Statistical approach for the retrieval of phytoplankton community structures from in situ fluorescence measurements

SHENGQIANG WANG,1,2,6 CONG XIAO,1 JOJI ISHIZAKA,3,7 ZHONGFENG QIU,1,2 DEYONG SUN,1,2 QIAN XU,4 YUANLI ZHU,4 YU HUAN,1 AND YUJI WATANABE5

1School of Marine Sciences, Nanjing University of Information Science & Technology, Nanjing 210044, China
2Jiangsu Research Center for Ocean Survey Technology, Nanjing University of Information Science & Technology, Nanjing 210044, China
3Institute for Space-Earth Environmental Research, Nagoya University, Nagoya 464-8601, Japan
4Graduate School of Environmental Studies, Nagoya University, Nagoya 464-8601, Japan
5KANSO Technos Co., LTD, Osaka 541-0052, Japan
6wsiqiang815@163.com
7joji.ishizaka@gmail.com

Abstract: Knowledge of phytoplankton community structures is important to the understanding of various marine biogeochemical processes and ecosystem. Fluorescence excitation spectra ($F(\lambda)$) provide great potential for studying phytoplankton communities because their spectral variability depends on changes in the pigment compositions related to distinct phytoplankton groups. Commercial spectrofluorometers have been developed to analyze phytoplankton communities by measuring the field $F(\lambda)$, but estimations using the default methods are not always accurate because of their strong dependence on norm spectra, which are obtained by culturing pure algae of a given group and are assumed to be constant. In this study, we proposed a novel approach for estimating the chlorophyll $a$ (Chl $a$) fractions of brown algae, cyanobacteria, green algae and cryptophytes based on a data set collected in the East China Sea (ECS) and the Tsushima Strait (TS), with concurrent measurements of in vivo $F(\lambda)$ and phytoplankton communities derived from pigments analysis. The new approach blends various statistical features by computing the band ratios and continuum-removed spectra of $F(\lambda)$ without requiring a priori knowledge of the norm spectra. The model evaluations indicate that our approach yields good estimations of the Chl $a$ fractions, with root-mean-square errors of 0.117, 0.078, 0.072 and 0.060 for brown algae, cyanobacteria, green algae and cryptophytes, respectively. The statistical analysis shows that the models are generally robust to uncertainty in $F(\lambda)$. We recommend using a site-specific model for more accurate estimations. To develop a site-specific model in the ECS and TS, approximately 26 samples are sufficient for using our approach, but this conclusion needs to be validated in additional regions. Overall, our approach provides a useful technical basis for estimating phytoplankton communities from measurements of $F(\lambda)$.

©2016 Optical Society of America

OCIS codes: (010.4450) Oceanic optics; (010.0280) Remote sensing and sensors.

References and links

1. A. Longhurst, S. Sathyendranath, T. Platt, and C. Caverhill, “An estimate of global primary production in the ocean from satellite radiometer data,” J. Plankton Res. 17(6), 1245–1271 (1995).
2. C. B. Field, M. J. Behrenfeld, J. T. Randerson, and P. Falkowski, “Primary production of the biosphere: integrating terrestrial and oceanic components,” Science 281(5374), 237–240 (1998).
3. E. Litchman, “Resource competition and the ecological success of phytoplankton,” in Evolution of Primary Producers in the Sea, P. Falkowski and A. H. Knoll, ed. (Academic Press, 2007).
4. T. Kanneda and J. Ishizaka, “Size-fractionated primary production estimated by a two-phytoplankton community model applicable to ocean color remote sensing,” J. Oceanogr. 61(4), 663–672 (2005).
30. H. Yamaguchi, H.-C. Kim, Y. B. Son, S. W. Kim, K. Okamura, Y. Kiyomoto, and J. Ishizaka, “Seasonal and summer interannual variations of SeaWiFS chlorophyll a in the Yellow Sea and East China Sea,” Prog. Oceanogr. 105, 22–29 (2012).

31. A. Isobe, M. Ando, T. Watanebe, T. Senjyu, S. Sugihara, and A. Manda, “Freshwater and temperature transports through the Tsushima-Korea Straits,” J. Geophys. Res.: Oceans 107(C7), 3065 (2002).

32. S. Wang, J. Ishizaka, H. Yamaguchi, S. Tripathy, M. Hayashi, Y. Xu, Y. Mino, T. Matsuno, Y. Watanebe, and S. Yoo, “Influence of the Changjiang River on the light absorption properties of phytoplankton from the East China Sea,” Biogeosciences 11(7), 1759–1773 (2014).

33. L. Van Heuvelen and C. S. Thomas, “Computer-assisted high-performance liquid chromatography method development with applications to the isolation and analysis of phytoplankton pigments,” J. Chromatogr. A 910(1), 31–49 (2001).

34. K. Furuya, M. Hayashi, Y. Yabushita, and A. Ishikawa, “Phytoplankton dynamics in the East China Sea in spring and summer as revealed by HPLC-derived pigment signatures,” Deep Sea Res. Part II Top. Stud. Oceanogr. 50(2), 367–387 (2003).

35. Z.-Y. Zhu, W.-M. Ng, S.-M. Liu, J. Zhang, J.-C. Chen, and Y. Wu, “Estuarine phytoplankton dynamics and shift of limiting factors: A study in the Changjiang (Yangtze River) Estuary and adjacent area,” Estuar. Coast. Shelf Sci. 84(3), 393–401 (2009).

36. S. W. Wright, R. L. van den Enden, I. Pearce, A. T. Davidson, F. J. Scott, and K. J. Westwood, “Phytoplankton community structure and stocks in the Southern Ocean (30–80 E) determined by CHEMTAX analysis of HPLC pigment signatures,” Deep Sea Res. Part II Top. Stud. Oceanogr. 57(9-10), 758–778 (2010).

37. M. Latasa, “Improving estimations of phytoplankton class abundances using CHEMTAX,” Mar. Ecol. Prog. Ser. 329, 13–21 (2007).

38. F. van der Meer, “Analysis of spectral absorption features in hyperspectral imagery,” Int. J. Appl. Earth Obs. Geoinf. 5(1), 55–68 (2004).

39. Z. Li, L. Li, K. Song, and N. Cassar, “Estimation of phytoplankton size fractions based on spectral features of remote sensing ocean color data,” J. Geophys. Res.: Oceans 118(3), 1445–1458 (2013).

40. I. Jolliffe, Principal component analysis (Springer, 2002).

41. J. Berkson, “Application of the logistic function to bio-assay,” J. Am. Stat. Assoc. 39, 357–365 (1944).

42. R. J. W. Brewin, S. Sathyendranath, T. Hirata, S. J. Lavender, R. M. Barciela, and N. J. Hardman-Mountford, “A three-component model of phytoplankton size class for the Atlantic Ocean,” Ecol. Modell. 221(11), 1472–1483 (2010).

43. T. Hirata, N. J. Hardman-Mountford, R. J. W. Brewin, J. Aiken, R. Barlow, K. Suzuki, T. Isada, E. Howell, T. Hashioka, M. Noguchi-Aita, and Y. Yamanaka, “Synchronic relationships between surface Chlorophyll-a and diagnostic pigments specific to phytoplankton functional types,” Biogeosciences 8(2), 311–327 (2011).

44. C.-F. J. Wu, “Jackknife, bootstrap and other resampling methods in regression analysis,” Ann. Stat. 14(4), 1261–1295 (1986).

45. S. E. Craig, C. T. Jones, W. K. W. Li, G. Lazin, E. Horne, C. Caverhill, and J. J. Cullen, “Deriving optical metrics of coastal phytoplankton biomass from ocean colour,” Remote Sens. Environ. 119, 72–83 (2012).

46. A. A. Taylor, B. Taylor, T. Dinter, B. Rötgers, and F. Steinmetz, “Using empirical orthogonal functions derived from remote sensing reflectance for the prediction of phytoplankton pigments concentrations,” Ocean Sci. 11(1), 139–158 (2015).

47. Y. Huot and M. Babin, “Overview of fluorescence protocols: theory, basic concepts, and practice,” in Chlorophyll a Fluorescence in Aquatic Sciences: Methods and Applications, D. J. Suggett, O. Präsl and M. A. Borowitzka, ed. (Springer, 2010).

48. S. Wang, J. Ishizaka, T. Hirawake, Y. Watanebe, Y. Zhu, M. Hayashi, and S. Yoo, “Remote estimation of phytoplankton size fractions using the spectral shape of light absorption,” Opt. Express 23(8), 10301–10318 (2015).

49. S. A. Kriss, S. E. Figary, G. L. Boyer, S. B. Watson, and M. R. Twiss, “Rapid in situ measures of phytoplankton communities using the bbe FluoProbe: evaluation of spectral calibration, instrument intercompatibility, and performance range,” Can. J. Fish. Aquat. Sci. 71(7), 1087–1095 (2014).

50. J. Harrison, T. E. Howell, S. B. Watson, and R. E. Smith, “Improved estimates of phytoplankton community composition based on in situ spectral fluorescence: use of ordination and field-derived norm spectra for the bbe FluoProbe,” Can. J. Fish. Aquat. Sci. 73(10), 1472–1482 (2016).

51. R. Alexander, P. Gikuma-Njuru, and J. Imberger, “Identifying spatial structure in phytoplankton communities using multi-wavelength fluorescence spectral data and principal component analysis,” Limnol. Oceanogr. Methods 10(6), 402–415 (2012).

52. A. M. Ciotti, M. R. Lewis, and J. J. Cullen, “Assessment of the relationships between dominant cell size in natural phytoplankton communities and the spectral shape of the absorption coefficient,” Limnol. Oceanogr. 47(2), 404–417 (2002).

53. A. Bricaud, C. Mejia, D. Blondeau-Patissier, H. Claustre, M. Crepon, and S. Thiria, “Retention of pigment concentrations and size structure of algal populations from their absorption spectra using multilayered perceptrons,” Appl. Opt. 46(8), 1251–1260 (2007).
1. Introduction

Phytoplankton are important organic organisms in oceanic waters due to their role as primary producers and as key players in biogeochemical cycling [1, 2]. Biological diversity is a typical trait for phytoplankton, and their community structure changes depending on various environmental conditions, such as temperature and nutrients [3]. Distinct phytoplankton taxonomic groups have different nutrient utilization efficiencies, photosynthesis rates, life cycles, biochemical requirements and roles in the marine food web [4, 5]. For instance, diatoms, which are the major utilizer of silicon, usually dominate in eutrophic waters and contribute approximately 20% of global carbon fixation, whereas cyanobacteria commonly exist in oligotrophic waters due to their high nitrite utilization rate. Additionally, some cyanobacteria are nitrogen fixers [5, 6]. Thus, knowledge of phytoplankton communities could improve our understanding of the functional roles of phytoplankton in various marine biogeochemical processes and in the marine ecosystem.

Several traditional methods are available to determine the phytoplankton community structure, such as flow cytometry, microscopy and pigments analysis by high-performance liquid chromatography (HPLC). The measurements of flow cytometry and microscopy methods have limitations with respect to their particle size detection range. Moreover, microscopic observations are time-consuming and often rely on the observer’s taxonomic experience [5]. HPLC pigment analyses are also popular methods for analyzing phytoplankton community structures [7]. Several marker pigments of distinct phytoplankton groups are selected to determine their fractional contribution to the chlorophyll $a$ concentration (Chl $a$) (units: mg m$^{-3}$) through statistical analysis [8]. A common disadvantage of all previously used methods to determine phytoplankton community structure is that their measurements depend on field water sampling. As such, the phytoplankton communities revealed by these methods are usually limited to discrete points that cannot accurately reflect their real spatial distribution.

In recent years, based on satellite ocean color observations, several bio-optical methods have been developed to estimate phytoplankton communities and their size structure [9]. However, the current passive remote sensing technology usually observes only surface waters.

Fluorescence emission is an important physiological trait of phytoplankton [10]. The light absorbed by Chl $a$ molecules is transferred for utilization in photosynthesis, and the excess energy is re-emitted as red-wavelength fluorescence or is dissipated as heat. Based on this trait, the measurement of in vivo fluorescence was introduced to detect changes in phytoplankton biomass in 1966 [11]. In addition to Chl $a$, the PSII pigment antenna of phytoplankton contains additional accessory pigments, such as chlorophylls, carotenoids and phycobilins, which absorb light from various regions of the spectrum [12]. Among the accessory pigments, photosynthetic pigments transfer almost all of the absorbed energy to Chl $a$; this energy is further transferred by Chl $a$ in the same way as described above [10]. The pigment composition is taxa specific, and different accessory pigments have significantly different light absorption spectra [12], which results in different fluorescence excitation spectra ($F(\lambda)$) (units: a.u.) for specific phytoplankton taxonomic groups [13]. Thus, $F(\lambda)$ have potential for the analysis of the phytoplankton community structure [14].

During the past decade, submersible spectrofluorometers, such as FluoroProbe (bbe moldaenke, USA) and Multi-Exciter (JFE Advantech Co., Ltd., Japan), were commercially developed to derive the phytoplankton composition by measuring in vivo $F(\lambda)$ [15, 16]. Because of their easy operation and rapid measurement, these spectrofluorometers have become increasingly popular in aquatic ecosystem investigations, for instance, assessing freshwater water quality [17], monitoring harmful algal blooms [18, 19], and investigating the
vertical distribution patterns of phytoplankton communities [20, 21]. These instruments utilize several modulated LED lamps with central wavelengths in the blue-green region to sequentially excite Chl \(a\) fluorescence (emission wavelength of approximately 680 nm). A detector is used to record the red fluorescence signals sequentially excited by the different LED lamps, and the recorded relative intensity of the fluorescence consequently forms the fluorescence excitation spectra \(F(\lambda)\). By culturing pure algae in the laboratory, \(F(\lambda)\) of specific phytoplankton groups are obtained; and these spectra are the so-called “norm spectra” and are stored in the instrument. Based on these “norm spectra”, spectral unmixing analysis is used as the default method to derive the Chl \(a\) amount of individual phytoplankton groups from the field \(F(\lambda)\) of mixed populations (see details in [15] and [16]).

Despite the increasingly common use of submersible spectrofluorometers in field surveys, the accuracy of the default estimations of phytoplankton communities is not satisfactory, as shown in this study later and as suggested by previous studies [22–24]. For instance, See et al. [22] found that the amount of Chl \(a\) of green algae and cryptophytes determined by FluoroProbe was overestimated when compared with the taxonomic analysis based on HPLC pigments. These inaccuracies are attributed to the spectral unmixing analysis, which strongly depends on the norm spectra [25]. As mentioned above, the norm spectra are usually obtained by culturing typical algae species representing distinct phytoplankton groups and are assumed to be constant. However, within a given phytoplankton group, phytoplankton species vary in different regions, especially from freshwater to oceanic water, which may produce considerable inter-variability for the norm spectra [13]. Thus, using constant norm spectra is not always appropriate to analyze the individual phytoplankton groups in various regions [25].

To reduce these uncertainties, researchers re-calibrated the norm spectra by culturing algae species that are common in their specific study regions [26, 27]. Although these re-calibrations improved the instrument performance, algae culture in the laboratory is time-consuming, labor-intensive and skill-required work. More importantly, the environmental conditions (e.g., light and nutrients) experienced by phytoplankton in the laboratory are different from those in nature. These environmental differences may cause differences in the light absorption and fluorescence emission spectra of cultured algae and those found in nature [24, 25], resulting in uncertainties in the determinations of phytoplankton communities when using the cultured norm spectra to analyze the field fluorescence excitation spectra.

In this study, we developed a novel approach that does not rely on \textit{a priori} knowledge regarding the norm spectra but rather uses purely statistical analysis of the features of field fluorescence excitation spectra \(F(\lambda)\) to determine the phytoplankton community structure. The model’s ability, robustness and sensitivity were evaluated and discussed. This approach is straightforward, and local models can be derived in new regions once concurrent measurements of \(F(\lambda)\) and phytoplankton community are available.

2. Materials and methods

2.1 Study area and sampling

The data used in this study were collected during cruises in the East China Sea (ECS) in July 2011 and 2013 and in the Tsushima Strait (TS) in July 2012 aboard the T/V \textit{Nagasaki maru} [Fig. 1]. The ECS is one of the largest marginal seas in the western North Pacific. In summer, it receives a huge amount of nutrient-rich freshwater from the Changjiang River, forming the Changjiang diluted water (CDW) [28], which enhances phytoplankton growth and modulated the community structure [29, 30]. The TS is a narrow, shallow strait between Korea and Japan that connects the ECS and the Sea of Japan. The water in the TS originates from the Kuroshio and the ECS in summer [31]. Field observations have suggested that phytoplankton groups in the TS possess the typical characteristics of those in the global oceans [32].

During the cruise, a submersible multi-excitation fluorometer (Multi-Exciter; JFE Advantech Co., Ltd., Japan) was used to measure the fluorescence excitation spectra.
instrument was equilibrated in surface waters for several minutes and was then slowly lowered in the water column to a maximum depth of approximately 90 m. At each station, water samples were collected for measurement of the phytoplankton pigment concentrations at 2–4 layers, including the surface, subsurface chlorophyll \(a\) maximum (SCM), above the SCM and below the SCM, using 5-l Niskin bottles mounted on a CTD/rosette system. In total, 141 samples, with concurrent measurements of the phytoplankton pigment concentrations and fluorescence spectra, were obtained.

Fig. 1. Locations of the sampling stations in the East China Sea (ECS) in 2011 and 2013 and in the Tsushima Strait (TS) in 2012.

2.2 Determination of the phytoplankton community based on pigment concentrations

Water samples (1–2 L) for the pigment analysis were filtered through 25 mm Whatman GF/F glass fiber filters under low vacuum pressure (<0.01 MPa) and dim light. The filters were immediately frozen in liquid nitrogen onboard and were stored in a deep freezer (−80 °C) onshore for subsequent laboratory analysis. The concentrations of 19 phytoplankton pigments were identified and quantified by reversed-phase HPLC with a Zorbax Eclipse XDB-C8 column (150 mm × 4.6 mm, 3.5 μm; Agilent Technologies) using the method of Van Heukelem and Thomas [33], calibrated by commercial pigment standards (Sigma Aldrich, St. Louis, USA and DHI, Hørsholm, Denmark). The identification of pigment was based on the retention time and its absorbance spectrum with a photodiode array detector.

The phytoplankton community structure was then derived from the HPLC pigment concentrations using CHEMTAX program (version 1.95) [8], which is widely used for phytoplankton taxonomy studies [34–36]. In brief, this program yields the contributions of each phytoplankton group to Chl \(a\) (in terms of the Chl \(a\) fraction) using the ratio matrices of their biomarker pigment to Chl \(a\) and the corresponding HPLC pigment concentrations as inputs. Based on the knowledge of the phytoplankton communities in adjacent areas reported by previous studies [34, 35], 8 phytoplankton groups were considered in this study, including diatoms, dinoflagellates, prymnesiophytes, chrysophytes, cyanobacteria, prasinophytes, chlorophytes and cryptophytes. Prochlorophytes were excluded due to their absence from the majority of stations of our cruises. To obtain the most accurate analysis results, the CHEMTAX program was tuned to our study region, as described below [37]. Firstly, we ran the CHEMTAX program using biomarker ratio matrices based on those reported in adjacent regions as the initial values [34]. Then, using the output ratio matrices yielded by the last run as the inputs, the CHEMTAX program was ran again. This process was repeated until the outputs became stable.

Subsequently, due to the similar characteristics of the fluorescence excitation spectra \(F(\lambda)\) of phytoplankton groups [15], the 8 phytoplankton groups identified from the HPLC pigments
were combined into 4 groups, including brown algae, cyanobacteria, green algae and cryptophytes. The Chl \( a \) fraction of brown algae was calculated as the sum of the Chl \( a \) fractions of diatoms, dinoflagellates, prymnesiophytes and chrysophytes, and the green algae represented the combination of chlorophytes and prasinophytes.

2.3 Fluorescence excitation spectra measurements

The fluorescence excitation spectra \( F(\lambda) \) of phytoplankton at 9 wavelengths (center wavelengths of 375, 400, 420, 435, 470, 505, 525, 570 and 590 nm) were determined from the Multi-Exciter measurements. The excitation wavelengths were designed in consideration of the maximum absorption of photosynthetic pigments. The Multi-Exciter measures chlorophyll fluorescence emission at 685 nm, which was alternately excited by the 9 excitation wavelengths (see detailed descriptions in [16]). Before the cruise, the instrument was calibrated by the manufacturer.

We also calculated the phytoplankton community using the default instrument method. Briefly, the default method uses a spectral unmixing model, which can be expressed as [16]

\[
F(\lambda) = \sum_{i=1}^{m} C_i N_i
\]

where \( F(\lambda) \) is the measured total fluorescence excitation spectra, \( m \) is the number of considered phytoplankton groups, which is equal to 3 in the default method, \( C_i \) (units: mg m\(^{-3}\)) is the Chl \( a \) of the \( i \)th phytoplankton group, and \( N_i \) (units: a.u. mg\(^{-1}\) m\(^3\)) is the norm spectrum of the \( i \)th phytoplankton group, which is obtained from the cultured algae in the laboratory.

Figure 2 shows the comparisons between the Chl \( a \) fractions of the three phytoplankton groups and the HPLC pigment values. Large estimations errors were observed. The factors responsible for the poor estimates will be discussed later.

2.4 Model development

The first-order variability in \( F(\lambda) \) is driven by phytoplankton biomass (strong relationship between \( F(\lambda) \) at 470 nm and Chl \( a \) was observed in our data set, with a \( R^2 \) value of 0.72), and the second-order variability (mainly reflected by spectral shape) is related to phytoplankton community changes. Therefore, this study attempted to analyze the phytoplankton community using the characteristics of the spectral shape of \( F(\lambda) \). We constructed two types of spectral features, the band ratio and continuum-removed spectra, to highlight the \( F(\lambda) \) shape variation. The band ratio (\( F_{\text{BR}}(\lambda_1, \lambda_2) \)) (dimensionless) was defined as

\[
F_{\text{BR}}(\lambda_1, \lambda_2) = \frac{F(\lambda_1)}{F(\lambda_2)} , \lambda_2 > \lambda_1
\]
In total, 36 $F_{BR}(\lambda_1, \lambda_2)$ were constructed from all possible band combinations. The continuum-removed spectra $F_{CR}(\lambda_1, \lambda_2, \lambda_3)$ (dimensionless) was calculated as

$$F_{CR}(\lambda_1, \lambda_2, \lambda_3) = \frac{F(\lambda_3)}{F(\lambda_1, \lambda_2, \lambda_3)} \lambda_3 > \lambda_2 > \lambda_1$$  \hspace{1cm} (3)$$

where $F_C (\lambda_1, \lambda_2, \lambda_3)$ is the minimal convex hull of a straight line fitted over the top of a spectrum. Continuum removal is a useful procedure to amplify the peak and trough of a spectrum and has been widely used in remote sensing studies [38, 39]. The physical meaning of continuum-removed spectra can be found in the studies of Van Der Meer [38] and Li et al. [39].

In this study, by considering all possible band combinations, we constructed 84 $F_{CR}(\lambda_1, \lambda_2, \lambda_3)$. By combining $F_{BR}(\lambda_1, \lambda_2)$ and $F_{CR}(\lambda_1, \lambda_2, \lambda_3)$, 120 spectral features were obtained. Considering the potential information redundancy of these 120 features, principal component analysis (PCA) was performed to compress the data dimensions. Here, we performed PCA on the entire data set composed of both $F_{BR}(\lambda_1, \lambda_2)$ and $F_{CR}(\lambda_1, \lambda_2, \lambda_3)$, rather than separate data sets of $F_{BR}(\lambda_1, \lambda_2)$ and $F_{CR}(\lambda_1, \lambda_2, \lambda_3)$, because 1) variation magnitudes of $F_{BR}(\lambda_1, \lambda_2)$ and $F_{CR}(\lambda_1, \lambda_2, \lambda_3)$ were comparable, and 2) model performances using the two manners of PCA were generally similar. To simplify the model, we performed PCA on the combined data set of $F_{BR}(\lambda_1, \lambda_2)$ and $F_{CR}(\lambda_1, \lambda_2, \lambda_3)$. In brief, PCA rotates the original axes to new, orthogonal axes, along which the variance decreases sequentially. The output of PCA consisted of two terms, the principal component (PC) modes and PC scores. PC modes are orthogonal eigenvectors that define the rotations of the axes, and PC scores are uncorrelated new variables that are the linear combinations of the original spectral features with the corresponding PC modes [40]. The PC scores were used to derive the Chl $a$ fractions of the phytoplankton groups. In this study, the sigmoid function was used to quantify the relationship between the Chl $a$ fractions of the phytoplankton groups and the PC scores and can be expressed as [41]

$$f = \frac{1}{1 + \exp[-(c_0 + \sum_{i=1}^{k} c_i S_i)]}$$  \hspace{1cm} (4)$$

where $f$ (dimensionless) stands for the Chl $a$ fractions of brown algae, cyanobacteria or green algae; $S_i$ (dimensionless) represents the $i$th PC score; $c_0$ and $c_i$ are dimensionless and represent the model regression coefficients; and $k$ (dimensionless) is the number of PCs. Because more than 99% of the variance of the original data set was explained by the first 8 PC scores, $k$ was set to 8. The selection of the sigmoid function in this study was made with the consideration that the sigmoid function has upper and lower limits of 1 and 0 [41], respectively, so unrealistic estimation (beyond the range of 0–1) of the Chl $a$ fraction values can be avoided. To determine the regression coefficients $c_0$ and $c_i$, nonlinear least square method was used in this study [41]. The fitting residuals were normally distributed and independent of $S$, implying that the regression analysis is reasonable.

For cryptophytes, their fractions were calculated as 1 minus the total fractions of brown algae, cyanobacteria and green algae. As clarified in section 2.2, in our study region, it is appropriate to consider 4 phytoplankton groups in total, i.e., brown algae, cyanobacteria, green algae and cryptophytes. Considering that the sum of the fractions of the 4 groups should be equal to 1, and also the relatively low abundance of cryptophytes in our data set, we used the difference method to estimate cryptophyte fractions. It must be admitted that the difference method is empirical without optical basis; and errors in estimations of brown algae, cyanobacteria and green algae will be introduced to estimation of cryptophytes. However, such an approach is useful and has been widely used for determining phytoplankton size structure [42, 43]. Meanwhile, cryptophyte fractions estimated using the difference method have generally performed well, as shown in sections 3.2 and 4.1 following. The fractions of
cryptophytes can also be estimated using the fitting method based on spectral features of $F(\lambda)$ (in this case, using the difference method for green algae); we found that the performances using the fitting method and difference method were generally comparable (data not shown).

The flow chart of the model is shown in Fig. 3. In general, our approach comprises three steps: 1) spectral features construction; 2) data dimension reduction by PCA and 3) function fitting. The model training procedure is conducted to obtain the PC modes and regression coefficients $c_0$ and $c_i$, which are then used in the model to estimate the phytoplankton group fractions.

**3. Results**

**3.1 Data distributions**

The phytoplankton composition showed wide dynamic ranges and significant variability [Table 1 and Fig. 4]. The Chl $a$ fraction of brown algae for all samples ranged from 0.263 to 0.986, with the largest mean value of 0.678 (standard variation of 0.193). The mean fraction of cryptophytes was the lowest, at 0.062, and the values of cyanobacteria and green algae were between the mean fractions of brown algae and cryptophytes. In the ECS, phytoplankton showed complex compositions during 2011 July. The brown algae of most samples was generally the dominant group, and cyanobacteria, green algae and cryptophytes had comparable fractions. However, in 2013, the phytoplankton communities were mainly composed of brown algae and green algae. In the TS, brown algae contributed the majority Chl $a$, and cyanobacteria were the second most important group.

Similarly, large variations were observed in both the magnitude and spectra shape of the fluorescence excitation spectra $F(\lambda)$ [Fig. 5]. The $F(\lambda)$ of most of samples possessed spectral peaks at 470 nm, and the peaks of some samples fell at 435 nm. Meanwhile, some samples had a second spectral peak at 570 nm. The $F(\lambda)$ at 443, 470 and 570 nm significantly varied from 0.147 to 17.225, from 0.140 to 24.082 and from 0.131 to 3.580, respectively [Table 1 and Fig. 5(a)]. The magnitude changes were mainly related to variation in the phytoplankton biomass, which was confirmed by the strong correlation between $F(470)$ and Chl $a$ ($r = 0.85, p < 0.01$) in log-log space (data not shown).
Table 1. Statistics of HPLC-derived Chl a fractions ($f$) of each phytoplankton group, fluorescence excitation spectra ($F(\lambda)$) and normalized $F(\lambda)$ by the spectral mean value ($N_F(\lambda)$) (N = 141). Max, Min, SD and CV indicate the values of maximum, minimum, standard deviation and coefficient of variation, respectively.

| Variable                  | Min  | Max  | Mean | SD  | CV (%) |
|---------------------------|------|------|------|-----|--------|
| $f$ (dimensionless)       |      |      |      |     |        |
| Brown algae               | 0.263| 0.986| 0.678| 0.193| 28.5   |
| Cyanobacteria             | 0.001| 0.637| 0.150| 0.138| 92.1   |
| Green algae               | 0.000| 0.415| 0.110| 0.090| 82.0   |
| Cryptophytes              | 0.000| 0.547| 0.062| 0.084| 134.1  |
| $F(\lambda)$ (a.u.)      |      |      |      |     |        |
| $F(435)$                  | 0.147| 17.225| 1.419| 1.910| 134.5  |
| $F(470)$                  | 0.140| 24.082| 1.745| 2.690| 154.2  |
| $F(570)$                  | 0.131| 3.580| 0.665| 0.480| 72.2   |
| $N_F(\lambda)$ (dimensionless) |      |      |      |     |        |
| $N_F(435)$                | 0.694| 1.746| 1.472| 0.178| 12.1   |
| $N_F(470)$                | 0.692| 2.402| 1.622| 0.367| 22.6   |
| $N_F(570)$                | 0.357| 2.480| 0.947| 0.441| 46.6   |

The variability in the magnitude of $F(\lambda)$ was significantly reduced by normalization using the spectral mean values, while the variation in the spectral shape became more obvious, especially the spectral peaks at 570 nm [Table 1 and Fig. 5(b)]. To investigate the spectral characteristics of different phytoplankton groups, we examined the normalized $F(\lambda)$ of several samples with the highest Chl a fractions of brown algae and cyanobacteria [Figs. 5(c) and 5(d)]. The normalized $F(\lambda)$ dominated by brown algae showed clear peaks at 470 nm, while for cyanobacteria-dominant samples, peaks at 570 nm were more obvious, which were generally consistent with those reported by previous studies [15, 16, 24].
3.2 Performance of the estimation models

Based on the analysis of the spectral characteristics of $F(\lambda)$, we constructed 120 spectral features to distinguish different phytoplankton groups. Subsequently, PCA was applied to reduce the data dimensions, and the PC scores and HPLC-derived Chl $a$ fractions of the phytoplankton groups were used to fit the regression function [Eq. (4)]. The fitted functions were then applied to derive the Chl $a$ fractions of different phytoplankton groups for the same data set used for model validation. We used the same data set for model calibration and validation in order to evaluate the model fitting performance. As shown in Fig. 6, the $F(\lambda)$-derived Chl $a$ fractions from the new model had good consistency with HPLC-derived values, even for green algae and cryptophytes, which had the lowest Chl $a$ fractions in our data set. The data were generally clustered along an approximately 1:1 line, with $R^2$ values of 0.64, 0.68, 0.36, and 0.49 and RMSE values of 0.117, 0.078, 0.072 and 0.060 for brown algae, cyanobacteria, green algae and cryptophytes, respectively [Table 2 and Fig. 6]. For most of the samples, the absolute difference between the $F(\lambda)$-derived Chl $a$ fractions and the HPLC-derived values was less than 0.2 (within the ± 0.2 fraction range, as shown in Fig. 6). These results indicate that the newly developed method has good ability for deriving the phytoplankton community from $F(\lambda)$. 

Fig. 5. Characteristics of the fluorescence excitation spectra $F(\lambda)$ (a), $F(\lambda)$ normalized by the spectral mean value $N_\lambda F(\lambda)$ (b), $N_\lambda F(\lambda)$ dominated by brown algae (c) and $N_\lambda F(\lambda)$ dominated by cyanobacteria (d).
Fig. 6. Scatterplots of HPLC-derived Chl a fractions of brown algae (a), cyanobacteria (b), green algae (c) and cryptophytes (d) versus the values derived from the fluorescence excitation spectra $F(\lambda)$. Black lines indicate the 1:1 line, and grey dashes indicate the ± 0.2 fraction range relative to the 1:1 line.

Table 2. Summary of the performances (RMSE) of the full models calibrated using the full data set, the East China Sea (ECS) models calibrated using the ECS data set and the Tsushima Strait (TS) models calibrated using the TS data set.

| Models                | Data set | Brown algae | Cyanobacteria | Green algae | Cryptophytes |
|-----------------------|----------|-------------|---------------|-------------|--------------|
| Full models           | ECS      | 0.128       | 0.092         | 0.084       | 0.076        |
|                       | TS       | 0.098       | 0.054         | 0.052       | 0.028        |
|                       | ECS and TS | 0.117       | 0.078         | 0.072       | 0.060        |
| ECS models or TS models | ECS   | 0.116       | 0.086         | 0.082       | 0.072        |
|                       | TS       | 0.055       | 0.033         | 0.027       | 0.022        |
|                       | ECS and TS | 0.095       | 0.069         | 0.064       | 0.056        |

To evaluate the performance of the model, a cross-validation was conducted. The models developed in this study are essentially for calibrating field spectrophotometers based on concurrent measurements of $F(\lambda)$ and Chl a fractions of phytoplankton groups derived from HPLC pigments. After calibrations using our approach, the field spectrophotometers can be used to determine phytoplankton communities accurately. We determined the adequate number of samples needed for model calibration by conducting a cross-validation using the variable jackknife procedure [44]. In brief, we randomly selected $d$ samples (the number of samples of $d$) using MATLAB™ from the full data set (the number of samples of $N$, i.e., $N = 141$) with HPLC pigment and $F(\lambda)$ measurements, which were used as the model validation data set. The remaining subset of $N - d$ samples was used for model calibration. The random selection of $d$ samples yielded a huge amount of possible combinations for the validation subset, which would significantly affect the overall computing time. Therefore, we only examined 1000 combinations for $d$ samples, similar to the suggestions of previous studies [45, 46]. For each combination, the model calibration subset was used to determine the PC modes and the regression coefficients $c_0$ and $c_i$ following the model training procedure [Fig. 3, left panel], and the spectral features of $F(\lambda)$ of the validation data set, together with the calibrated
PC modes and $c_0$ and $c_i$, were used to determine the Chl $a$ fractions of each phytoplankton group [Fig. 2, right panel]. Comparison between the $F(\lambda)$-derived fractions and the HPLC-derived values was then performed, and the RMSE values were calculated. This procedure was repeated for 1000 combinations of $d$ samples. The values of the jackknife RMSE (RMSE_J) were calculated as the mean RMSE from 1000 trials. RMSE_J was also scaled to the RMSE from the model calibrated using the full data set ($N = 141$). Furthermore, to determine the minimum number of samples for model training, we varied $d$ from 11 to 136 in intervals of 5 and conducted the above procedure for each value of $d$.

As shown in Fig. 7, similar variation patterns of the jackknife RMSE versus training sample number were observed for the four phytoplankton groups. The jackknife RMSE decreased rapidly from 0.250, 0.251, 0.246 and 0.119 for 11 training samples to 0.152, 0.126, 0.117 and 0.088 for 26 training points for the Chl $a$ fraction estimations of brown algae, cyanobacteria, green algae and cryptophytes, respectively. For training samples larger than 26, the jackknife RMSE gradually became stable but continued to decrease slightly with increasing number of training samples. When the number of training samples reached 136, the jackknife RMSE decreased to 0.120, 0.086, 0.077 and 0.056, similar to the RMSE of the models calibrated on the full data set. These results suggest that the models developed in this study were generally robust for determining the Chl $a$ fractions of phytoplankton groups. These findings imply that when using our new approach in the ECS and TS, approximately 26 samples are generally sufficient for model calibration and additional samples may only yield small improvements in model performance; this conclusion should be further validated in additional regions.

![Fig. 7. Number of training points versus Jackknife RMSE (RMSE_J) (left axis) and the ratio of RMSE_J to RMSE of the model fitted on the full data set (right axis) for estimating the Chl $a$ fractions of brown algae (a), cyanobacteria (b), green algae (c) and cryptophytes (d).]

**3.3 Analysis of the model sensitivity to uncertainties in $F(\lambda)$ bands**

To evaluate the models’ sensitivities to each $F(\lambda)$ band and the model’s robustness, we conducted a sensitivity analysis experiment. During the experiment, random errors, which uniformly varied within $\pm 20\%$ of the original $F(\lambda)$ values, were added to one band of $F(\lambda)$ for all samples at the one time. The synthetic $F(\lambda)$ values were used to derive the Chl $a$ fractions of
phytoplankton groups using the models calibrated based on the original $F(\lambda)$ data set. The relative changes in the RMSE values of the estimated fractions were calculated. We repeated this process 1000 times, and the mean values of the relative errors were calculated by averaging the changes across the 1000 repetitions. This procedure was conducted for each band of $F(\lambda)$.

For the Chl $a$ fraction estimation model of brown algae, the most sensitive wavelength was 435 nm, whereas for cyanobacteria, in addition to the blue bands, 570 nm showed large sensitivity (Fig. 8). In contrast to the estimation models of brown algae and cyanobacteria, the model for estimating the Chl $a$ fraction of green algae was most sensitive to 420 and 470 nm. The different wavelength sensitivities of these models may be related to the distinct spectral characteristics of $F(\lambda)$ of different phytoplankton groups. Meanwhile, Fig. 8 shows that, except in cases of the effects of uncertainties in $F(\lambda)$ at 435 and 470 nm for the fraction estimation model of brown algae and the effects of uncertainties in $F(\lambda)$ at 435 nm for the fraction estimation model of cyanobacteria, when random errors ranging from −20% to 20% were incorporated into $F(\lambda)$, the relative errors of the models for estimating the Chl $a$ fraction of phytoplankton groups were less than 20%, especially for green algae. These findings indicate that the Chl $a$ fraction estimation models are generally robust to uncertainties in $F(\lambda)$.

3.4 Preliminary model application

We preliminarily applied the developed models to one section profile (red line marked in Fig. 1) of $F(\lambda)$ measured during a 2013 cruise in the ECS as a case study. The selected section comprised profile measurements of $F(\lambda)$ at 7 stations. To remove potential measurement noise, the vertical resolution of $F(\lambda)$ at each station was resampled from 0.3 to 0.4 m into 1 m. Subsequently, the estimation models were applied to derive the Chl $a$ fractions of the phytoplankton groups.

As shown in Fig. 9, clear differences among the Chl $a$ fractions of phytoplankton groups were observed. In general, brown algae had the highest Chl $a$ fractions in the whole water column, and the Chl $a$ contribution from cryptophytes was the lowest. Brown algae often showed relatively higher Chl $a$ fractions at the subsurface chlorophyll $a$ maximum layer than the surface and bottom waters. However, at the west stations, which are often influenced by the nutrient-rich freshwater of the Changjiang River, abundant brown algae were detected in the surface water. For cyanobacteria, high fractions were only observed at the surface layers of the east stations, which are influenced by Kuroshio oligotrophic waters. Green algae contributed relatively little Chl $a$ and were uniformly distributed in the water column. These variation patterns of the phytoplankton group distributions, which were derived from $F(\lambda)$ using the
models developed in this study, were generally consistent with those of previous studies conducted in adjacent regions [34]. Understanding the causes of the distribution patterns of phytoplankton groups is beyond the scope of this study; however, nutrient and light conditions related to water stratification and Changjiang River discharge are possible causes, as suggested by previous studies [35].

![Fig. 9. Profile distributions of the Chl a fractions of brown algae (a), cyanobacteria (b), green algae (c) and cryptophytes (d) derived from the fluorescence excitation spectra.](image)

4. Discussion

4.1 New approach for the determination of phytoplankton community structures

Phytoplankton fluorescence is an important property that links the biology of phytoplankton to optics [10]. The spectral shape of the fluorescence excitation spectra $F(\lambda)$ is strongly related to the phytoplankton community through the dependence of light absorption on the pigment composition [10, 13]. In this study, we observed large spectral variation in the fluorescence excitation and distinct spectral shapes for samples dominated by brown algae and cyanobacteria [Table 1, Fig. 5]. In general, the most obvious spectral differences were reflected in the spectral peaks at 435 and 470 nm, which are caused by changes in the accessory pigment composition, or in green regions, which are related to the phycobilins that commonly exist in cyanobacteria. These findings agree with those documented by previous studies [13, 25, 47] and provide a solid optical basis for deriving phytoplankton community structures from $F(\lambda)$.

To capture spectral variation related to the phytoplankton community, we blended various spectral features using calculations of the band ratio and the continuum removal, which are useful tools that are commonly used in spectral analysis [39]. Although 120 spectral features were constructed, not all of these features had considerable variability. Thus, PCA was used to reduce the data dimensions and to select the features with the greatest variability to form new orthogonal variables. Note that the purpose of using PCA here was different from previous studies [45, 46, 48], which used PCA to decompose the total spectral variation of remote sensing reflectance or light absorption into individual components. The first 8 PC modes explained more than 99% of the variance, and some modes were actually poorly correlated with changes in the Chl $a$ fractions of some phytoplankton groups. Excluding these PC modes did not impact the model’s ability to derive the phytoplankton community. This phenomenon was also true for the constructed spectral features that have minimal variation. To maintain consistency between the models for different phytoplankton groups and to ensure the easy operability of the new approach, we suggest using all the constructed spectral features and the first 8 PC modes when implementing our approach.

We must state here that our intent was not to develop a model that could be globally applied to derive phytoplankton community structures because the $F(\lambda)$ of the defined phytoplankton group may vary to some extent due to changes in phytoplankton populations within the same
group from location to location [13, 25]. In this study, although we combined the 8 phytoplankton groups derived from HPLC pigment analysis into 4 groups based on spectral similarity, we observed differences among the $F(\lambda)$ values that were dominated by brown algae, especially between the ECS samples and the TS samples. In the ECS, brown algae-dominant $F(\lambda)$ had peaks at 435 nm, whereas in the TS, the spectral peaks shifted to 470 nm (data not specifically shown). This shift was probably caused by the changes in the phytoplankton population within the brown algae because brown algae in the ECS were usually a mix of diatoms, dinoflagellates, prymnesiophytes and chrysophytes, while in the TS, they were mainly composed of diatoms. The inter-variability of the $F(\lambda)$ of the same group may reduce the estimation accuracy of the phytoplankton community when using a global model in the ECS and TS. This was confirmed by the following experimental comparison. Using the approach of this study, local models were developed for the ECS and the TS, respectively. These local models significantly improved the estimations of the phytoplankton community, especially in the TS [Table 2 and Fig. 10]. For all samples from the ECS and the TS, RMSE values decreased from 0.117, 0.078, 0.072 and 0.060 to 0.095, 0.069, 0.064 and 0.056 for brown algae, cyanobacteria, green algae and cryptophytes, respectively. Therefore, site-specific models derived using our approach may produce more accurate model estimations. A more detailed definition of phytoplankton groups than the 4 groups used in this study may reduce the inter-spectral variability of a defined phytoplankton group and thereby improve the estimations of the phytoplankton community. Because the abundances of dinoflagellates, prymnesiophytes and chrysophytes are low in our data set, it is difficult to evaluate this hypothesis at this stage.

Fig. 10. Scatterplots of the HPLC-derived Chl $a$ fractions of brown algae (a), cyanobacteria (b), green algae (c) and cryptophytes (d) versus the values derived from the fluorescence excitation spectra $F(\lambda)$ using the models calibrated on the ECS and TS data sets. Black lines indicate the 1:1 line, and grey dashes indicate the ± 0.2 fraction range relative to the 1:1 line.
Several factors may possibly impact the approach developed in this study. Changes in the physiological states of phytoplankton related to varying environmental conditions may alter pigment composition and thereby $F(\lambda)$ [47]. Such spectral variability may produce uncertainty in deriving the phytoplankton community. It is known that separation of the effects of phytoplankton community and physiology on $F(\lambda)$ is not easy. In our data set, because of the lack of appropriate measurements such as time-series observations from a fixed station over an entire day, at this stage, it is difficult to quantify to the extent to which physiology vs. community regulates $F(\lambda)$; we believe $F(\lambda)$ should be mainly controlled by the phytoplankton community based on our observations. To possibly reduce physiological effects, we suggest that data sampling cover a large variation of the phytoplankton community and environmental conditions as conducted in this study, which should ensure more accurate model calibration. Moreover, colored dissolved organic matter (CDOM) also yields fluorescence at 680 nm, although its emission peak is at approximately 500 nm and the emission at 680 nm is usually weak in oceanic conditions [47]. The data set used in this study covered clear water from the TS and river-influenced waters from the ECS, in which CDOM varied greatly (absorption of CDOM at 440 nm ranged from 0.015 to 0.159 $m^{-1}$), while our models still produced good estimations for the phytoplankton community [Figs. 6 and 10], similar to the findings of Kring et al. [49]. These results indicate that our approach is robust to changes in CDOM in our regions, but its applicability to waters with greater variation of CDOM must be determined.

In this study, the phytoplankton community used for model training and validation was determined from HPLC pigments by CHEMTAX program, which may not completely reflect the actual phytoplankton composition. CHEMTAX analysis depends both on photosynthetic and photoprotective pigments, while $F(\lambda)$ is only theoretically associated with photosynthetic pigments. This would cause a disjunction between the fluorescence method and CHEMTAX analysis for cyanobacteria estimations in this study, since cyanobacteria determined by CHEMTAX is mainly based on zeaxanthin, which is a photoprotective pigment and does not regulate $F(\lambda)$, while spectral features of $F(\lambda)$ used for cyanobacteria estimations are probably based on phycobilins, which commonly exist in cyanobacteria and may mainly regulate their $F(\lambda)$. This disjunction may bring some uncertainties in estimations of actual cyanobacteria fractions. However, for model calibration and validation, although we observed phytoplankton community determined by HPLC pigment analysis, our approach does not have to depend on this method. Once additional actual phytoplankton communities can be measured using methods other than HPLC pigment analysis, our approach can be accordingly adapted to develop estimation models of the phytoplankton community.

4.2 Comparisons with other approaches and implications

Compared with determinations of phytoplankton communities by flow cytometry, microscopy and HPLC pigment analysis, spectrofluorometer measurements are rapid and low-cost. More importantly, spectrofluorometer measurements are almost continuous, and thus, phytoplankton communities derived from $F(\lambda)$ can more objectively reflect their actual distributions in water columns, while those determined by flow cytometry, microscopy and HPLC pigment analysis are limited to discrete water depths because of the strong dependences of their measurements on water sampling. However, the current default methods of commercial spectrofluorometers yield large uncertainties in the estimations of phytoplankton communities, as found by previous studies [22, 23, 50] and as shown in this study [Fig. 2]. These uncertainties are attributed to the assumption of a constant value of the factory-set norm spectra obtained by culturing pure algae in the laboratory. In contrast, our approach does not require a priori knowledge of the norm spectra, and the site-specific models can be easily developed based on synoptically-collected $F(\lambda)$ and the phytoplankton community. This allows our approach to accommodate the inter-variability of the $F(\lambda)$ of the same phytoplankton group induced by possible changes in phytoplankton species within the given group and the environmental conditions.
In addition, as an improvement to the study of Alexander et al. [51], Harrison et al. [50] recently proposed an approach to obtain norm spectra directly from field measurements of $F(\lambda)$ by statistical analysis, while this approach relies on the fact that the collected data set contains samples overwhelmingly dominated by one defined phytoplankton group (Chl $a$ fraction close to 1). Note that in natural conditions, it is usually difficult to meet this requirement. For instance, in our data set, the highest Chl $a$ fraction of cyanobacteria was less than 0.7, and the fractions of green algae and cryptophytes were much lower. In contrast, our approach does not need specific samples as required by the approach of Harrison et al. [50]. For low-abundance phytoplankton groups, e.g., green algae and cryptophytes in this study, our approach could yield satisfactory estimations of the Chl $a$ fractions [Figs. 6 and 10]. We must admit that our approach requires synchronous measurements of field $F(\lambda)$ and the phytoplankton community derived from HPLC pigment analysis, while the analysis in section 3.2 suggests that approximately 26 samples are sufficient for model calibration [Fig. 7]. This suggestion should be effective in the ECS and TS, while its validity in additional regions need to be further examined.

Except the fluorescence excitation spectra $F(\lambda)$, the light absorption of phytoplankton ($a_{ph}(\lambda)$) is also suggested to be related to the phytoplankton pigment composition and cell size [12]. Therefore, models have been proposed for the retrieval of the phytoplankton pigment concentrations, community and size structure from $a_{ph}(\lambda)$ [52–54]. These studies suggest that if $a_{ph}(\lambda)$ can be accurately measured using field instruments, such as WET labs AC-9 or AC-S, the pigment concentrations, community and/or size structure can be determined in the field. However, the field measurement of $a_{ph}(\lambda)$ suffers from the influences of non-algae particles and CDOM to a greater extent than fluorescence [16], which increases the difficulty of obtaining accurate measurements of $a_{ph}(\lambda)$ using field instruments. These characteristics suggest that fluorescence is a better trait than $a_{ph}(\lambda)$ to estimate phytoplankton communities in the field.

Knowledge of phytoplankton communities is crucial to understanding various marine biogeochemical processes and ecosystems. The approach proposed in this study can be easily implemented to calibrate commercial spectrofluorometers to yield accurate estimations of phytoplankton communities. Recently, autonomous underwater vehicles and gliders have been successfully applied to water surveys [55]. Based on these platforms, spectrofluorometers can be used to study the spatial variability in phytoplankton communities across regions of interest. In addition, ship-borne and air-borne laser fluorometers have increasingly attracted attention in the measurement of fluorescence not only in surface but also profile with certain depth from space [56]. Once multispectral fluorescence excitation could be measured from space, the approach of this study could be adopted to derive phytoplankton communities from space. Overall, our approach provides a practical basis for the determination of phytoplankton communities based on fluorescence excitation spectra.

5. Conclusions

In this study, we proposed an innovative approach to determine phytoplankton community structures from field fluorescence excitation spectra. In contrast to the widely used spectral unmixing analysis method, this approach does not require a priori knowledge of the norm spectra of pure algae but uses statistical spectral features of the band ratios and continuum removal. The evaluation of the models derived using our approach showed good performance, with RMSE values of 0.117, 0.078, 0.072 and 0.060 for brown algae, cyanobacteria, green algae and cryptophytes, respectively. The models described here were generally robust when errors were introduced to the bands of the fluorescence excitation spectra. Our approach can be used to study the spatial variation of phytoplankton communities based on field measurements of fluorescence excitation spectra. This was confirmed by the reasonable spatial distributions of the derived phytoplankton communities obtained by applying the new approach in the East China Sea. In general, our approach can be easily implemented in other regions.
site-specific model, we found that approximately 26 samples with concurrent measurements of the phytoplankton community and fluorescence excitation spectra were sufficient in the case of the East China Sea and the Tsushima Strait, but this suggestion and the model's applicability need to be validated in additional locations.

Funding

Natural Science Foundation of Jiangsu Province (BK20150914, BK20151526); National Natural Science Foundation of China (NSFC) (41506200, 41576172, 41276186); Natural Science Foundation of the Jiangsu Higher Education Institutions of China (15KJB170015); National Key Research and Development Program of China (2016YFC1400901, 2016YFC1400904); Global Observation Mission-Climate (GCOM-C) Project of the Japan Aerospace Exploration Agency (JAXA).

Acknowledgments

We would like to thank the captains, officers and crew of the T/V Nagasaki maru for their support and assistance during the field sampling and measurements. We would also like to thank the anonymous reviewers for their constructive comments towards improving this manuscript.