Retrieval of phytoplankton pigments and functional types from underway spectrophotometry in the Fram Strait

Yangyang Liu[1,2,*], Emmanuel Boss[4], Alison Chase[4], Yanqun Pan[6], Hongyan Xi[1], Rüdiger Röttgers[5], and Astrid Bracher[1,3]

[1]*Alfred Wegener Institute Helmholtz Center for Polar and Marine Research, Bremerhaven, Germany*
[2]*Faculty of Biology and Chemistry, University of Bremen, Bremen, Germany*
[3]*Institute of Environmental Physics, University of Bremen, Germany*
[4]*School of Marine Sciences, University of Maine, Orono, ME, USA*
[5]*Helmholtz-Zentrum Geesthacht Center for Materials and Coastal Research, Geesthacht, Germany*
[6]*State Key Laboratory of Estuarine and Coastal Research, East China Normal University, Shanghai, China*
[*]*Corresponding author: Yangyang Liu, yangyang.liu@awi.de*

Abstract

Gaussian decomposition and Singular Value Decomposition combined with Non-Negative Least Squares (SVD-NNLS) method are compared and combined to estimate the concentration of 18 phytoplankton pigments (names and abbreviations are displayed in Table 2) from phytoplankton absorption spectra. Results show that both methods tend to overestimate pigment concentration. Gaussian decomposition method provides robust estimation of TChl-a, TChl-b, Chl-c1/2, PSC and PPC. The estimates of TChl-a, Fuco, Diato, β-Caro, Prasino, TChl-b, Zea, Viola and Lut from SVD-NNLS show reasonable estimation accuracy, while the other pigments are subjected to relatively large prediction errors. The estimated pigments concentrations are further exploited based on Diagnostic Pigment Analysis to derive four phytoplankton functional types, i.e. diatoms, prymnesiophytes, green algae and prokaryotes. By the application of these two methods to the particulate absorption spectra collected by underway spectrophotometry during three summer cruises in 2015 - 2017 in the Fram Strait, continuous surface phytoplankton functional types are estimated along the cruise course.

1 Introduction

Knowledge on phytoplankton pigments (some are abbreviated as Table 2) is critical to understanding the impacts of the changing environment on primary productivity (Uitz et al., 2009), phytoplankton diversity and taxonomic composition. In remote sensing applications, the pigment databases have been extensively used in developing, validating or refining bio-optical algorithms for estimating phytoplankton biomass and functional types (Bracher et al., 2017, and references therein). These databases are mainly based on high-performance liquid chromatography (HPLC) analysis of discrete water samples. This technique is limited by repeat frequency and spatial coverage. While HPLC pigment analysis remains indispensable, it is invaluable to explore methods that enable easier availability of high resolution pigment data.

Attempts have been made to quantify the concentration of various phytoplankton pigments from optical measurements (e.g. absorption or reflectance spectra) since they currently remain the only means of collecting synoptic scale information. Among these methods, the decomposition of spectra into Gaussian functions (Hoepffner & Sathyendranath, 1993; Lohrenz et al., 2003; Chase et al., 2013) and spectral reconstruction (e.g. Moisan et al., 2011) are commonly used. The shipboard underway spectrophotometry greatly facilitates the acquisition of

particulate absorption spectra ($a_p(\lambda)$) with unprecedented temporal and spatial resolution (e.g. Dall'Olmo et al., 2009, 2011, 2012; Slade et al., 2010; Westberry et al., 2010; Boss et al., 2013; Chase et al., 2013; Werdell et al., 2013; Brewin et al., 2016; Liu et al., 2018). It has been proved to successfully provide continuous surface TChl-a data along cruise tracks (Westberry et al., 2010; Slade et al., 2010; Boss et al., 2013; Werdell et al., 2013; Brewin et al., 2016; Liu et al., 2018). Furthermore, Gaussian decomposition has been performed by Chase et al. (2013) to retrieve major pigment groups from a large global underway AC-S derived $a_p(\lambda)$ data set. A widespread investigations of phytoplankton abundance and distribution can be expected if the pigment data derived from this technique is further exploited.

The Fram Strait is the only deep connection between the North Atlantic and Arctic Oceans (Figure 1(a)). The studies of phytoplankton community composition in this region are mainly based on discrete water sample analysis or moored sediment traps. *In situ* optical measurements are insufficient. In particular, no local algorithms for phytoplankton composition nor pigment inversion from light absorption measurements have been developed.

In this study, we investigate the performances of two approaches, namely Gaussian decomposition and the singular value decomposition combined with non-negative least squares (SVD-NNLS), in determining the concentrations of either individual pigments or pigment groups (e.g. PPC or PSC) from $a_p(\lambda)$ obtained from underway spectrophotometry during three Fram Strait cruises. Their applicability to the Fram Strait and its vicinity was assessed. Finally, these continuous underway pigment data were analysed to obtain the contribution of four phytoplankton functional types, i.e. diatoms, prymnesiophytes, green algae and prokaryotes, to total biomass.

2 Data and Methods

2.1 Data Collection

Data were collected during three expeditions on R/V Polarstern: PS93.2 (July-August 2015), PS99.2 (June-July 2016) and PS107 (July-August 2017). Sampling sites were located in the Fram Strait and its vicinity, ranging from approximately latitudes 72° to 80°N and longitudes 10°W to 15°E (Figure 1(a)).

Continuous $a_p(\lambda)$ and discrete pigment concentration measurements of the surface water were performed for each expedition. Sampling methods and data analysis have been described in detail by Liu et al. (2018). Table 2 depicts the names and abbreviations of each pigment as well as the pigment groups used in this study. In addition, absorption coefficient of non-algal particles ($a_{NAP}(\lambda)$) in discrete water samples was measured for the determination of the spectral slope that will be used in subsequent data analysis. $a_{NAP}(\lambda)$ has been recognized to vary approximately in an exponentially decaying function (Bricaud et al., 1998; Babin et al., 2003):

$$a_{NAP}(\lambda) = a_{NAP}(400)\, e^{-S(\lambda - 400)} \tag{1}$$

where S is the spectral slope of the $a_{NAP}(\lambda)$ spectrum. Equation 1 was fitted to $a_{NAP}(\lambda)$ for data between 380-700 nm excluding the 400-480 nm and 620-700 nm ranges using non-linear least squares method (Babin et al., 2003). The median value of S for all three expeditions is 0.016, which is used in the decomposition of AC-S derived $a_p(\lambda)$ to obtain $a_{ph}(\lambda)$ (detailed in 2.2.2).

2.2 Retrieval of Phytoplankton Pigments

2.2.1 Statistics

AC-S derived $a_p(\lambda)$ averaged within the period of 10 minutes before and after HPLC sampling time were collocated with HPLC pigments data. $a_{ph}(\lambda)$ was then obtained by numerical decomposition. In total, 299 match-ups were obtained, which were subsequently used as the pigment retrieval data set.

For the development of pigment retrieval models, all the match-up points were used as training data. Statistics for applying the model on the training data include the slope (S) and the intercept (I) of Model-1 Bisquare robust linear regression, the determination coefficient (R^2), the root mean square error (RMSE), the mean absolute error (MAE) and Relative Percentage Difference (RPD) (Equations see Liu et al. (2018)). For model evaluation, ten-fold cross-validation was performed (MATLAB function *crossvalind*) to estimate likely performance of each model on out-of-sample data, according to Kohavi et al. (1995). The pigment inversion data set was split into 10 equal partitions (or "folds"), with one fold being the testing set and the union of the other folds being the training set. Statistics were iteratively calculated for 10 times, using a different fold as the testing set each time. The model prediction error for out-of-sample data was defined as the average of the 10 sets of statistics.

2.2.2 Gaussian Decomposition

Following Chase et al. (2013), AC-S $a_p(\lambda)$ was decomposed into 12 Gaussian functions and one $a_{NAP}(\lambda)$ exponential function expressed by Equation 1 in the range of 400-700 nm. Each Gaussian function ($a_{gaus}(\lambda)$) represents the absorption by a certain phytoplankton pigment or pigment group. The absorption by the water-soluble photosynthetic pigment phycoerythrin was also represented as a Gaussian function. The peak location and width of each Gaussian function shown in Table 2 were defined with fixed values based on known pigment absorption shapes (Bricaud et al., 2004).

The amplitude of $a_{NAP}(400)$ was derived and used to reconstruct $a_{NAP}(\lambda)$ according to Equation 1. $a_{ph}(\lambda)$ was obtained by differencing $a_p(\lambda)$ and $a_{NAP}(\lambda)$. To evaluate the goodness of the decomposition, $a_p(\lambda)$ was reconstructed by the linear combination of derived $a_{gaus}(\lambda)$ for all the pigments and $a_{NAP}(\lambda)$ and then compared to that before decomposition. The relative median error of all the match-up points at each band is calculated as:

$$e_{median}(\lambda) = median\ of\ \frac{|a_p(\lambda)_{AC-S} - a_p(\lambda)_{reconstruct}|}{a_p(\lambda)_{AC-S}} \times 100 \tag{2}$$

The amplitude of each Gaussian function $a_{gaus}(\lambda_0)$ was compared to HPLC pigment concentrations by fitting the power function using Bisqaure robust non-linear least squares method.

2.2.3 Singular Value Decomposition - Non-Negative Least Squares (SVD-NNLS)

Phytoplankton absorption spectra $a_{ph}(\lambda)$ can be reconstructed as the linear combination of the absorption spectra of individual pigments that equals to the pigment-specific absorption coefficient multiplied by pigment concentration (Bidigare et al., 1987). When there is more than one samples, this can be written in matrix form as:

$$\begin{bmatrix} c_{i=1,j=1} & \cdots & c_{i=m,j=1} \\ \vdots & \ddots & \vdots \\ c_{i=1,j=n} & \cdots & c_{i=m,j=n} \end{bmatrix} \begin{bmatrix} \widetilde{a}^*_{i=1}(\lambda) \\ \vdots \\ \widetilde{a}^*_{i=m}(\lambda) \end{bmatrix} = \begin{bmatrix} a_{ph,j=1}(\lambda) \\ \vdots \\ a_{ph,j=n}(\lambda) \end{bmatrix} \Longleftrightarrow C \cdot \widetilde{A} = A_{ph} \tag{3}$$

where c is the observed pigment concentration, $\widetilde{a}^*(\lambda)$ is the derived pigment-specific absorption coefficient, n is the number of samples, i is the sample index, m is the number of pigments measured in each sample and j is the pigment index. At this point, the observed collocated pigment concentration (matrix C) and $a_{ph}(\lambda)$ (matrix A_{ph}) are involved in a set of linear algebraic equations. The solution of these linear equations is the unknown matrix $\widetilde{A}$ with the elements being the inverted pigment-specific absorption coefficient $\widetilde{a}^*_{i,i=1:m}(\lambda)$. The derived $\widetilde{A}$ was then used combined with A_{ph} to derive matrix $\widetilde{C}$ in Equation 11 - the estimated pigmtent concentration.

$$\begin{bmatrix} \widetilde{a}^*_{i=1}(\lambda_1) & \cdots & \widetilde{a}^*_{i=m}(\lambda_1) \\ \vdots & \ddots & \vdots \\ \widetilde{a}^*_{i=1}(\lambda_l) & \cdots & \widetilde{a}^*_{i=m}(\lambda_l) \end{bmatrix} \begin{bmatrix} \widetilde{c}_{i=1,j=1...n} \\ \vdots \\ \widetilde{c}_{i=m,j=1...n} \end{bmatrix} = \begin{bmatrix} a_{ph,j=1...n}(\lambda_1) \\ \vdots \\ a_{ph,j=1...n}(\lambda_l) \end{bmatrix} \Longleftrightarrow \widetilde{A} \cdot \widetilde{C} = A_{ph} \quad (4)$$

where $\widetilde{a}^*(\lambda)$ is the derived pigment-specific absorption coefficient by solving Equation 10, $\widetilde{c}$ is the estimated pigment concentration, n is the number of samples, i is the sample index, m is the number of pigments measured in each sample, j is the pigment index and l is the number of wavelengths.

Based on the above theory, Moisan et al. (2011) developed the SVD-NNLS approach to estimate phytoplankton pigment concentration, i.e. solving Equation 3 with SVD and Equation 4 with NNLS. Here, their SVD-NNLS approach was adapted and tested using our data set. $a_{ph}(\lambda)$ was normalized by the pigment package effect index at 675 nm (denoted as $\hat{a}_{ph}(\lambda)$) (Moisan et al., 2011).

3 Results

3.1 Characteristics of the absorption-pigment data set

Figure 1(b) illustrates the $a_p(\lambda)$ spectra derived from underway AC-S measurements collocated to HPLC data. These spectra show two absorption maxima in blue ($\sim$ 440 nm) and red ($\sim$ 675 nm) spectrum domains. Additionally, most $a_p(\lambda)$ spectra also exhibit a shoulder peak in the region of 460-500 nm. The blue spectrum domain is absorbed by chlorophylls and carotenoids and the red domain is primarily absorbed by chlorophyll-a and to a lesser extent by chlorophyll-b and chlorophyll-c. The magnitudes of $a_p(440)$ and $a_p(675)$ vary in the range of 0.007-0.258 m^{-1} (median: 0.043 m^{-1}) and 0.0008-0.086 m^{-1} (median: 0.013 m^{-1}), respectively. TChl-a-specific particulate absorption coefficient ($a_p^*(\lambda)$) at 440 nm ($a_p^*(440)$) varies from 0.031 to 0.289 $m^2\ mg^{-1}$ (median: 0.059 $m^2\ mg^{-1}$), whilst $a_p^*(675)$ varies from 0.008 to 0.105 $m^2\ mg^{-1}$ (median: 0.018 $m^2\ mg^{-1}$).

TChl-a concentration determined by HPLC at the discrete water samples spans the range 0.065-3.868 $mg\ m^{-3}$, with 9.4%, 84.3% and 6.3% of the data <0.2 $m^2\ mg^{-1}$ (oligotrophic), between 0.2 and 2 $m^2\ mg^{-1}$ (mesotrophic) and >2 $m^2\ mg^{-1}$ (eutrophic), respectively (trophic state level referred to Bricaud et al. (2004)). Only TChl-a, Fuco, Hex and PSC have a mean concentration greater than 0.2 $mg\ m^{-3}$. Other than that, the pigments with mean concentration greater than 0.05 $mg\ m^{-3}$ are Chl-c1/2, Diadino, TChl-b, and PPC. Large standard deviation were observed within individual pigments, and the ratio of standard deviation to mean value are in the range of 0.58 to 2.04.

3.2 Gaussian Decomposition

As shown in Figure 2(a), the $e_{median}(\lambda)$ of the reconstructed $a_p(\lambda)$ is small. It is less than 10% for the range 400-690 nm and reaches the largest value $\sim$21% at $\sim$698 nm, indicating a good

accuracy of the decomposition. Significant correlations were found between the Gaussian function amplitudes and the corresponding pigment concentration ($R^2 > 0.49$) (Table 1). Among all the Gaussian functions representing TChl-a, the amplitude at 435 nm provides the best power relationship with TChl-a concentration ($R^2 = 0.86, RMSE = 0.310$), closely followed by the one at 675 nm. $a_{gaus}(638)$ is much better correlated with Chl-c1/2 than $a_{gaus}(584)$, while $a_{gaus}(660)$ provides slightly better correlation with TChl-b than $a_{gaus}(470)$. PSC is well correlated with $a_{gaus}(523)$, whereas $a_{gaus}(492)$ and PPC has the poorest relationship. Our data are comparable to the results from the Tara expedition (Chase et al., 2013) except for TChl-b. This is probably because of a small TChl-b range of values in this study.

Subsequently, $a_{gaus}(435)$, $a_{gaus}(660)$, $a_{gaus}(638)$, $a_{gaus}(523)$ and $a_{gaus}(492)$ were used for predicting the concentration of TChl-a, TChl-b, Chl-c1/2, PSC and PPC. The measured and estimated TChl-a, TChl-b and Chl-c1/2 using all the match-up data (training set) show consistency with each other (Figure 3). Statistics based on ten-fold cross validation is shown in Table 2. Overall, all five pigments were reasonably retrieved, but overestimated (indicated by positive RPD). TChl-a has the least prediction error and was overestimated by 12.2%. TChl-b has the second least prediction error, but with a low R^2 possibly because of a narrower data range than the other four pigments (Table 1). Similarly, PPC was in a relatively small data range but better estimated than PSC and Chl-c1/2 in terms of MAE and RPD.

3.3 SVD-NNLS

The pigment-specific absorption coefficients ($\widetilde{a}_i^*(\lambda)$) for the 18 pigments were obtained by solving Equation 3 using SVD. Each $\widetilde{a}_i^*(\lambda)$ spectra varies smoothly across the full bands. Negative coefficients are possible because they are solved by pure matrix inversion, which is different from the real *in vivo* $a_i^*(\lambda)$. The calculated $\widetilde{a}_i^*(\lambda)$ for various pigments were used to solve Equation 4 to estimate the corresponding pigment concentration ($\widetilde{c}_i$). $\hat{a}_{ph}(\lambda)$ were reconstructed by use of $\widetilde{a}_i^*(\lambda)$ and $\widetilde{c}_i$ according to Equation 4. The $e_{median}(\lambda)$ of the reconstruction for all the match-up points was calculated according to Equation 2. As displayed in Figure 2(b), similar to Gaussian decomposition method (Figure 2(a)), the $e_{median}(\lambda)$ is less than 6% for the range 400-690 nm and reaches the largest value ~16% at ~698 nm, indicating a good accuracy of the decomposition and that the negative $\widetilde{a}_i^*(\lambda)$ did not render the spectral reconstruction.

For comparison, the estimated TChl-a, TChl-b and Chl-c1/2 from training set were plotted again the HPLC measured ones (Figure 3). The overlapping of the data points in Figure 3(a-c) reveals similar training accuracy obtained by SVD-NNLS as by Gaussian decomposition. The statistics for pigment retrieval are shown in Table 2. Except for Allo, α-Caro, But, Viola, Lut, Zea and TChl-b, the estimates of all the other pigments show strong correlation with the HPLC measured values ($R^2 > 0.25$). Overall, all the pigments were overestimated, as implied by a positive RPD. TChl-a has the smallest prediction error (RMSE=0.316, MAE=0.215) and was overestimated by 39.6%. Furthermore, Fuco, Diato, β-Caro and Prasino have both relatively lower estimation errors and strong measure-estimation correlations. Additionally, TChl-b, Zea, Viola and Lut show reasonable estimation accuracy. In contrast, Chl-c3, α-Caro, Hex and But were poorly estimated, with larger RMSE and MAE than other pigments and RPD greater than 200%.

3.4 Phytoplankton functional types retrievals

Based on the statistics in Table 2, the concentration of various pigments were estimated from the underway AC-S $a_p(\lambda)$ data where there were no collocated HPLC data. TChl-b was estimated using Gaussian decomposition, due to its outperformance over SVD-NNLS for TChl-b, whereas Fuco, Peri, Hex, But, Allo and Zea were estimated using SVD-NNLS. These pigment estimates

were subsequently analysed with Diagnostic Pigment Analysis according to Losa et al. (2017) for the derivation of the contribution of diatoms, prymnesiophytes, green algae and prokaryotes to total biomass. Figure 4 shows an example of the comparison of the time series of the four phytoplankton functional types derived from AC-S and HPLC. The functional types estimated from AC-S data exhibit a general agreement with those from HPLC, however, yet much more highly resolved during the cruise periods.

4 Discussion

In this study, both methods invert pigments by the use of the relationships between phytoplankton absorption and pigments concentration. These relationships are mainly modulated by pigment package effect which is not so far accounted for. It is a function of phytoplankton cell size and pigment composition. For the Gaussian decomposition method, the absorption of individual pigments or pigment groups are separated and related to corresponding pigment concentration. Therefore, the data scatter around the regression line between $a_{gaus}(\lambda_0)$ and pigment concentration (Table 1) is not affected by the differences in pigment composition but mainly by cell size. In addition, to some extent the covarying absorption by more than one pigment that failed to be separated by this method can also enhance this scatter. For SVD-NNLS, Moisan et al. (2011) have found that after correcting $a_{ph}(\lambda)$ with the package effect factor at 675 nm, the reconstruction of $a_{ph}(\lambda)$ using inverted pigment-specific absorption and inverted pigment concentration shows a dramatic improvement as compared to the use of inversions done using the original measured absorption spectra. Note that this correction is a simplification of package effect correction by assuming that the package effect factor is wavelength-independent and equals to that at 675 nm. Therefore, better inversion accuracy can be expected by further deciphering package effect and quantifying the influence of cell size to the absorption - concentration relationship.

5 Conclusions

In this study, we tested the applicability of two approaches, i.e. Gaussian decomposition and SVD-NNLS for retrieving phytoplankton pigments from $a_p(\lambda)$ collected by underway spectrophotometry in the Fram Strait, European Arctic Ocean. Although both methods tend to overestimate specific pigment concentration, the combination of the methods still provide robust estimates for many of the tested 18 phytoplankton pigments and two pigment groups (PSC and PPC). The Gaussian decomposition method provides robust estimation of TChl-a, TChl-b, Chl-c1/2, PSC and PPC. The SVD-NNLS method allows a reasonably good estimation of TChl-a, Fuco, Diato, β-Caro, Prasino, TChl-b, Zea, Viola and Lut. Other pigments are subjected to relatively large prediction errors. Based on Diagnostic Pigment Analysis, diatoms, prymnesiophytes, green algae and prokaryotes were determined by combining the results from the two inversion methods. This enables not only the obtaining of high resolution phytoplankton pigment data but also a generation of a continuous high resolution data set on TChl-a concentration of key phytoplankton groups in this region. By the application of these two methods to underway spectrophotometry data sampled during three summer cruises in 2015 - 2017 in the Fram Strait, continuous surface phytoplankton functional types were estimated along the cruise course.

Acknowledgement

References

Babin, M., Stramski, D., Ferrari, G. M., Claustre, H., Bricaud, A., Obolensky, G., & Hoepffner, N. (2003). Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around europe. *Journal of Geophysical Research: Oceans*, *108*(C7).

Bidigare, R. R., Smith, R., Baker, K., & Marra, J. (1987). Oceanic primary production estimates from measurements of spectral irradiance and pigment concentrations. *Global Biogeochemical Cycles*, *1*(3), 171–186.

Boss, E., Picheral, M., Leeuw, T., Chase, A., Karsenti, E., Gorsky, G., ... Claustre, H. (2013). 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.

Bracher, A., Bouman, H. A., Brewin, R. J., Bricaud, A., Brotas, V., Ciotti, A. M., ... others (2017). Obtaining phytoplankton diversity from ocean color: a scientific roadmap for future development. *Frontiers in Marine Science*, *4*, 55.

Brewin, R. J., Dall'Olmo, G., Pardo, S., van Dongen-Vogels, V., & Boss, E. S. (2016). Underway spectrophotometry along the atlantic meridional transect reveals high performance in satellite chlorophyll retrievals. *Remote sensing of environment*, *183*, 82–97.

Bricaud, A., Claustre, H., Ras, J., & Oubelkheir, K. (2004). Natural variability of phytoplanktonic absorption in oceanic waters: Influence of the size structure of algal populations. *Journal of Geophysical Research: Oceans*, *109*(C11).

Bricaud, A., Morel, A., Babin, M., Allali, K., & Claustre, H. (1998). Variations of light absorption by suspended particles with chlorophyll a concentration in oceanic (case 1) waters: Analysis and implications for bio-optical models. *Journal of Geophysical Research: Oceans*, *103*(C13), 31033–31044.

Chase, A., Boss, E., Zaneveld, R., Bricaud, A., Claustre, H., Ras, J., ... Westberry, T. K. (2013). Decomposition of in situ particulate absorption spectra. *Methods in Oceanography*, *7*, 110–124.

Dall'Olmo, G., Boss, E., Behrenfeld, M., & Westberry, T. (2012). Particulate optical scattering coefficients along an atlantic meridional transect. *Optics express*, *20*(19), 21532–21551.

Dall'Olmo, G., Boss, E., Behrenfeld, M. J., Westberry, T., Courties, C., Prieur, L., ... Moutin, T. (2011). Inferring phytoplankton carbon and eco-physiological rates from diel cycles of spectral particulate beam-attenuation coefficient. *Biogeosciences*, *8*(11), 3423–3439.

Dall'Olmo, G., Westberry, T., Behrenfeld, M., Boss, E., & Slade, W. (2009). Significant contribution of large particles to optical backscattering in the open ocean. *Biogeosciences*, *6*(6), 947.

Hoepffner, N., & Sathyendranath, S. (1993). Determination of the major groups of phytoplankton pigments from the absorption spectra of total particulate matter. *Journal of Geophysical Research: Oceans*, *98*(C12), 22789–22803.

Kohavi, R., et al. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai* (Vol. 14, pp. 1137–1145).

Liu, Y., Röttgers, R., Ramírez-Pérez, M., Dinter, T., Steinmetz, F., Nöthig, E.-M., ... Bracher, A. (2018). Underway spectrophotometry in the fram strait (european arctic ocean): a highly resolved chlorophyll a data source for complementing satellite ocean color. *Optics Express*, *26*(14), A678–A696.

Lohrenz, S. E., Weidemann, A. D., & Tuel, M. (2003). Phytoplankton spectral absorption as influenced by community size structure and pigment composition. *Journal of Plankton Research*, *25*(1), 35–61.

Losa, S. N., Soppa, M. A., Dinter, T., Wolanin, A., Brewin, R. J., Bricaud, A., ... others (2017). Synergistic exploitation of hyper-and multi-spectral precursor sentinel measurements to determine phytoplankton functional types (synsenpft). *Frontiers in Marine Science*, *4*, 203.

Moisan, J. R., Moisan, T. A., & Linkswiler, M. A. (2011). An inverse modeling approach to estimating phytoplankton pigment concentrations from phytoplankton absorption spectra. *Journal of Geophysical Research: Oceans*, *116*(C9).

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

Slade, W. H., Boss, E., Dall'Olmo, G., Langner, M. R., Loftin, J., Behrenfeld, M. J., ... Westberry, T. K. (2010). Underway and moored methods for improving accuracy in measurement of spectral particulate absorption and attenuation. *Journal of Atmospheric and Oceanic Technology*, *27*(10), 1733–1746.

Uitz, J., Claustre, H., Griffiths, F. B., Ras, J., Garcia, N., & Sandroni, V. (2009). A phytoplankton class-specific primary production model applied to the kerguelen islands region (southern ocean). *Deep Sea Research Part I: Oceanographic Research Papers*, *56*(4), 541–560.

Werdell, P. J., Proctor, C. W., Boss, E., Leeuw, T., & Ouhssain, M. (2013). Underway sampling of marine inherent optical properties on the tara oceans expedition as a novel resource for ocean color satellite data product validation. *Methods in Oceanography*, *7*, 40–51.

Westberry, T. K., Dall'Olmo, G., Boss, E., Behrenfeld, M. J., & Moutin, T. (2010). Coherence of particulate beam attenuation and backscattering coefficients in diverse open ocean environments. *Optics Express*, *18*(15), 15419–15425.
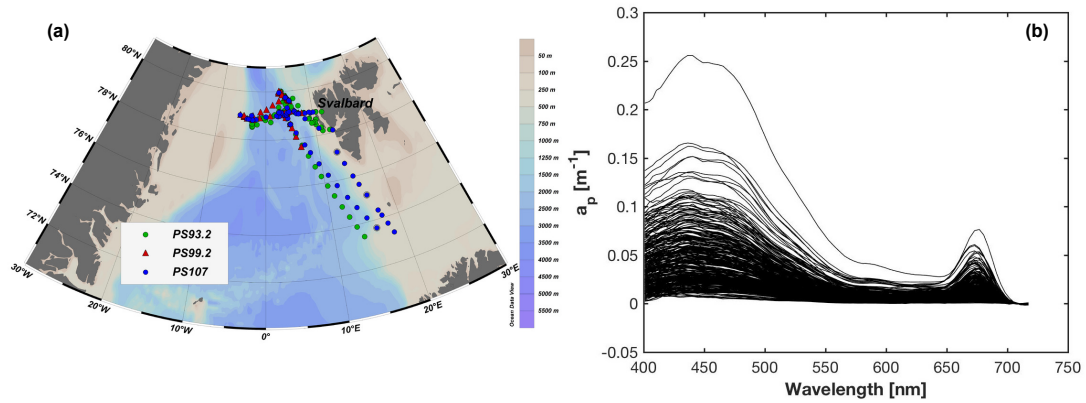
Figure 1: (a) Expedition tracks for PS93.2 (July-August 2015), PS99.2 (June-July 2016) and PS107 (July-August 2017) for collocated AC-S and HPLC measurements; (b) AC-S derived $a_p(\lambda)$ spectra collocated with HPLC measurements.
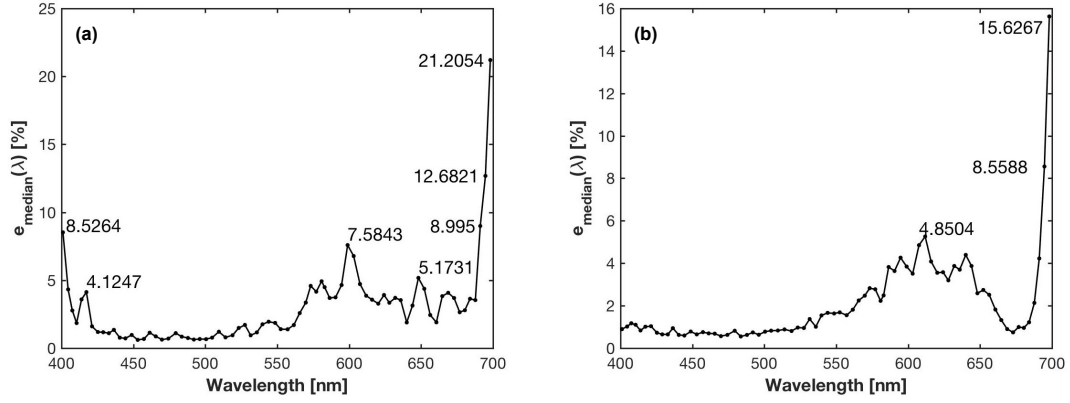
Figure 2: The relative median error $e_{median(\lambda)}$ of the reconstructed $a_p(\lambda)$ from (a) Gaussian decomposition method and (b) SVD-NNLS method.
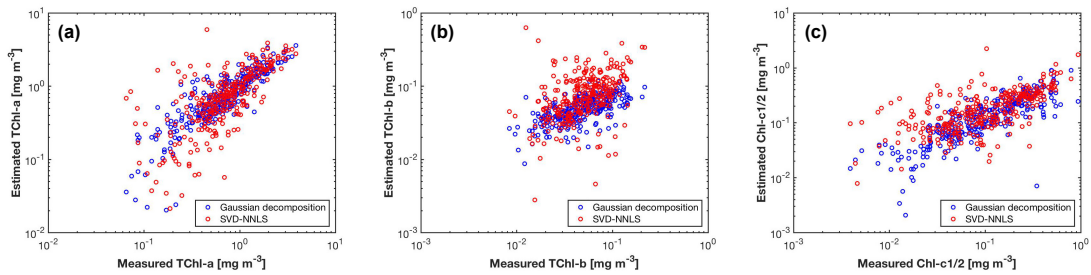

Figure 3: Comparison of HPLC measured and Gaussian decomposition or SVD-NNLS estimated concentration of TChl-a (a), TChl-b (b) and Chl-c1/2 (c) for all the match-up points.
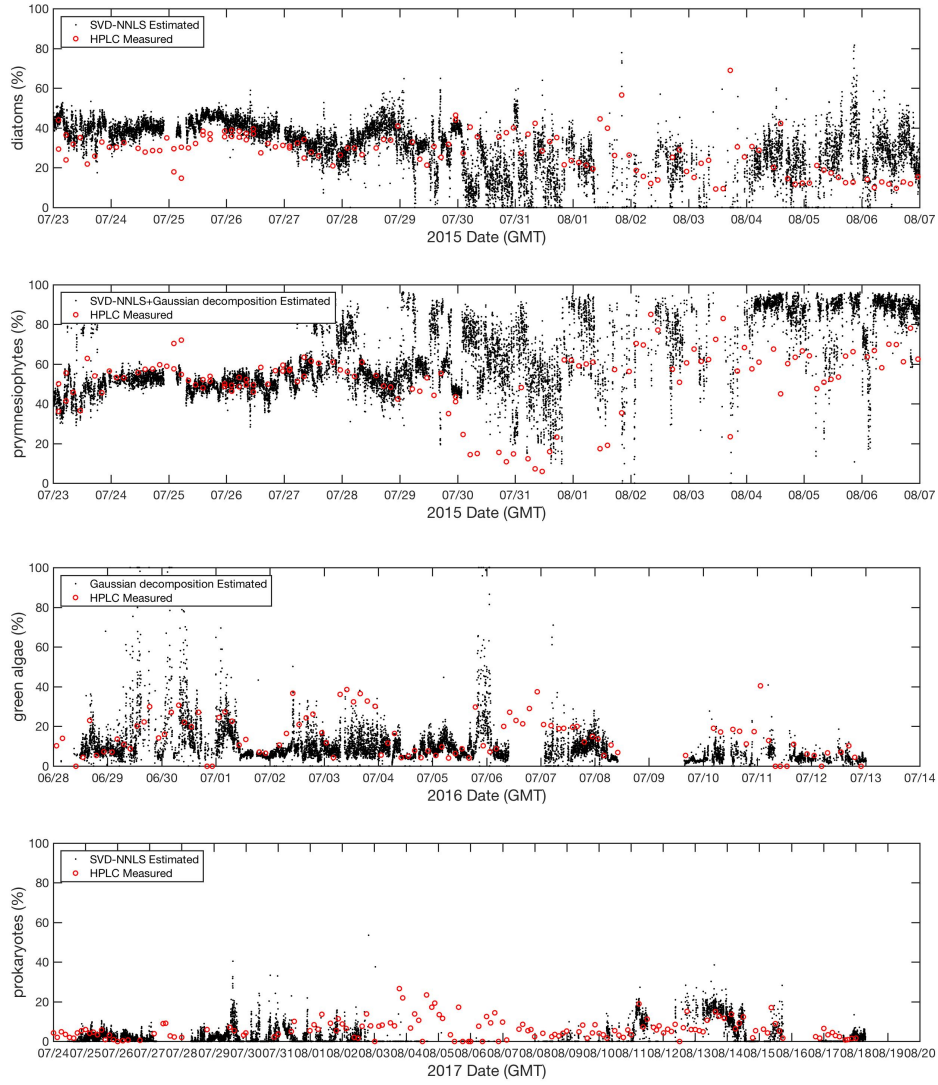
Figure 4: Time series of the contribution of diatoms, prymnesiophytes, green algae and prokaryotes to total biomass derived from both AC-S (black dots) and HPLC (red circle).

Table 1: The center wavelengths and widths of the Gaussian functions for phytoplankton pigments and pigment groups, and the statistics for the power function regression: $pigment\ concentration = A * a_{gaus}(\lambda_0)^B$.

$\lambda_0(nm)$	σ	pigment	A	B	R^2	$RMSE$
406	16	TChl-a	17.60	0.90	0.75	0.391
434	12	TChl-a	41.60	1.12	0.86	0.310
453	12	0.03(TChl-b)+0.07(Chl-c1/2)[a]	1.18	1.23	0.92	0.006
470	13	TChl-b	0.38	0.50	0.52	0.029
492	16	PPC	1.21	0.54	0.49	0.090
523	14	PSC	25.05	0.92	0.76	0.306
550	14	phycoerythrin	-	-	-	-
584	16	Chl-c1/2	11.54	0.84	0.68	0.101
617	13	TChl-a	19.78	0.56	0.65	0.461
638	11	Chl-c1/2	49.94	1.03	0.81	0.079
660	11	TChl-b	0.65	0.44	0.56	0.027
675	10	TChl-a	19.72	0.76	0.82	0.363

[a]Chase et al. (2013); Bricaud et al. (2004).

Table 2: The names and abbreviations of phytoplankton pigments and pigment groups analyzed in this study, and the statistics of phytoplankton pigments retrieval using Gaussian decomposition, SVD-NNLS, principal component analysis and artificial neural network, respectively based on ten-fold cross validation. RMSE and MAE are in log10 space, while R^2 and RPD in linear space. Note: TChl-a=monovinyl- + divinyl- chlorophyll-a + chlorophyllide-a; TChl-b=monovinyl- + divinyl- chlorophyll-b; PSC=Fuco + But + Hex + Peri, PPC=Allo + Diadino + Diato + Zea + α- + β-Caro (Roy et al., 2011).

pigments	abbreviation	Gaussian decomposition				SVD-NNLS			
		R^2	RMSE	MAE	RPD (%)	R^2	RMSE	MAE	RPD (%)
alloxanthin	Allo	-	-	-	-	0.20	0.408	0.294	114.6
chlorophyll-c1/2	Chl-c1/2	0.69	0.264	0.207	40.0	0.51	0.416	0.308	157.0
chlorophyll-c3	Chl-c3	-	-	-	-	0.35	0.648	0.500	484.6
α-carotene	α-Caro	-	-	-	-	0.16	0.528	0.425	217.6
β-carotene	β-Caro	-	-	-	-	0.26	0.306	0.229	60.8
diadinoxanthin	Diadino	-	-	-	-	0.26	0.464	0.330	171.5
diatoxanthin	Diato	-	-	-	-	0.28	0.401	0.321	48.9
fucoxanthin	Fuco	-	-	-	-	0.52	0.326	0.243	53.8
19'-hexanoyloxyfucoxanthin	Hex	-	-	-	-	0.47	0.485	0.344	227.1
19'-butanoyloxyfucoxanthin	But	-	-	-	-	0.16	0.739	0.585	589.8
neoxanthin	Neo	-	-	-	-	0.43	0.568	0.480	123.3
lutein	Lut	-	-	-	-	0.10	0.461	0.354	47.6
peridinin	Peri	-	-	-	-	0.32	0.500	0.395	181.5
prasinoxanthin	Prasino	-	-	-	-	0.75	0.344	0.316	60.6
violaxanthin	Viola	-	-	-	-	0.14	0.413	0.307	89.4
zeaxanthin	Zea	-	-	-	-	0.18	0.354	0.279	74.1
total chlorophyll-a	TChl-a	0.78	0.182	0.131	12.2	0.55	0.316	0.215	39.6
total chlorophyll-b	TChl-b	0.40	0.200	0.158	15.4	0.14	0.403	0.286	100.2
photosynthetic carotenoids	PSC	0.60	0.246	0.211	49.8	-	-	-	-
photoprotective carotenoids	PPC	0.39	0.278	0.188	34.3	-	-	-	-