino=Diadinoxanthin, Diato=Diatoxanthin, Neo=Neoxanthin, Chl b=Chlorophyll-b, Prasino= Prasinoxaxnthin, Viola=Violaxanthin, Allo=Alloxanthin, Zea= Zeaxanthin, DV Chl a=Divinyl Chlorophyll-a, Chl c2=Chlorophyll c2, Chl c1=Chlorophyll c1, Fuco=Fucoxanthin.

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

Author contributions. SP and PT conceived of the study. SP analysed the data and wrote the manuscript with significant input from PT. AP collected the HPLC data, performed the CHEMTAX analysis and provided feedback on the manuscript, WB collected the Optical data and provided feedback on the manuscript. BM measured the concentration of DMS and DMSP, and collected some of the optical data used in this paper.

Competing interests. The authors declare that they have no conflict of interest

Acknowledgements. We would like to thank the crew and officers of the CCGS *John P. Tully* for assistance during Line P and La Perouse cruises, and the science staff at the Institute of Ocean Sciences and Fisheries and Oceans Canada for their efforts in maintaining a consistent, long-term sampling program in the Eastern Subarctic Pacific. This project was funded by grants provided by the Natural Sciences and Engineering Council of Canada

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

References

Alvain, S., Moulin, C., Dandonneau, Y., and Breon, F.-M.: Remote sensing of phytoplankton groups in case 1 waters from global SeaWiFS imagery, Deep Sea Research Part I: Oceanographic Research Papers, 52, 1989–2004, 2005.

Anderson, T. R.: Plankton functional type modelling: running before we can walk?, Journal of plankton research, 27, 1073–1081, 2005.

Archer, D., Winguth, A., Lea, D., and Mahowald, N.: What caused the glacial/interglacial atmospheric pCO2 cycles?, Reviews of Geophysics, 38, 159–189, 2000.

Asher, E. C., Dacey, J. W., Jarniková, T., and Tortell, P. D.: Measurement of DMS, DMSO, and DMSP in natural waters by automated sequential chemical analysis, Limnology and Oceanography: Methods, 13, 451–462, 2015.

Barwell-Clarke, J. and Whitney, F. a.: Institute of Ocean Sciences nutrient methods and analysis, Citeseer, 1996.

Behrenfeld, M. J., O'Malley, R. T., Boss, E. S., Westberry, T. K., Graff, J. R., Halsey, K. H., Milligan, A. J., Siegel, D. A., and Brown, M. B.: Revaluating ocean warming impacts on global phytoplankton, Nature Climate Change, 6, 323–330, 2016.

Borcard, D., Gillet, F., and Legendre, P.: Unconstrained Ordination, chap. 5, pp. 151–201, Springer, 2018.

Boss, E., Twardowski, M. S., and Herring, S.: Shape of the particulate beam attenuation spectrum and its inversion to obtain the shape of the particulate size distribution, Applied Optics, 40, 4885–4893, 2001.

Boss, E., Picheral, M., Leeuw, T., Chase, A., Karsenti, E., Gorsky, G., Taylor, L., Slade, W., Ras, J., and Claustre, H.: The characteristics of particulate absorption, scattering and attenuation coefficients in the surface ocean; Contribution of the Tara Oceans expedition, Methods in Oceanography, 7, 52–62, 2013.

Boss, E. S., Collier, R., Pegau, W., Larson, G., and Fennel, K.: Measurements of spectral optical properties and their relation to biogeochemical variables and processes in Crater Lake, Crater Lake National Park, OR, in: Long-term Limnological Research and Monitoring at Crater Lake, Oregon, pp. 149–159, Springer, 2007.

Boyd, P. and Harrison, P.: Phytoplankton dynamics in the NE subarctic Pacific, Deep Sea Research Part II: Topical Studies in Oceanography, 46, 2405–2432, 1999.

Brewin, R. J., Dall'Olmo, G., Pardo, S., van Dongen-Vogels, V., and Boss, E. S.: Underway spectrophotometry along the Atlantic Meridional Transect reveals high performance in satellite chlorophyll retrievals, Remote sensing of environment, 183, 82–97, 2016.

Bricaud, A., Claustre, H., Ras, J., and Oubelkheir, K.: Natural variability of phytoplanktonic absorption in oceanic waters: Influence of the size structure of algal populations, Journal of Geophysical Research: Oceans, 109, 2004.

Burt, W. J., Westberry, T. K., Behrenfeld, M. J., Zeng, C., Izett, R. W., and Tortell, P. D.: Carbon: Chlorophyll ratios and net primary productivity of Subarctic Pacific surface waters derived from autonomous shipboard sensors, Global Biogeochemical Cycles, 32, 267–288, 2018.

Cael, B., Chase, A., and Boss, E.: Information content of absorption spectra and implications for ocean color inversion, Applied Optics, 59, 3971–3984, 2020.

Catlett, D. and Siegel, D.: Phytoplankton pigment communities can be modeled using unique relationships with spectral absorption signatures in a dynamic coastal environment, Journal of Geophysical Research: Oceans, 123, 246–264, 2018.

Cetinic, I., Perry, M., D'asaro, E., Briggs, N., Poulton, N., Sieracki, M., and Lee, C.: A simple optical index shows spatial and temporal heterogeneity in phytoplankton community composition during the 2008 North Atlantic Bloom Experiment, Biogeosciences, 12, 2179–2194, 2015.

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

Chase, A., Boss, E., Zaneveld, R., Bricaud, A., Claustre, H., Ras, J., Dall'Olmo, G., and Westberry, T. K.: Decomposition of in situ particulate absorption spectra, Methods in Oceanography, 7, 110–124, 2013.

Chase, A., Boss, E., Haëntjens, N., Culhane, E., Roesler, C., and Karp-Boss, L.: Plankton Imagery Data Inform Satellite-Based Estimates of Diatom Carbon, Geophysical Research Letters, 49, e2022GL098 076, 2022.

Chase, A. P., Kramer, S. J., Haëntjens, N., Boss, E. S., Karp-Boss, L., Edmondson, M., and Graff, J. R.: Evaluation of diagnostic pigments to estimate phytoplankton size classes, Limnology and Oceanography: Methods, 18, 570–584, 2020.

Ciotti, A. M., Lewis, M. R., and Cullen, J. J.: Assessment of the relationships between dominant cell size in natural phytoplankton communities and the spectral shape of the absorption coefficient, Limnology and Oceanography, 47, 404–417, 2002.

Crawford, W. R. and Dewey, R. K.: Turbulence and mixing: Sources of nutrients on the Vancouver Island continental shelf, Atmosphere-ocean, 27, 428–442, 1989.

Del Bel Belluz, J., Peña, M. A., Jackson, J. M., and Nemcek, N.: Phytoplankton composition and environmental drivers in the northern Strait of Georgia (Salish Sea), British Columbia, Canada, Estuaries and Coasts, 44, 1419–1439, 2021.

DFO: Institute of Ocean Sciences Data Archive, http://www.pac.dfo-mpo.gc.ca/science/oceans/data-donnees/index-eng.html. Data obtained on 2023/06/04, 2023.

El Hourany, R., Abboud-abi Saab, M., Faour, G., Mejia, C., Crépon, M., and Thiria, S.: Phytoplankton diversity in the Mediterranean Sea from satellite data using self-organizing maps, Journal of Geophysical Research: Oceans, 124, 5827–5843, 2019.

Falkowski, P. G., Barber, R. T., and Smetacek, V.: Biogeochemical controls and feedbacks on ocean primary production, science, 281, 200–206, 1998.

Fisher, R. A.: The use of multiple measurements in taxonomic problems, Annals of eugenics, 7, 179–188, 1936.

Fox, J., Kramer, S. J., Graff, J. R., Behrenfeld, M. J., Boss, E., Tilstone, G., and Halsey, K. H.: An absorption-based approach to improved estimates of phytoplankton biomass and net primary production, Limnology and Oceanography Letters, 7, 419–426, 2022.

Geider, R., MacIntyre, H., and Kana, T.: Dynamic model of phytoplankton growth and acclimation: responses of the balanced growth rate and the chlorophyll a: carbon ratio to light, nutrient-limitation and temperature, Marine Ecology Progress Series, 148, 187–200, 1997.

Guo, Y., Hastie, T., and Tibshirani, R.: Regularized linear discriminant analysis and its application in microarrays, Biostatistics, 8, 86–100, 2007.

Haëntjens, N., Boss, E. S., Graff, J. R., Chase, A. P., and Karp-Boss, L.: Phytoplankton size distributions in the western North Atlantic and their seasonal variability, Limnology and Oceanography, 67, 1865–1878, 2022.

Henson, S. A., Cael, B., Allen, S. R., and Dutkiewicz, S.: Future phytoplankton diversity in a changing climate, Nature communications, 12, 5372, 2021.

Hirata, T., Hardman-Mountford, N., Brewin, R., Aiken, J., Barlow, R., Suzuki, K., Isada, T., Howell, E., Hashioka, T., Noguchi-Aita, M., et al.: Synoptic relationships between surface Chlorophyll-a and diagnostic pigments specific to phytoplankton functional types, Biogeosciences, 8, 311–327, 2011.

Hoepffner, N. and Sathyendranath, S.: Effect of pigment composition on absorption properties of phytoplankton, Mar. Ecol. Prog. Ser, 73, 11–23, 1991.

Isada, T., Hirawake, T., Kobayashi, T., Nosaka, Y., Natsuike, M., Imai, I., Suzuki, K., and Saitoh, S.-I.: Hyperspectral optical discrimination of phytoplankton community structure in Funka Bay and its implications for ocean color remote sensing of diatoms, Remote Sensing of Environment, 159, 134–151, 2015.

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

Keller, M. D.: Dimethyl sulfide production and marine phytoplankton: the importance of species composition and cell size, Biological oceanography, 6, 375–382, 1989.

Kostadinov, T., Siegel, D., and Maritorena, S.: Retrieval of the particle size distribution from satellite ocean color observations, Journal of Geophysical Research: Oceans, 114, 2009.

Kramer, S. J. and Siegel, D. A.: How can phytoplankton pigments be best used to characterize surface ocean phytoplankton groups for ocean color remote sensing algorithms?, Journal of Geophysical Research: Oceans, 124, 7557–7574, 2019.

Kramer, S. J., Siegel, D. A., and Graff, J. R.: Phytoplankton community composition determined from co-variability among phytoplankton pigments from the NAAMES field campaign, Frontiers in Marine Science, 7, 215, 2020.

Krause, J. W., Schulz, I. K., Rowe, K. A., Dobbins, W., Winding, M. H., Sejr, M. K., Duarte, C. M., and Agustí, S.: Silicic acid limitation drives bloom termination and potential carbon sequestration in an Arctic bloom, Scientific Reports, 9, 8149, 2019.

Liu, Y., Boss, E., Chase, A., Xi, H., Zhang, X., Röttgers, R., Pan, Y., and Bracher, A.: Retrieval of phytoplankton pigments from underway spectrophotometry in the Fram Strait, Remote Sensing, 11, 318, 2019.

Mackey, M., Mackey, D., Higgins, H., and Wright, S.: CHEMTAX-a program for estimating class abundances from chemical markers: application to HPLC measurements of phytoplankton, Marine Ecology Progress Series, 144, 265–283, 1996.

Mao, Z., Stuart, V., Pan, D., Chen, J., Gong, F., Huang, H., and Zhu, Q.: Effects of phytoplankton species composition on absorption spectra and modeled hyperspectral reflectance, Ecological Informatics, 5, 359–366, 2010.

McLachlan, G. and Peel, D.: Finite Mixture Models, John Wiley, New York. doi, 10, 0471721 182, 2000.

McNabb, B. J. and Tortell, P. D.: Improved prediction of dimethyl sulfide (DMS) distributions in the northeast subarctic Pacific using machine-learning algorithms, Biogeosciences, 19, 1705–1721, 2022.

McParland, E. L. and Levine, N. M.: The role of differential DMSP production and community composition in predicting variability of global surface DMSP concentrations, Limnology and Oceanography, 64, 757–773, 2019.

Morel, A. and Bricaud, A.: Theoretical results concerning light absorption in a discrete medium, and application to specific absorption of phytoplankton, Deep Sea Research Part A. Oceanographic Research Papers, 28, 1375–1393, 1981.

Mouw, C. B., Hardman-Mountford, N. J., Alvain, S., Bracher, A., Brewin, R. J., Bricaud, A., Ciotti, A. M., Devred, E., Fujiwara, A., Hirata, T., et al.: A consumer's guide to satellite remote sensing of multiple phytoplankton groups in the global ocean, Frontiers in Marine Science, 4, 41, 2017.

Orkney, A., Platt, T., Narayanaswamy, B. E., Kostakis, I., and Bouman, H. A.: Bio-optical evidence for increasing Phaeocystis dominance in the Barents Sea, Philosophical Transactions of the Royal Society A, 378, 20190 357, 2020.

Peña, M. A. and Varela, D. E.: Seasonal and interannual variability in phytoplankton and nutrient dynamics along Line P in the NE subarctic Pacific, Progress in Oceanography, 75, 200–222, 2007.

Peña, M. A., Fine, I., and Callendar, W.: Interannual variability in primary production and shelf-offshore transport of nutrients along the northeast Pacific Ocean margin, Deep Sea Research Part II: Topical Studies in Oceanography, 169, 104 637, 2019a.

Peña, M. A., Nemcek, N., and Robert, M.: Phytoplankton responses to the 2014–2016 warming anomaly in the northeast subarctic Pacific Ocean, Limnology and Oceanography, 64, 515–525, 2019b.

Queré, C. L., Harrison, S. P., Colin Prentice, I., Buitenhuis, E. T., Aumont, O., Bopp, L., Claustre, H., Cotrim Da Cunha, L., Geider, R., Giraud, X., et al.: Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models, Global Change Biology, 11, 2016–2040, 2005.

https://doi.org/10.5194/egusphere-2023-2851
Preprint. Discussion started: 8 December 2023

Robinson, C. M., Huot, Y., Schuback, N., Ryan-Keogh, T. J., Thomalla, S. J., and Antoine, D.: High latitude Southern Ocean phytoplankton have distinctive bio-optical properties, Optics Express, 29, 21 084–21 112, 2021.

Roy, S., Llewellyn, C. A., Egeland, E. S., and Johnsen, G.: Phytoplankton pigments: characterization, chemotaxonomy and applications in oceanography, Cambridge University Press, 2011.

Slade, W. H., Boss, E., Dall'Olmo, G., Langner, M. R., Loftin, J., Behrenfeld, M. J., Roesler, C., and Westberry, T. K.: Underway and moored methods for improving accuracy in measurement of spectral particulate absorption and attenuation, Journal of Atmospheric and Oceanic Technology, 27, 1733–1746, 2010.

Smetacek, V.: Role of sinking in diatom life-history cycles: ecological, evolutionary and geological significance, Marine biology, 84, 239–251, 1985.

Uitz, J., Claustre, H., Morel, A., and Hooker, S. B.: Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll, Journal of Geophysical Research: Oceans, 111, 2006.

van Oijen, T., van Leeuwe, M. A., Gieskes, W. W., and de Baar, H. J.: Effects of iron limitation on photosynthesis and carbohydrate metabolism in the Antarctic diatom Chaetoceros brevis (Bacillariophyceae), European Journal of Phycology, 39, 161–171, 2004.

Werdell, P. J., Behrenfeld, M. J., Bontempi, P. S., Boss, E., Cairns, B., Davis, G. T., Franz, B. A., Gliese, U. B., Gorman, E. T., Hasekamp, O., et al.: The Plankton, Aerosol, Cloud, ocean Ecosystem mission: status, science, advances, Bulletin of the American Meteorological Society, 100, 1775–1794, 2019.

Whitney, F., Crawford, W., and Harrison, P.: Physical processes that enhance nutrient transport and primary productivity in the coastal and open ocean of the subarctic NE Pacific, Deep Sea Research Part II: Topical Studies in Oceanography, 52, 681–706, 2005.

Yoch, D. C.: Dimethylsulfoniopropionate: its sources, role in the marine food web, and biological degradation to dimethylsulfide, Applied and environmental microbiology, 68, 5804–5815, 2002.

Zapata, M., Rodríguez, F., and Garrido, J. L.: Separation of chlorophylls and carotenoids from marine phytoplankton: a new HPLC method using a reversed phase C8 column and pyridine-containing mobile phases, Marine Ecology Progress Series, 195, 29–45, 2000.

Zeng, C., Rosengard, S. Z., Burt, W., Peña, M. A., Nemcek, N., Zeng, T., Arrigo, K. R., and Tortell, P. D.: Optically-derived estimates of phytoplankton size class and taxonomic group biomass in the Eastern Subarctic Pacific Ocean, Deep Sea Research Part I: Oceanographic Research Papers, 136, 107–118, 2018.

Zhang, X., Huot, Y., Gray, D., Weidemann, A., and Rhea, W.: Biogeochemical origins of particles obtained from the inversion of the volume scattering function and spectral absorption in coastal waters, Biogeosciences, 10, 6029–6043, 2013.

Zhang, X., Huot, Y., Bricaud, A., and Sosik, H. M.: Inversion of spectral absorption coefficients to infer phytoplankton size classes, chlorophyll concentration, and detrital matter, Applied Optics, 54, 5805–5816, 2015.