Multivariate approach for the retrieval of phytoplankton size structure from measured light absorption spectra in the Mediterranean Sea (BOUSSOLE site)

Emanuele Organelli,* Annick Bricaud, David Antoine, and Julia Uitz

Laboratoire d'Océanographie de Villefranche, UMR 7093, CNRS and Université Pierre et Marie Curie, Paris 6, Villefranche sur Mer 06238, France

*Corresponding author: emanuele.organelli@obs-vlfr.fr

Received 12 November 2012; revised 5 February 2013; accepted 11 February 2013; posted 22 February 2013 (Doc. ID 179785); published 4 April 2013

Models based on the multivariate partial least squares (PLS) regression technique are developed for the retrieval of phytoplankton size structure from measured light absorption spectra (BOUSSOLE site, northwestern Mediterranean Sea). PLS-models trained with data from the Mediterranean Sea showed good accuracy in retrieving, over the nine-year BOUSSOLE time series, the concentrations of total chlorophyll a [Tchl a], of the sum of seven diagnostic pigments and of pigments associated with micro, nano, and picophytoplankton size classes separately. PLS-models trained using either total particle or phytoplankton absorption spectra performed similarly, and both reproduced seasonal variations of biomass and size classes derived by high performance liquid chromatography. Satisfactory retrievals were also obtained using PLS-models trained with a data set including various locations of the world's oceans, with however a lower accuracy. These results open the way to an application of this method to absorption spectra derived from hyperspectral and field satellite radiance measurements.

OCIS codes: 010.4450, 010.1030, 010.7340, 010.0010.

1. Introduction

Phytoplankton are a major component of ocean's biogeochemical cycles, especially in the epipelagic zone where they regulate the total amount of carbon and other elements in the oceans [1]. When analyzing biogeochemical fluxes in the oceans, however, it is inadequate to consider phytoplankton as a single variable (i.e., chlorophyll a) because the various phytoplankton groups (e.g., diatoms, coccolithophores, cyanobacteria) have different roles in many marine biogeochemical processes, such as carbon fixation and export, nitrogen fixation and silicon uptake [2–4].

This is the rationale for the development of a new generation of bio-optical products able to identify different phytoplankton types, in order to continuously analyze changes in algal communities at regional and global scale [5–7], and in view of refining biogeochemical models. Currently, several bio-optical methods are proposed to analyze and quantify the temporal and spatial variability of phytoplankton communities in the world's oceans. These approaches, using inherent or apparent optical properties (IOPs and AOPs), focus on the retrieval of products such as phytoplankton types [8–10], size classes [11–15], dominant size class [16–18], phytoplankton size distribution [19,20], or phytoplankton pigments [21–23].

Many efforts have been dedicated to the development of products for the retrieval of the phytoplankton size structure. Partitioning phytoplankton into their

1559-128X/13/112257-17$15.00/0

micro, nano, and picocomponents [24] is considered a good ecological indicator [6] with fundamental implications in a biogeochemical and trophic (food web) context [25–27]. The rationale is that the cell size influences many eco-physiological processes such as the sinking rate and the nutrient uptake [28], or the pigment packaging within the cell [29,30]. The latter, in particular, drives modifications in the spectral characteristics of the light absorption coefficients [31] that can be actually used as the basis of methods for the retrieval of the algal community size structure from space-derived or *in situ* absorption measurements (see [27,32] and references therein).

A further approach to extract information on the size structure of algal communities from light absorption properties is the multivariate partial least squares (PLS) regression technique [33,34]. This technique, which is frequently used in chemistry for spectroscopy analysis, has been only scarcely applied in oceanography. The first PLS application was performed about 10 years ago to determine concentrations of chlorophyll and phaeo-pigments in solution from their absorbance spectra [35]. Progressively, the application of the PLS technique was extended to the retrieval of algal classes abundance either from fluorescence [36] or absorption spectra [37–39]. In particular, Stæhr and Cullen [38] showed the remarkable skill of the PLS technique in predicting the fraction of chlorophyll biomass of the harmful algae *Karenia mikimotoi* both in controlled and in natural conditions. On the basis of the observed low sensitivity of PLS to absorption spectral variations induced by different irradiances, Stæhr and Cullen [38] also recommended the PLS for the detection in the natural environment of phytoplankton types other than *K. mikimotoi*, provided that the algorithm is developed using a large number of samples in order to achieve retrievals with a high degree of confidence. These considerations, in addition to the uncertainties and the various sources of errors still observed in the application of several current approaches [32] for the retrieval of phytoplankton size classes from optical data, are the rationales for testing the potential of the PLS technique in this field.

In the framework of the BIOoptics and CARbon Experiment (BIOCAREX) and BOUée pour l'acquisition de Séries Optiques à Long termE (BOUSSOLE) projects, we developed a new algorithm based on the multivariate PLS technique in order to retrieve information on phytoplankton pigments and size structure from a long time series of hyperspectral absorption measurements performed monthly at the BOUSSOLE site (northwestern Mediterranean Sea) since 2003. In view of a possible application of such a method to various IOPs derived from inversion of AOPs (see [40]), the prediction ability of the PLS is investigated both for total particle or phytoplankton absorption measurements. For the development of the PLS models, we used an extensive data set of phytoplankton and particle light absorption spectra coupled with high performance liquid chromatography (HPLC) pigment measurements collected from the first optical depth of the world's oceans. A nine-year time series of measurements at the BOUSSOLE site is then used for testing the models. Finally, changes in the phytoplankton community structure observed from the application of the new models to the entire BOUSSOLE time series are discussed and compared with those retrieved from HPLC pigment measurements.

2. Methods

A. Sampling

Samples used to train models (see Subsection 2.D) were collected between 1991 and 2004 during 12 oceanographic cruises in different seasons and across the world's oceans [Table 1, Fig. 1]. In order to ensure the homogeneity of the data set with respect to the processing procedure, additional data from other publicly available data sets were not used in this work. The data from the cruises carried out between 1991 and 2001 were described and used in Bricaud *et al.* [30] while those from the BIOSOPE cruise can be found in Bricaud *et al.* [41]. Information on the additional data collected during the AOPEX

Table 1. Cruises, Location, Sampling Period, Number of Samples (*n*) and [Tchl *a*] range for the First Optical Depth, for the Data Used to Train Models

Cruise	Location	Period	*n*	[Tchl *a*] Range mg m^{-3}
EUMELI 3	Tropical North Atlantic	Oct. 1991	5	0.073–0.340
FLUPAC	Equatorial and subequatorial Pacific	Sep.–Oct. 1994	11	0.039–0.236
OLIPAC	Equatorial and subequatorial Pacific	Nov. 1994	34	0.072–0.291
MINOS	Eastern and western Mediterranean Sea	May 1996	24	0.028–0.070
ALMOFRONT II	Alboran Sea (Mediterranean Sea)	Dec. 1997–Jan. 1998	59	0.202–1.185
PROSOPE (upw)	Morocco upwelling	Sep. 1999	10	2.03–4.04
PROSOPE (Med)	Eastern and western Mediterranean Sea	Sep.–Oct. 1999	102	0.020–0.221
POMME 1	North Atlantic	Feb.–March 2001	116	0.105–0.933
POMME 2	North Atlantic	March–May 2001	125	0.254–1.44
POMME 3	North Atlantic	Aug.–Oct. 2001	125	0.039–0.395
AOPEX	Tyrrhenian Sea (Mediterranean Sea)	Aug. 2004	43	0.047–0.092
BIOSOPE	South Pacific	Nov.–Dec. 2004	62	0.017–1.481

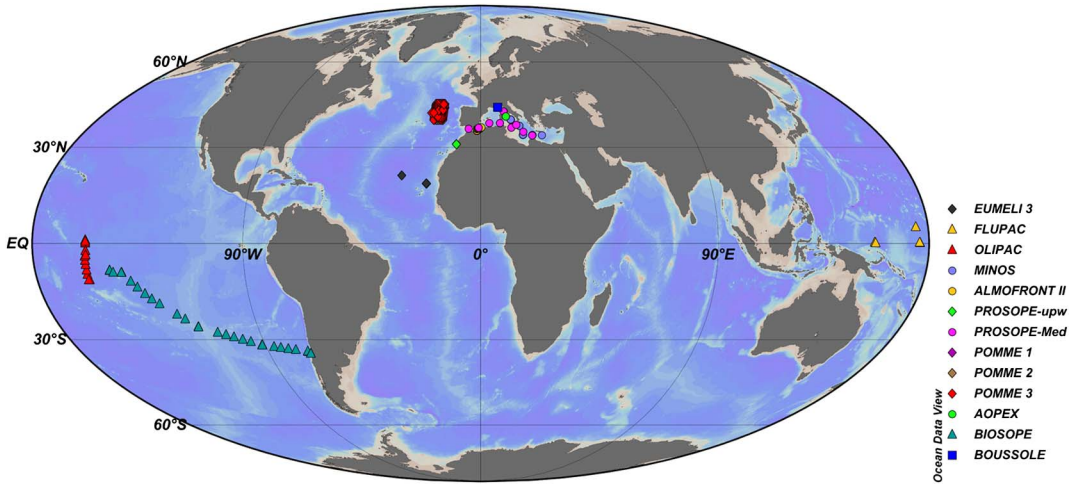

Fig. 1. (Color online) Map of the stations where data were collected. Stations are displayed according to their geographical distribution (square = BOUSSOLE site, circle = Mediterranean Sea, diamond = Atlantic Ocean, triangle = Pacific Ocean) and to the oceanographic cruise during which they were visited. The map is drawn by the Ocean Data View software (Schlitzer, R., Ocean Data View, http://odv.awi.de, 2012).

cruise in August 2004 is available in Antoine *et al.* [42]. At the BOUSSOLE site in the Mediterranean Sea (7°54'E, 43°22'N; Fig. 1), samples for particulate absorption measurements have been collected monthly since 2003 [42,43].

All the data considered here were collected in Case-1 waters as defined by Morel and Prieur [44]. Water collection was performed at various depths of the 0–400 m water column during up-cast CTD profiles of temperature, conductivity, and chlorophyll fluorescence performed by a CTD-fluorometer rosette system equipped with Niskin bottles. Seawater samples were collected and filtered for subsequent determination of phytoplankton pigments and particulate absorption spectra. Only samples collected within the first optical depth [45] are analyzed here in order to minimize the influence of photoacclimation and its possible effects on the pigment packaging and thus on the spectral shape of the phytoplankton light absorption [29]. The first optical depth was computed for each station as being $Z_{eu}/4.6$, where the euphotic depth Z_{eu} is the depth at which the photosynthetically available radiation is reduced to 1% of its value just below the surface. The euphotic depth was either calculated from radiometric measurements (downward irradiance profiles) or estimated from the measured chlorophyll profile following Morel and Maritorena [46]. From a total of 6657 samples, 1211 belong to the first optical depth: 727, from various areas of the world's oceans, are used for model training and 484, from the BOUSSOLE time series, are used for the test.

B. Algal Pigment Measurements

Algal pigment measurements were carried out by HPLC. Seawater samples (up to 5.6 L) were filtered through 25 mm Whatman glass-fiber filters (GF/F), immediately frozen in liquid nitrogen and subsequently stored in the laboratory at −80°C until analysis. HPLC procedures are described in Claustre and Marty [47] for the EUMELI 3 cruise, Ras *et al.* [48] for the BIOSOPE cruise, and Vidussi *et al.* [49] for all other cruises. The procedure used for the AOPEX and BOUSSOLE cruises is comparable to that described by Vidussi *et al.* [50] (see [42]). Chlorophyll *a* and divinyl-chlorophyll *a* were fully resolved for all cruises but EUMELI3. Hereafter, the sum of chlorophyll *a*, divinyl-chlorophyll *a*, and chlorophyllide *a* concentrations is named total chlorophyll *a* concentration and noted [Tchl *a*].

Seven major diagnostic pigments (DPs) were selected as being representative of the three phytoplankton size classes (micro, nano, and pico phytoplankton). According to Vidussi *et al.* [50], these pigments are fucoxanthin (Fuco), peridinin (Perid), alloxanthin (Allo), 19′-butanoyloxyfucoxanthin (19′-BF), 19′-hexanoyloxyfucoxanthin (19′-HF), zeaxanthin (Zea), and chlorophyll *b* + divinyl chlorophyll *b* (Chl *b* + DV Chl *b*). The concentrations of these biomarker pigments were used to calculate the biomass proportions associated with micro, nano, and picophytoplankton size classes [11]:

$$\%\text{microphytoplankton} = 100(1.41[\text{Fuco}] + 1.41[\text{Perid}])/\text{DP}, \quad (1)$$

$$\%\text{nanophytoplankton} = 100(0.60[\text{Allo}] + 0.35[19' - \text{BF}] + 1.27[19' - \text{HF}])/\text{DP}, \quad (2)$$

$$\%\text{picophytoplankton} = 100(0.86[\text{Zea}] + 1.01[\text{Chl}\,b + DV\text{Chl}\,b])/\text{DP}, \quad (3)$$

where DP is the sum of the weighted concentrations of the seven bio-marker pigments.

The numerical coefficients used to compute the contribution of the three size classes to the taxonomic structure of the algal community were calculated by multiple regression on a global data set by Uitz *et al.* [11]. They actually represent the average ratios between [Tchl *a*] and each marker pigment. As already noted [11,51], such a distribution of DPs may yield some errors and uncertainty in the evaluation of the algal size classes because some pigments can be shared by various phytoplankton groups and some groups can be found in more than one size class. In spite of these possible sources of error and ambiguity, this method has been shown to provide reasonable information on the size structure and taxonomic composition of algal communities at global scale [30,41,48,52–54]. Note, however, that slight modifications in the repartition of pigments within size classes have been recently proposed by Brewin *et al.* [14] and Hirata *et al.* [21].

A size index (SI) was derived from Eqs. (1)–(3) in order to assess the variations of the dominant size class of the phytoplankton communities as [30]:

$$\begin{aligned} \mathrm{SI} = {} & (1(\%\text{picophytoplankton}) \\ & + 5(\%\text{nanophytoplankton}) \\ & + 50(\%\text{microphytoplankton}))/100, \end{aligned} \tag{4}$$

where 1, 5, and 50 μm are central size values for each size class.

As already acknowledged, SI is only a rough indicator of the size because of the unique central size used to represent each size class [30]. Nevertheless, it is a single parameter able to represent the dominant size of the phytoplankton communities.

C. Spectral Light Absorption Measurements

Particle absorption spectra ($a_p(\lambda)$) were measured using the "quantitative filter pad technique" (QFT) except for the FLUPAC cruise where the "glass-slide technique" [55] was used. The procedure is extensively described by Antoine *et al.* [42] for AOPEX and BOUSSOLE cruises, Bricaud *et al.* [41] for the BIOSOPE cruise, and Bricaud *et al.* [30,56] for all the other cruises. Briefly, seawater samples (up to 11.2 L) were filtered through 25 mm Whatman GF/F filters, immediately frozen in liquid nitrogen and then stored in a −80°C freezer in laboratory until the analysis. Spectra were measured every 1 nm in the visible-near infrared range by a spectrophotometer equipped with an integrating sphere. A blank wet filter was used as a reference. Optical densities were shifted to 0 in the near infrared, and then transformed into absorption coefficients (in m^{-1}). All spectra were corrected for the path length amplification effect (β-effect) using the algorithms given by Allali *et al.* [57] for samples collected during the OLIPAC, MINOS, PROSOPE (Mediterranean part), POMME 3, BIOSOPE (oligotrophic and mesotrophic waters) cruises, and by Bricaud and Stramski [58] for all the other samples. Absorption spectra collected during the FLUPAC cruise were not corrected for the β-effect, which occurs only when the QFT is used. Finally, the particulate absorption spectra $a_p(\lambda)$ were decomposed into phytoplankton ($a_{\mathrm{phy}}(\lambda)$) and nonalgal particle ($a_{\mathrm{NAP}}(\lambda)$) absorption spectra using the numerical decomposition described by Bricaud and Stramski [58], except for the samples of the EUMELI 3 and BIOSOPE cruises where the chemical procedure described by Kishino *et al.* [59] was used. In the present study, only the absorption values between 400 and 700 nm are considered.

D. Retrieval of Phytoplankton Size Structure from Absorption Spectra

The retrieval of pigment information and size structure of algal communities in the surface layer of the BOUSSOLE site from absorption spectra can be achieved by the development of a model based on the multivariate PLS regression technique. PLS is a multivariate analysis technique that relates by regression a data matrix of predictor variables (X) to a data matrix of response variables (Y). Basically, PLS consists of two steps: first, a model explaining the relations between dependent and independent variables has to be found (training step). Practically, the PLS technique decomposes an X matrix using the dependent variables in order to obtain model parameters and select the best number of latent variables (i.e., components) that maximize the covariance between X and Y variables. Second, the parameters of the PLS model can be used for the prediction of dependent variables from several independent variables of a new data set (testing step) [33,34].

Here we used the fourth-derivative absorption spectra as the independent variables. The fourth-derivative analysis introduced by Bidigare *et al.* [60] was performed (in the range 400–700 nm) using a finite approximation algorithm that computes the changes in curvature of a given spectrum within an interval $\Delta\lambda$ [$\Delta\lambda = \lambda_2 - \lambda_1$, where $\lambda_2 > \lambda_1$; see example in Fig. 2]. The fourth-derivative was chosen over the second-derivative because it enables a better separation of absorption bands and the quantification of

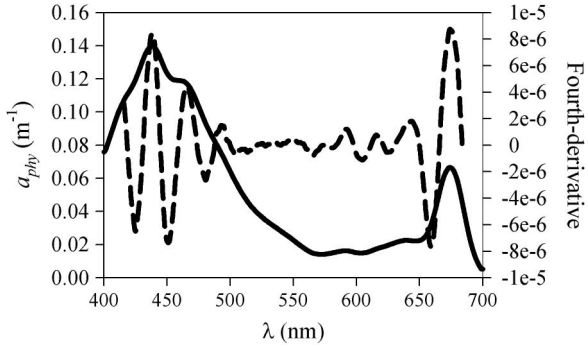

Fig. 2. Example of a smoothed phytoplankton absorption spectrum (solid curve) of the BOUSSOLE time series and its fourth-derivative (dashed curve).

pigments while the second-derivative was observed to only provide qualitative identification of pigments [60]. Because the fourth-derivative analysis is sensitive to the signal-to-noise ratio, the "mean filter" described by Tsai and Philpot [61] was used to smooth the absorption spectra before computation of the derivatives. Briefly, this filter assigns the mean value of all points within a sampling interval to the middle point of the window. In this study $\Delta\lambda$ was set to 8 nm for the derivative analysis and 9 nm was the size selected for the "mean filter," according to the range of optimal values showed in the analysis performed by Torrecilla *et al.* [23]. Finally, fourth-derivative absorption spectra composed of 269 wavelengths (from 416 to 684 nm) with 1 nm resolution were obtained and used.

The weighted concentrations of the seven DPs associated with the three phytoplankton size classes (see Subsection 2.B for details) and the total chlorophyll *a* concentrations are used as the dependent variables. Hence, five response variables were chosen: concentration of [Tchl *a*], sum of the concentrations of the seven DPs, and sum of the concentrations of the DPs associated with each size class separately.

The classical approach of the PLS (PLS1), which applies to a single variable at a time, is used to develop the models. Models were trained using two different data sets and tested on the BOUSSOLE data set. A flowchart summarizes the distribution and use of all the data in this study (Fig. 3). The first training data set comprises 716 simultaneous HPLC pigment and light absorption measurements ($a_p(\lambda)$ and $a_{\mathrm{phy}}(\lambda)$) collected during the cruises listed in Table 1 and includes samples collected at global scale (hereafter denoted GLOCAL). In order to assess also the performances of regional trained PLS models, the second training data set is built using data from the Mediterranean Sea only as collected during the MINOS, ALMOFRONT II, PROSOPE, and AOPEX cruises (hereafter denoted MedCAL). Using these cruises only, the MedCAL data set would essentially include oligo- to mesotrophic waters whereas the BOUSSOLE site, on which the model will be tested, also exhibits eutrophic waters during the spring phytoplankton bloom. Therefore, a small number ($n = 11$) of high-chlorophyll samples from the BOUSSOLE time series were also included in the MedCAL data set. These samples, when removed from the time series, did not substantially change its

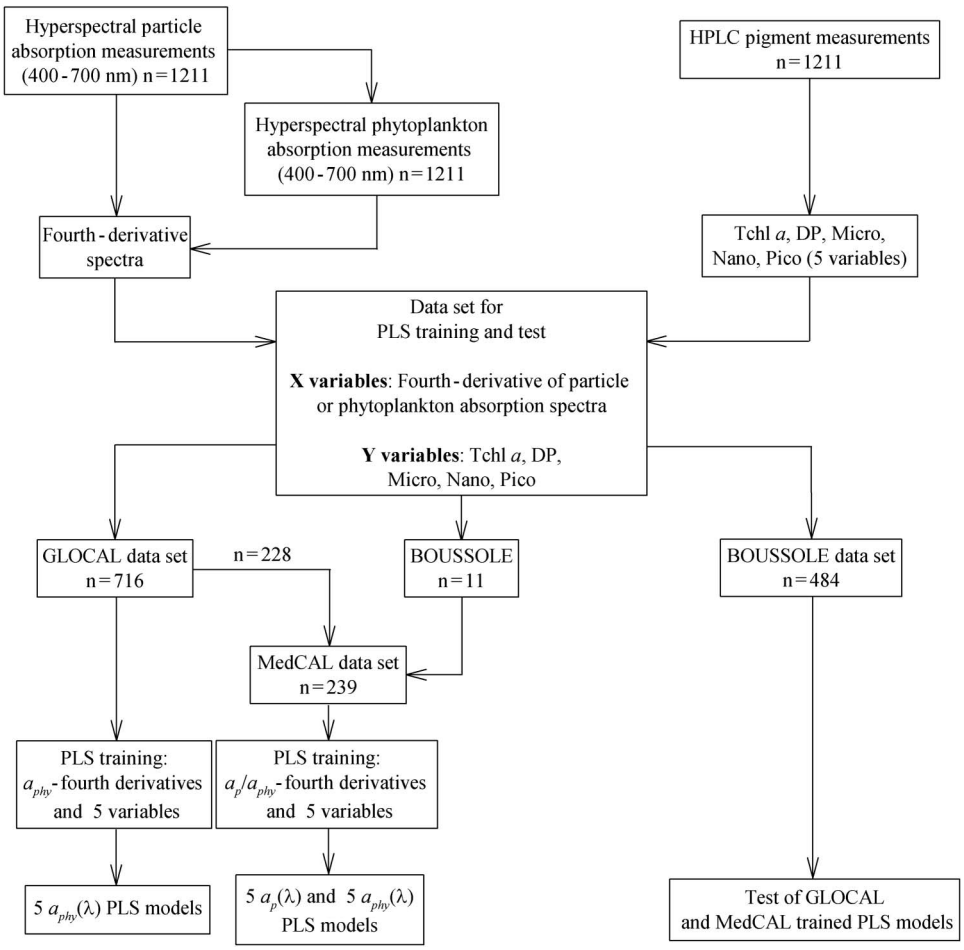

Fig. 3. Flowchart displaying distribution and use of HPLC pigment and spectral light absorption data for subsequent training and test of the *PLS* regression method.

temporal trend. Finally, the MedCAL data set includes 239 simultaneous HPLC pigment and light absorption measurements (see flowchart in Fig. 3).

The models were trained with PLS including leave-one-out (LOO) cross-validated predictions. Briefly, LOO validation computes a model by removing one data point at a time from the training data set and uses the fitted model to predict the value of the left out data point. The LOO cross-validation is used here to estimate the expected accuracy level of the predictive model. In order to determine the optimal number of components that minimized the error of prediction, the root mean square error of prediction (RMSEP) between LOO predicted and HPLC measured values was computed and the best number of components was selected for the lowest RMSEP value [62]. When the lowest RMSEP value occurred with a high number of components, to avoid overfitting, the number of components after which the error of prediction did not significantly decrease was considered as optimal [63]. PLS models were tested using the BOUSSOLE data set (see flowchart in Fig. 3) and their performances in predicting pigment information and size classes at the BOUSSOLE site were evaluated using the coefficient of determination (r^2), the RMSEP and the systematic error (BIAS). RMSEP and BIAS were computed as follows:

$$\mathrm{RMSEP} = \left(\sum_{i=1}^{n}(x_i - \bar{x}_i)^2/n\right)^{1/2} \quad (5)$$

and

$$\mathrm{BIAS} = \sum_{i=1}^{n}(\bar{x}_i - x_i)/n \quad (6)$$

where x_i was the measured value and $\bar{x}_i$ the value predicted by the models.

All the PLS analyses presented in this study were carried out by the "pls" package [63] for the free statistical software *R* (www.r-project.org).

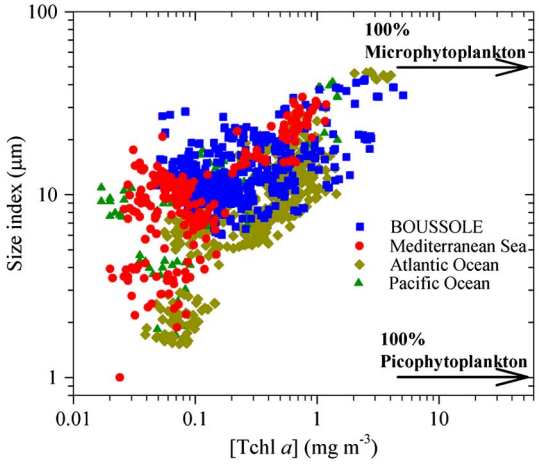

Fig. 4. (Color online) Variations of the size index (SI) derived from the relative contributions of micro, nano, and picophytoplankton [Eqs. (1)–(4)] as a function of [Tchl *a*] for BOUSSOLE, compared to various areas.

3. Results and Discussion

A. Size Characteristics of Algal Communities

The main bio-optical characteristics of the data sets used in the present study are reported and discussed by Antoine *et al.* [42] for the BOUSSOLE and the AOPEX cruises and by Bricaud *et al.* [30,41] for all the other cruises in Table 1. Here, we describe only the variations in the size structure of the algal communities when relevant to the results of the PLS application. To address this question, the variations of the size index (SI) as a function of [Tchl *a*] are analyzed.

The variations of the SI as a function of [Tchl *a*] for the cruises listed in Table 1 and for the BOUSSOLE data set are shown in Fig. 4. The previous study of

Table 2. PLS Parameters of $a_p(\lambda)$-and $a_{phy}(\lambda)$-Models Trained Using HPLC Pigment Measurements and Absorption Spectral Values Included in the MedCAL Data Set (n = 239), from Left to Right: Number of Components (N), RMSEP (mg m^{-3}), Explained Variance (%) for Independent [r^2X (%)] and Dependent [r^2Y (%)] Variables[a]

	LOO Prediction						
	N	RMSEP	r^2X (%)	r^2Y (%)	r^2	b	a
$a_p(\lambda)$ Models							
Tchl *a*	4	0.1038	96.12	98.63	0.97	0.99	0.005
DP	3	0.0879	95.40	97.99	0.97	0.98	0.006
Micro	4	0.1031	96.10	95.20	0.85	0.90	0.014
Nano	4	0.0789	95.50	94.88	0.84	0.87	0.012
Pico	6	0.0221	97.64	95.76	0.87	0.88	0.006
$a_{phy}(\lambda)$ Models							
Tchl *a*	3	0.1086	95.63	98.56	0.96	1.00	0.004
DP	2	0.0857	95.18	97.12	0.97	0.98	0.007
Micro	4	0.1085	96.31	96.56	0.84	0.91	0.010
Nano	4	0.0832	96.24	95.06	0.82	0.86	0.010
Pico	5	0.0207	97.29	95.62	0.88	0.88	0.010

[a]Statistical parameters for linear regressions between leave-one-out (LOO) predicted and measured pigment concentrations: determination coefficient (r^2), regression slope (b) and y-intercept (a).

Bricaud *et al.* [30] stated that, despite a general trend of covariation of SI with [Tchl *a*], the world's oceans are characterized by a different distribution of the three algal size classes for a given [Tchl *a*]. These results were then confirmed when data from the South Pacific Ocean (BIOSOPE cruise) were added to the data set and the dominant algal size in the clearest waters was revealed to be three times larger than those found for the same [Tchl *a*] level in the Mediterranean Sea [41]. The minimum [Tchl *a*] value during the nine-year BOUSSOLE time series was 0.047 mg m^{-3} (Fig. 4), which is slightly larger than [Tchl *a*] measured in extremely clear, picophytoplankton-dominated, Mediterranean waters observed during the PROSOPE cruise [30], or during other surveys in the Ligurian Sea [64] and other Mediterranean areas [50,54]. Nanophytoplankton was the dominant size class even in the clearest waters of the BOUSSOLE site (SI values close to 10 μm). This is a typical community structure observed also for samples from other areas of the Mediterranean Sea (see [30] for details) while, at similar [Tchl *a*], picophytoplankton is more present in the Atlantic and Pacific Oceans. The contribution of microphytoplankton increased with [Tchl *a*] at BOUSSOLE and SI values were similar to those found in the Mediterranean Sea, but generally higher than those observed in the North Atlantic [Fig. 4]. Samples with the highest [Tchl *a*] (up to 5 mg m^{-3}) were dominated by microphytoplankton, with SI values up to 42 μm, similar to those observed for the diatom-dominated waters of the Morocco upwelling [30]. These atypical eutrophic conditions for the Mediterranean Sea, observed at BOUSSOLE especially in 2005, have been recently reported by Marty and Chiavérini [65] in the Ligurian Sea at the Dyfamed station (near the BOUSSOLE site) as an effect of a more intense winter mixing compared to other years. Spring blooms characterized by nanophytoplankton (18–21 μm) were also observed at BOUSSOLE, as already reported in the same area during the bloom period at the Dyfamed time series [64].

The above observations suggest that the distribution of the three size classes at the BOUSSOLE site is, for a given [Tchl *a*], consistent with most of the Mediterranean samples, whereas major differences appear with respect to the Atlantic and Pacific Oceans. However, some particularities of the BOUSSOLE site with respect to other sampled areas of the Mediterranean Sea have to be taken into account, i.e., the absence of very clear picophytoplankton-dominated waters and the presence of eutrophic conditions and nanophytoplankton-dominated [Tchl *a*] maxima.

B. Retrieval of Phytoplankton Community Structure from the MedCAL Data Set

In the following sections, we present and compare the performances of PLS-models trained using either the total particle or the phytoplankton light absorp-

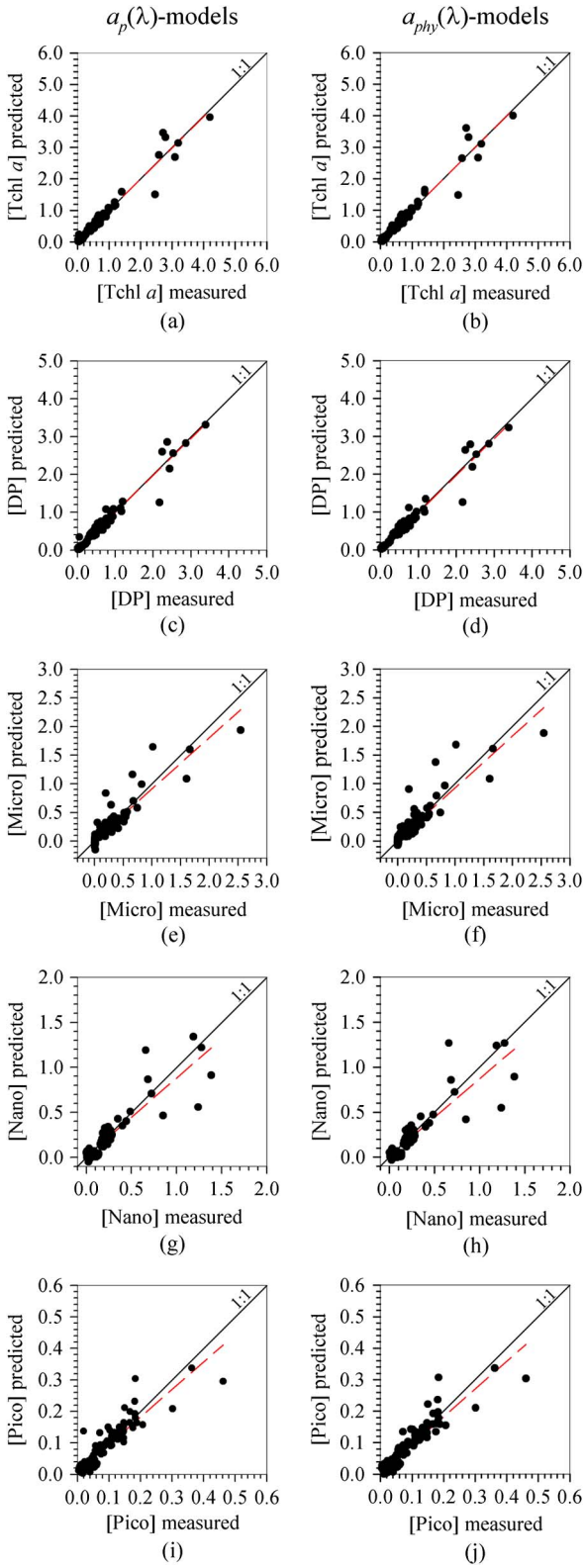

Fig. 5. (Color online) Cross-validated LOO predictions (in mg m^{-3}, $n = 239$) of the five variables ([Tchl *a*], DP, Micro, Nano, and Pico) as derived by the PLS models trained using HPLC pigment measurements and $a_p(\lambda)$ (left column) or $a_{phy}(\lambda)$ (right column) values included in the MedCAL data set (see Subsection 2.D for details) versus measured concentrations. The solid lines indicate the 1:1 ratio, the dashed lines show linear regressions between predicted and measured concentrations.

tion spectra (hereafter referred to as $a_p(\lambda)$- and $a_{phy}(\lambda)$-models) included in the MedCAL data set.

1. Selection of PLS-Models

The results of the PLS-models trained using the particle and phytoplankton absorption spectra included in the MedCAL data set are summarized in Table 2. The optimal number of components that minimized the error of prediction was different between the $a_p(\lambda)$- and $a_{phy}(\lambda)$-models and among the predicted variables. For the $a_p(\lambda)$-models, four PLS components were generally found to be optimal, as they explained more than 95% of the variance of the dependent variables, except for the picophytoplankton size class, which required six components (Table 2). In the case of the $a_{phy}(\lambda)$-models, a lower number of components was generally found to be optimal for the variables: two and three components were revealed to be sufficient for [Tchl a] and DP variables, while at least four PLS components were required to minimize the prediction error for the variables micro, nano, and pico (Table 2). These optimal numbers of components explained more than 95% of the variance both for the independent and dependent variables of all the five models (Table 2).

The cross-validated predictions for the $a_p(\lambda)$- and $a_{phy}(\lambda)$-models are shown in Figure 5 and the main parameters of the regression lines are reported in Table 2. In the plots showing predictions versus observations for the five variables (Fig. 5), predicted values are close to the 1:1 line, even if a high scatter can be observed for the cross predictions of the DPs associated with the phytoplankton size classes. All regression slopes (b) display values higher than 0.87 and 0.86 for the $a_p(\lambda)$- and $a_{phy}(\lambda)$-models, respectively. In the case of the [Tchl a] and DP variables, b values are the highest, close to 1. The determination coefficients (r^2) are high ($r^2 > 0.82$) for all variables, and they reach values up to 0.97 for the cross predictions of [Tchl a] and DP. The PLS $a_p(\lambda)$- and $a_{phy}(\lambda)$-models are therefore able to predict adequately all the variables used in this study, although the prediction accuracy is lower for the three variables associated with the algal size structure than it is for [Tchl a] and DP.

2. MedCAL-Trained Model Results

In this section, we compare the ability of the MedCAL trained PLS $a_p(\lambda)$- and $a_{phy}(\lambda)$-models (Table 2) in predicting the pigment concentrations and retrieving the algal size structure from the BOUSSOLE time series of particle and phytoplankton absorption spectra ($n = 484$). The parameters of linear regressions between predicted and measured pigment concentrations, the RMSEP and the BIAS values used to assess and compare the accuracy of the PLS models are reported in Table 3. Due to the large ranges of variation of [Tchl a] and pigment concentrations (three orders of magnitude), regressions between predicted and measured pigment concentrations are displayed in log–log scale in Fig. 6 for predictions obtained by $a_p(\lambda)$- and $a_{phy}(\lambda)$-models (left and right columns, respectively).

The most accurate predictions are obtained for [Tchl a] and the total DPs concentrations ($r^2 = 0.91$). All predicted values are close to the identity line (1:1) across the range of measured variables (Figs. 6(a)–6(d)), as shown by regression slopes close to 1 ($b > 0.98$, Table 3). More importantly, both $a_p(\lambda)$- and $a_{phy}(\lambda)$-models showed their ability in predicting the concentrations of the DPs associated with the micro, nano, and picophytoplankton size classes (Fig. 6). The predicted values are significantly correlated with the measured values ($r^2 > 0.52$) and the points are close to the identity line 1:1 as confirmed by the regression slopes ($b > 0.90$, Table 3). Analysis of the RMSEP and BIAS values reveals that the prediction accuracy is different among the variables but substantially unchanged between $a_p(\lambda)$- and $a_{phy}(\lambda)$-models (Table 3). Actually, both these models show varying prediction ability according to the pigment concentration. Indeed, the analysis in logarithm

Table 3. Statistical Parameters of Comparison between the HPLC Measured and PLS Pigment Concentrations Predicted by the $a_p(\lambda)$- and $a_{phy}(\lambda)$-Models Trained with the MedCAL Data Set and Tested on the BOUSSOLE Time Series ($n = 484$)[a]

	BOUSSOLE Prediction				
	r^2	b	a	RMSEP	BIAS
$a_p(\lambda)$ Models					
Tchl a	0.91	0.98	0.06	0.1690	0.0518
DP	0.91	1.03	0.04	0.1383	0.0510
Micro	0.75	0.91	0.06	0.1389	0.0477
Nano	0.66	0.98	0.04	0.1234	0.0378
Pico	0.54	0.94	0.01	0.0460	0.0039
$a_{phy}(\lambda)$ Models					
Tchl a	0.91	0.98	0.06	0.1681	0.0540
DP	0.91	1.02	0.05	0.1393	0.0550
Micro	0.75	0.90	0.04	0.1322	0.0297
Nano	0.65	0.97	0.04	0.1250	0.0355
Pico	0.52	0.93	0.01	0.0470	0.0030

[a]The various parameters are, from left to right: determination coefficient (r^2), regression slope (b), y-intercept (a), RMSEP (mg m^{-3}) and systematic error (BIAS, in mg m^{-3}).

scale of the regressions between predicted and measured concentrations (Fig. 6) shows a tendency of the models to underestimate very low (close to zero) concentrations, especially for [Tchl a], DP, nano, and pico. An opposite trend is observed for the lowest predicted fractions of microphytoplankton, which appear generally overestimated by both the $a_p(\lambda)$- (Fig 6(e)) and $a_{\mathrm{phy}}(\lambda)$- (Fig. 6(f)) models.

These observations suggest that in order to obtain an accurate retrieval of biomass and size structure of the algal communities at the BOUSSOLE site, both $a_p(\lambda)$- and $a_{\mathrm{phy}}(\lambda)$-models trained using the Mediterranean data set can be used interchangeably. It must be kept in mind, however, that only the particle absorption spectra were directly measured from seawater samples while the phytoplankton absorption spectra were computed by numerical decomposition [58]. Practically, the numerical decomposition leads to the estimation of the phytoplankton light absorption by the removal of an estimated contribution of the nonalgal particle (NAP) absorption represented with an exponential model. This exponential characteristic yields a fourth-derivative of NAP absorption characterized by exponential shape and magnitude close to zero, so that the fourth-derivative spectral features of particle and phytoplankton light absorption are very similar. Therefore, $a_{\mathrm{phy}}(\lambda)$-models might show higher performances than observed here if nonalgal absorption was measured instead of being estimated. However, the errors observed for predicted pigment concentrations in the clearest waters can be related to a reduction in efficiency of the fourth-derivative tool rather than to uncertainties in the $a_{\mathrm{phy}}(\lambda)$ estimation as these prediction errors were observed both for $a_p(\lambda)$ and $a_{\mathrm{phy}}(\lambda)$. This uncertainty in the pigment prediction may be driven by a reduced capability of the fourth-derivative analysis in highlighting the spectral absorption signatures of the DPs associated with a size class when close to zero. Another possible source of error is the presence, in the absorption spectrum of the algal community, of the signatures of non-taxonomic pigments. These signatures that are also present in the fourth-derivative absorption spectra, could actually overlap the absorption bands of DPs associated with size classes and alter, therefore, the correlation between the magnitude of the fourth derivative pigment band and the concentration of a pigment [60].

C. Retrieval of Phytoplankton Community Structure from the GLOCAL Data Set

Here we present the models trained using HPLC pigment and absorption data included in the GLOCAL data set and we discuss their prediction ability using the BOUSSOLE time series. As the previous results showed similar performances for the $a_p(\lambda)$- and $a_{\mathrm{phy}}(\lambda)$-models, we focus only on the results obtained from phytoplankton absorption spectra.

Four PLS components were found to be optimal for modeling and explaining ~95% of the variance

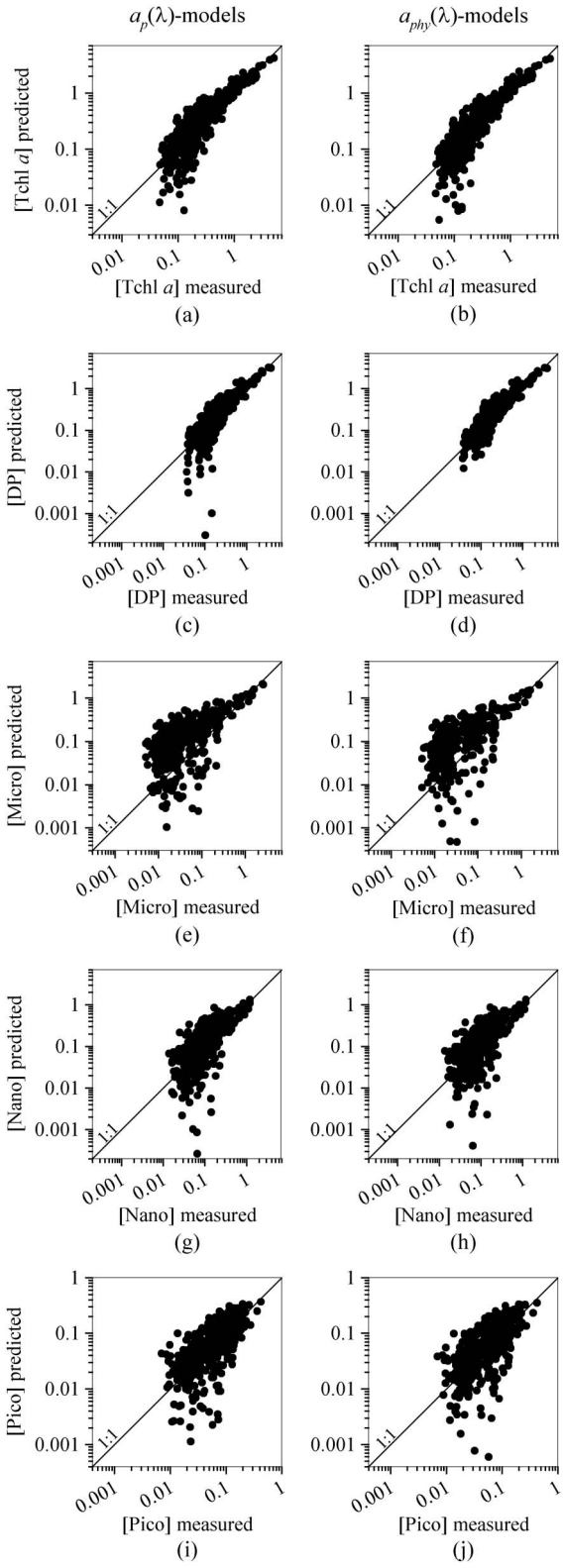

Fig. 6. Relationships between the predicted and measured concentrations (in mg m^{-3}) of the five variables (Tchl a, DP, Micro, Nano, and Pico) for the BOUSSOLE data set. A few predicted negative values are disregarded. Pigment concentrations are predicted by the PLS models trained using HPLC pigment measurements and $a_p(\lambda)$ (left column) or $a_{\mathrm{phy}}(\lambda)$ values (right column) included in the MedCAL data set. The 1:1 ratio is shown as a solid line.

Table 4. PLS Parameters of $a_{phy}(\lambda)$-Models Trained Using HPLC Pigment Measurements and Absorption Spectral Values Included in the GLOCAL Data Set ($n = 716$), from Left to Right: Number of Components (N), RMSEP (mg m^{-3}), Explained Variance (%) for Independent [r^2X (%)] and Dependent [r^2Y (%)] Variables[a]

		LOO Prediction					
$a_{phy}(\lambda)$ Models	N	RMSEP	r^2X (%)	r^2Y (%)	r^2	b	a
Tchl a	4	0.1145	88.88	94.96	0.94	0.94	0.02
DP	4	0.1025	89.32	95.14	0.94	0.95	0.02
Micro	7	0.0813	93.59	94.73	0.93	0.94	0.01
Nano	8	0.0618	94.85	91.74	0.89	0.91	0.01
Pico	8	0.0306	95.08	80.15	0.76	0.77	0.02

[a]Statistical parameters for linear regressions between leave-one-out (LOO) predicted and measured pigment concentrations: determination coefficient (r^2), regression slope (b) and y-intercept (a).

of [Tchl a] and DP (Table 4). A high number of components (seven at least) was required to minimize the prediction error of the three variables (micro, nano, and pico) associated with the phytoplankton size classes and to account for more than 80% of the variance of the data set (Table 4). Similar to the $a_{phy}(\lambda)$-models trained with the MedCAL data set, the cross-validated predictions (Fig. 7; Table 4) showed high determination coefficients and regression slopes, which are, respectively characterized by values higher than 0.89 and 0.91 (except for picophytoplankton where they are lower, 0.76 and 0.77).

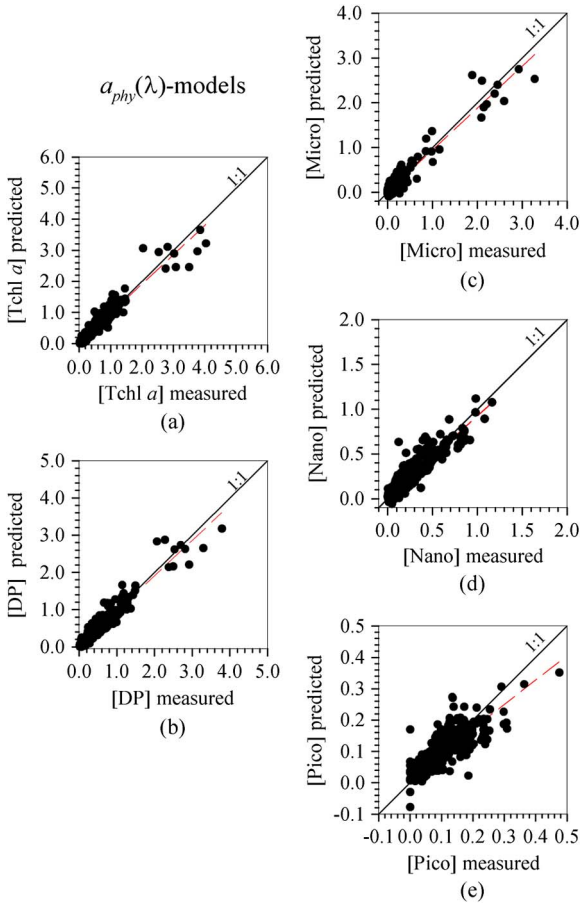

Fig. 7. (Color online) Cross-validated predictions (in mg m^{-3}; $n = 716$) of the 5 variables ([Tchl a], DP, Micro, Nano, and Pico) versus measured concentrations. LOO predictions result from the PLS models trained with HPLC pigment concentrations and $a_{phy}(\lambda)$ values included in the GLOCAL data set (see Subsection 2.D for details). The solid lines indicate the 1:1 ratio, the dashed lines show linear regressions between predicted and measured concentrations.

As for the MedCAL PLS models, the models trained with the GLOCAL data set and tested on the BOUSSOLE data also showed a good capability in predicting the algal biomass and total DPs content at the BOUSSOLE site (Table 5). The RMSEP and BIAS values reveal that the accuracy of [Tchl a] predicted by the GLOCAL PLS-model (Table 5) is very similar to that observed for the MedCAL one (Table 3). However, the prediction of DP is slightly more accurate and less biased when the MedCAL PLS-model is used instead of the GLOCAL one. More importantly, the GLOCAL PLS models are less efficient in retrieving the size structure of the BOUSSOLE algal communities (Fig. 8) than MedCAL PLS-models. The predicted values are actually correlated with the measured values ($r^2 > 0.42$, Table 5), but the predictions are systematically overestimated for microphytoplankton (Fig. 8(c)) and underestimated for the nano and picophytoplankton size classes (Figs. 8(d) and 8(e)).

As the PLS models utilize the spectral signatures of DPs to retrieve their concentrations, one would expect that the signature of a pigment does not vary regionally, so that the PLS could perform similarly regardless of the location of the data used for the training. In order to explore this issue, a comparison between the fourth-derivative absorption spectra collected at the BOUSSOLE site and those sampled from the Mediterranean Sea, the Atlantic and Pacific Oceans has been performed. For each location, we split the fourth-derivative spectra into four groups according to the level of biomass ([Tchl a] in mg m^{-3}: Tchl $a \leq 0.1$; $0.1 <$ Tchl $a \leq 0.5$; $0.5 <$ Tchl $a \leq 1$; Tchl $a > 1$) and then we compared the respective averages of fourth-derivative spectra within each level. As the results are similar for all the biomass levels, only the fourth-derivative absorption spectra for Tchl $a \leq 0.1$ mg m^{-3} and $0.1 <$ Tchl $a \leq 0.5$ mg m^{-3} are shown as examples (Fig. 9).

These comparisons show that the amplitude and position of the bands of the fourth-derivative

Table 5. Statistical Parameters of Comparison between the HPLC Measured and PLS Pigment Concentrations Predicted by the $a_{phy}(\lambda)$-Models Trained with the GLOCAL Data Set and Tested on the BOUSSOLE Time Series ($n = 484$)[a]

	BOUSSOLE Prediction				
$a_{phy}(\lambda)$ Models	r^2	b	a	RMSEP	BIAS
Tchl a	0.91	1.01	0.05	0.1669	0.0565
DP	0.93	1.08	0.04	0.1402	0.0660
Micro	0.70	1.18	0.12	0.2353	0.1367
Nano	0.48	0.44	0.04	0.1266	−0.0358
Pico	0.42	0.60	0.01	0.0440	−0.0100

[a]The various parameters are, from left to right: determination coefficient (r^2), regression slope (b), y-intercept (a), RMSEP (mg m^{-3}) and systematic error (BIAS, in mg m^{-3}).

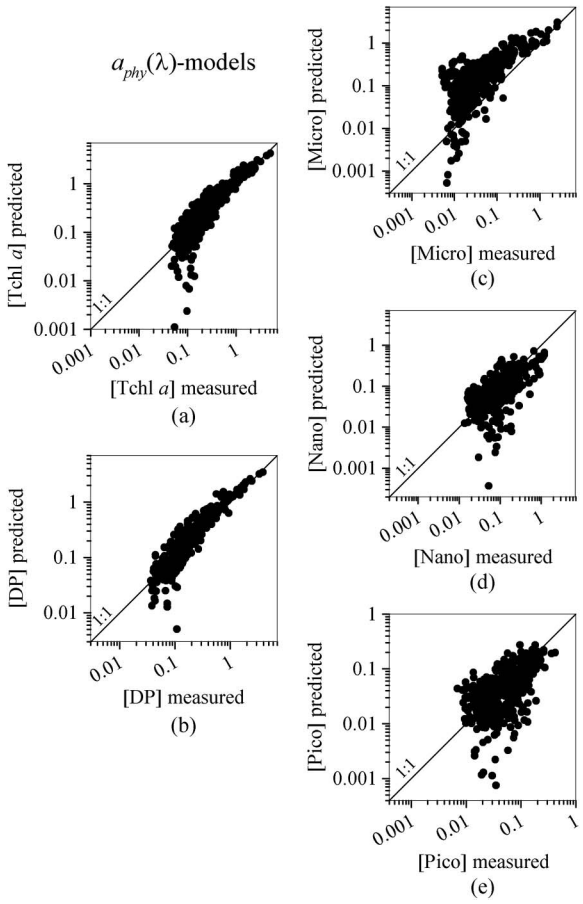

Fig. 8. Comparison between the predicted and measured concentrations (in mg m^{-3}) of the five variables (Tchl a, DP, Micro, Nano, and Pico) for the BOUSSOLE data set. A few predicted negative values are disregarded. Predicted concentrations are obtained by the PLS models trained using HPLC pigment measurements and $a_{phy}(\lambda)$ values included in the GLOCAL data set. The 1:1 ratio is shown as a solid line.

absorption peaks for BOUSSOLE are close to those of Mediterranean samples (Figs. 9(a) and 9(b)), whereas they reveal differences with those from the Atlantic and Pacific Oceans [Figs. 9(c)–9(f)]. We observed a shift of the pigment absorption bands to higher or lower wavelengths in the Atlantic and Pacific Oceans, respectively, compared to Mediterranean data: this effect can be essentially attributed to the displacement of the band center of a given pigment that can occur between different algal groups [66], possibly as a result of different interactions between pigments and proteins. More importantly, we observed a remarkable variation in the amplitudes of pigment absorption bands between the Mediterranean data (BOUSSOLE included) those from other regions (Fig. 9). At the first-order, the amplitudes of these bands are ruled, for a given sample, by the concentrations of the various pigments. However, they are also driven by the variations in algal size and intracellular pigment concentrations, which occur even within a narrow chlorophyll range: for instance, the ultra-oligotrophic waters collected in the Pacific Ocean (BIOSOPE cruise) are characterized by larger algal cells in comparison to other areas with similar chlorophyll ranges [41], which leads to a higher package effect and lower absorption bands per unit of pigment concentration (Fig. 9). In addition, the level of package effect is influenced by the incident irradiance, as the photoacclimation state of algal cells rules their intracellular pigment concentration. The variety of locations and sampling periods actually emphasizes these physiological variations in the phytoplankton populations, and consequently yields modifications in the spectral absorption characteristics. Therefore training the PLS-models with regional data sets actually reduces these sources of variability, leading to a more accurate retrieval of the algal size structure.

D. Comparison between HPLC- and PLS-Derived Variations over the BOUSSOLE Time Series

Temporal variations of chlorophyll a concentration [Tchl a], the total DPs and the concentrations of DPs associated with the three phytoplankton size classes (micro, nano, and picophytoplankton) as derived from HPLC measurements and from the MedCAL PLS models (see Table 2) are displayed in Fig. 10. The model-predicted concentrations of the different variables well reproduce those obtained from HPLC pigment measurements over the entire BOUSSOLE time series.

More importantly, even the short-term and seasonal fluctuations of algal biomass and size classes as retrieved from the nine-year series of HPLC pigment measurements are well reproduced by the MedCAL trained PLS-models (Fig. 11). As $a_p(\lambda)$ and $a_{phy}(\lambda)$ PLS models showed similar performances,

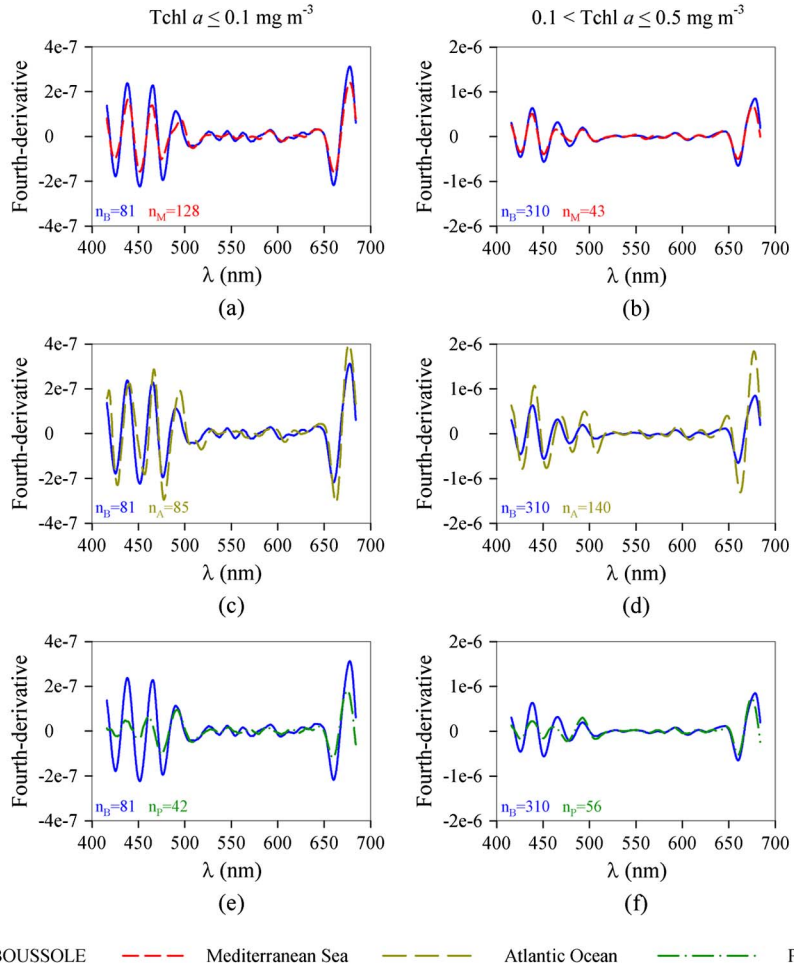

Fig. 9. (Color online) Comparison between fourth-derivatives of phytoplankton absorption spectra sampled at the BOUSSOLE site and those from: (a), (b) the Mediterranean Sea; (c), (d) the Atlantic Ocean; (e), (f) the Pacific Ocean. For each area, the averages of all samples with Tchl $a \leq 0.1$ mg m^{-3} (left column) and of all samples with $0.1 <$ Tchl $a \leq 0.5$ mg m^{-3} (right column) are reported. n_B, n_M, n_A, and n_P are the number of spectra used to calculate the average spectrum for each region (BOUSSOLE, the Mediterranean Sea, the Atlantic Ocean, and the Pacific Ocean, respectively).

Fig. 11 displays the annual cycle retrieved from HPLC pigment analysis and from the $a_p(\lambda)$ PLS models only. The algal biomass shows relatively marked seasonal variations at the BOUSSOLE site (Figs. 11(a) and 11(b)). According to previous observations in the Mediterranean Sea [64,67,68], maximal [Tchl a] occurs at the end of winter and during spring. The spring phytoplankton bloom starts generally at the beginning of March and ends at the end of May and the maximal [Tchl a] concentrations occur from mid-March to mid-April. The period from June to the beginning of October is generally characterized by very low concentrations of [Tchl a], while a progressive increase can be observed in wintertime. The absolute concentrations of DP and pigments associated with size classes change in accordance with the algal biomass (Fig. 11). However, some seasonal divergences between the three size classes can be observed. For example, microphytoplankton is strongly present at the BOUSSOLE site especially from the end of winter to the end of spring (Figs. 11(e) and 11(f)). Its maximal occurrence is between mid-March and mid-April during the spring bloom, then it decreases to very low concentrations during the rest of the year. The absolute abundances of nano and picophytoplankton generally follow the seasonal trend of the biomass (Fig. 11). After a recurrent maximal abundance in late winter and early spring, a significant increase can be observed in summer and from October to December.

Discrepancies are, however, observed in some instances between the pigment predictions and observations. This is particularly the case of nano and picophytoplankton (Fig. 11). Although the HPLC-measured seasonal fluctuations of these two algal classes are fully reproduced by PLS, their concentrations are on several occasions largely overestimated by the model during winter. Such an overestimation is also evidenced for microphytoplankton from June to December (Figs. 11(e) and 11(f)). However, it must be kept in mind that at this time of the year the c oncentrations of pigments associated with microphytoplankton are generally close to zero and, therefore, as discussed in the previous sections, the PLS

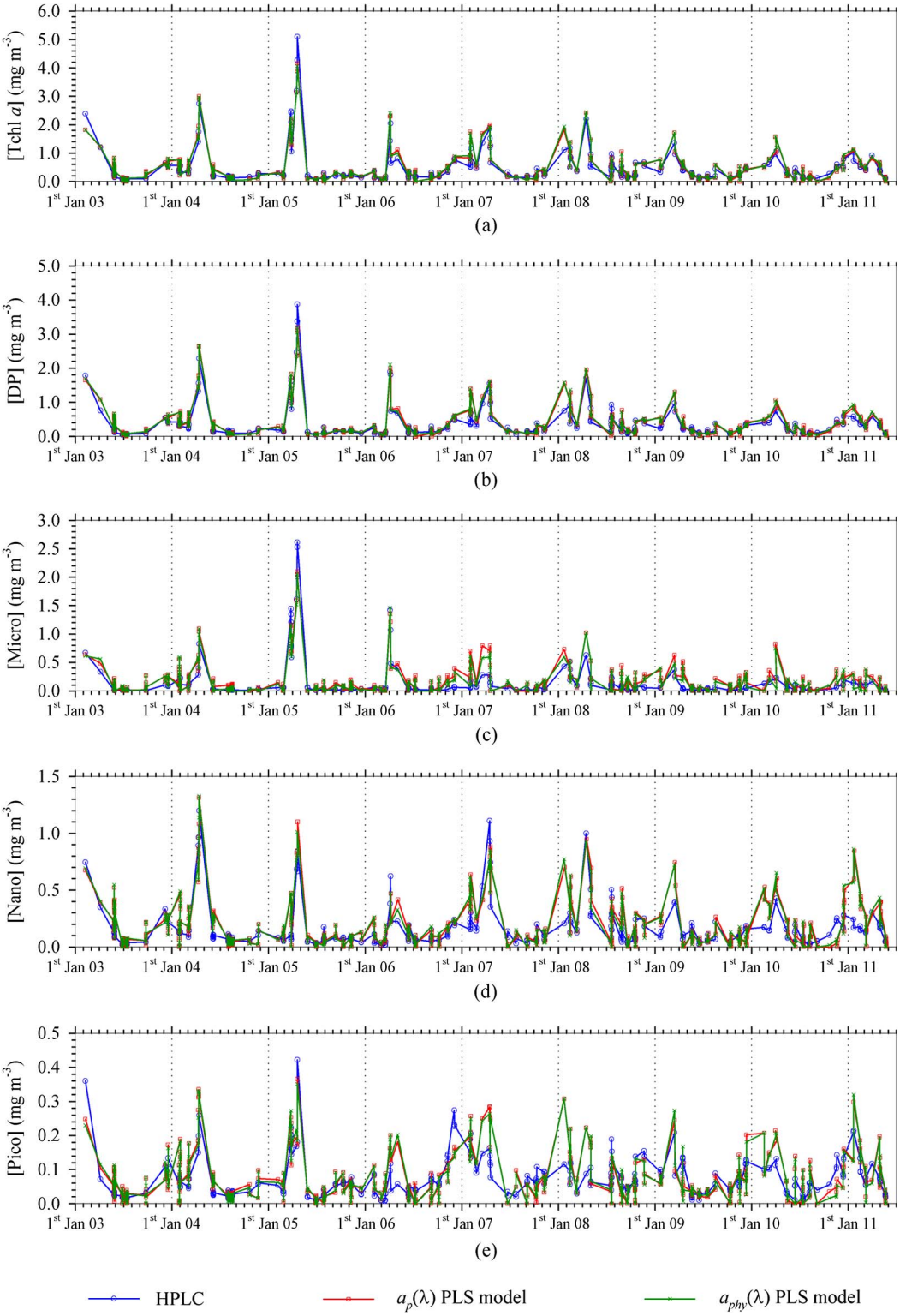

Fig. 10. (Color online) Entire BOUSSOLE time series (January 2003–May 2011) of pigment concentrations as derived from HPLC pigment measurements (blue line) and from PLS models trained using $a_p(\lambda)$ (red line) or $a_{\mathrm{phy}}(\lambda)$ (green line) values included in the MedCAL data set. A few predicted negative values are replaced by zero. The plot shows the time series for: (a) [Tchl a], (b) DP, (c) Micro, (d) Nano, and (e) Pico.

technique coupled with the fourth-derivative analysis of absorption spectra shows a lower prediction accuracy than for higher concentrations. In spite of this, the consistency between the seasonal and annual evolutions of algal biomass and size classes retrieved from PLS-models and HPLC pigment measurements emphasizes the potential of the PLS models presented for the retrieval and analysis of the temporal changes

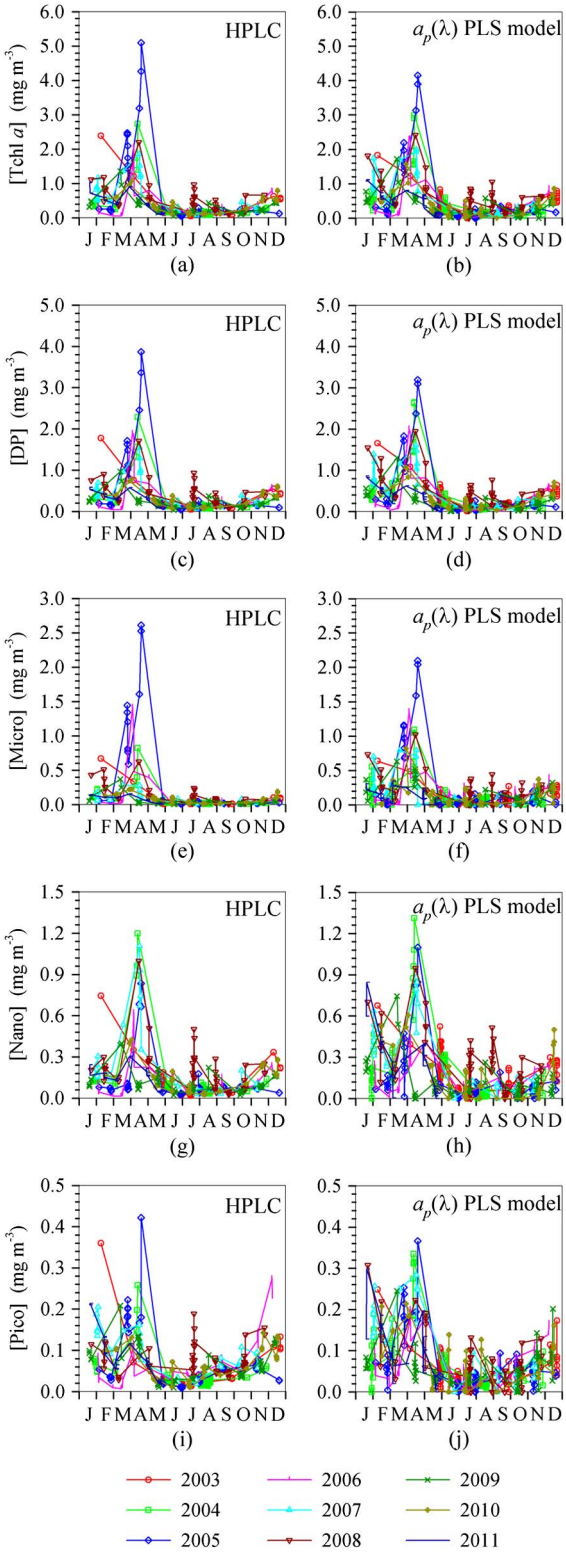

Fig. 11. (Color online) Annual cycle of [Tchl a], DP, Micro, Nano, and Pico concentrations, over the period January 2003–May 2011, as derived from: HPLC pigment measurements (left column) and $a_p(\lambda)$-PLS models trained with the MedCAL data set (right column). A few predicted negative values are replaced by zero.

in the phytoplankton community structure using particle or phytoplankton light absorption, especially in absence of HPLC measurements.

4. Conclusions

The retrieval of algal biomass and size structure from *in vivo* hyperspectral absorption measurements can be achieved by application of the multivariate PLS regression technique. As expected, PLS models trained using a regional data set, including data from the Mediterranean Sea only (MedCAL), provided the best prediction over the BOUSSOLE time series.

Satisfactory [Tchl a] and DP predictions emerged also from PLS models trained using a data set assembled from various locations in the world's oceans (GLOCAL). However, the retrieval of size classes by these models was less efficient because of the larger variability in band position and amplitude observed between the fourth-derivative spectra of the Mediterranean communities (BOUSSOLE included) in comparison with those of the Atlantic and Pacific Oceans. In spite of this, we suggest that future works should test the performances of such models on data from different locations in the world's oceans rather than from a single site. So, the applicability of GLOCAL PLS models in detecting phytoplankton size classes could effectively be evaluated.

The prediction abilities of the $a_p(\lambda)$- and $a_{\mathrm{phy}}(\lambda)$-models are very similar. However, it must be kept in mind that phytoplankton absorption spectra were obtained by numerical decomposition [58] in the present study. This suggests that better performances on the retrieval of the algal size structure might be achieved by PLS if measured phytoplankton absorption spectra (e.g., using chemical pigment extraction from filters [59]) are used. The use of the particle absorption measurements has the advantage (compared to the HPLC pigment analysis) that continuous profiling systems for measuring *in situ* hyperspectral absorption are becoming accessible (e.g., HOBILABS a-sphere and WET Labs ac-s). This, actually, leads to a faster retrieval of the algal size structure and to the possibility to detect the phytoplankton community structure with a fine vertical resolution within the water column. Nevertheless, HPLC pigment analysis remains indispensable for the validation of results.

In addition, the similar performances of the PLS technique for particle and phytoplankton absorption trained-models actually emphasize that such a technique could be applied to the absorption coefficients as inverted from AOPs such as the reflectance or the remote sensing reflectance. Nevertheless, the application of the PLS to IOPs derived from satellite ocean-color observations might largely depend on the uncertainties of the retrievals driven by inaccurate radiometric and atmospheric corrections as error sources [43] or more importantly on the limited availability of hyperspectral imagery. In regard to this, hyperspectral sensors have been recently launched onboard satellites (HICO, see [69,70]) or are planned in the near future (e.g., NASA's PACE mission, www.decadal.gsfc.nasa.gov).

The advantage of using hyperspectral data for increasing the accuracy of the taxonomic and algal size structure retrievals is intuitive and has been evidenced in several studies [22,23,71–73]. In addition, derivative analysis [60] provides a deep evaluation of the smallest variations in the spectral shape of hyperspectral IOPs and AOPs. Despite a basic noise that may occur during the measurement process, small spectral modifications are induced by the variations in pigment composition and concentration, and thus in the taxonomic and size composition of the algal communities [66,74]. Therefore, the efforts pursued to improve the retrieval of IOPs from hyperspectral reflectance data (e.g., BIOCAREX and BOUSSOLE projects) actually increase the chance to achieve a more accurate retrieval of the phytoplankton community structure and use the PLS method as an effective tool for monitoring continuously the changes in the algal community structure.

This study is a contribution to the BIOCAREX project, which was funded by the Agence Nationale de la Recherche (ANR), and to the BOUSSOLE project. Multiple organizations funded the BOUSSOLE project and provided technical and logistic support: European Space Agency (ESA), Centre National d'Etudes Spatiales (CNES), Centre National de la Recherche Scientifique (CNRS), National Aeronautics and Space Administration (NASA), Institut National des Sciences de l'Univers (INSU), Université Pierre et Marie Curie (UPMC), Observatoire Océanologique de Villefranche sur Mer (OOV). The authors are grateful to the members of the BOUSSOLE technical staff (http://www.obs-vlfr.fr/Boussole/html/people/tech_staff.php) for lab analyses and monthly cruises development, and to the captains and crews of the research vessels (*Téthys-II*, *Le Suroît, Antea, Europe*) for ship measurements and sampling. The training data set was previously acquired in the frame of several projects funded by the PROOF French program (EUMELI, EPOPE, FRONTAL, PROSOPE, POMME) and the LEFE-CYBER French program (BIOSOPE).

References

1. J. A. Raven and P. G. Falkowski, "Ocean sink for atmospheric CO_2," Plant Cell Environ. **22**, 741–755 (1999).
2. O. Aumont, E. Meier-Reimer, S. Blain, and P. Monfray, "An ecosystem model of the global ocean including Fe, Si, P colimitations," Glob. Biogeochem. Cycles **17**, 23–29 (2003).
3. C. Le Quéré, S. P. Harrison, I. C. Prentice, E. T. Buitenhuis, O. Aumont, L. Bopp, H. Claustre, L. Cotrim da Cunha, R. Geider, X. Giraud, C. Klaas, K. E. Kohfeld, L. Legendre, M. Manizza, T. Platt, R. B. Rivkin, S. Sathyendranath, J. Uitz, A. J. Watson, and D. Wolf-Gladrow, "Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models," Glob. Chang. Biol. **11**, 2016–2040 (2005).
4. R. R. Hood, E. A. Laws, R. A. Armstrong, N. R. Bates, C. W. Brown, C. A. Carlson, F. Chai, S. C. Doney, P. G. Falkowski, R. A. Feely, M. A. Friedrichs, M. R. Landry, J. K. Moore, D. M. Nelson, T. L. Richardson, B. Salihoglu, M. Schartau, D. A. Toole, and J. D. Wiggert, "Pelagic functional group modeling: progress, challenges and prospects," Deep Sea Res. Part II **53**, 459–512 (2006).
5. T. Platt, S. Sathyendranath, and V. Stuart, "Why study biological oceanography?" Aquabiology **28**, 542–557 (2006).
6. T. Platt and S. Sathyendranath, "Ecological indicators for the pelagic zone of the ocean from remote sensing," Remote Sens. Environ. **112**, 3426–3436 (2008).
7. S. Sathyendranath and T. Platt, "Ocean-colour radiometry: achievements and future perspectives," in *Oceanography from Space-Revisited*, V. Barale, J. F. R. Gower, and L. Alberotanza, eds., (Springer, 2010), pp. 349–359.
8. S. Alvain, C. Moulin, Y. Dandonneau, and F. M. Bréon, "Remote sensing of phytoplankton groups in case 1 waters from global SeaWiFS imagery," Deep Sea Res. Part I **52**, 1989–2004 (2005).
9. J. Aiken, J. R. Fishwick, S. Lavender, R. Barlow, G. F. Moore, H. Sessions, S. Bernard, J. Ras, and N. J. Hardman-Mountford, "Validation of MERIS reflectance and chlorophyll during the BENCAL cruise October 2002: preliminary validation of new demonstration products for phytoplankton functional types and photosynthetic parameters," Int. J. Remote Sens. **28**, 497–516 (2007).
10. D. E. Raitsos, S. J. Lavender, C. D. Maravelias, J. Haralabous, A. J. Richardson, and P. C. Reid, "Identifying four phytoplankton functional types from space: an ecological approach," Limnol. Oceanogr. **53**, 605–613 (2008).
11. J. Uitz, H. Claustre, A. Morel, and S. B. Hooker, "Vertical distribution of phytoplankton communities in open ocean: an assessment based on surface chlorophyll," J. Geophys. Res. **111**, C08005 (2006).
12. E. Devred, S. Sathyendranath, V. Stuart, H. Maas, O. Ulloa, and T. Platt, "A two-component model of phytoplankton absorption in the open ocean: theory and applications," J. Geophys. Res. **111**, C03011 (2006).
13. T. Hirata, J. Aiken, N. J. Hardman-Mountford, T. J. Smyth, and R. G. Barlow, "An absorption model to determine phytoplankton size classes from satellite ocean colour," Remote Sens. Environ. **112**, 3153–3159 (2008).
14. 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. Model. **221**, 1472–1483 (2010).
15. A. Fujiwara, T. Hirawake, K. Suzuki, and S. I. Saitoh, "Remote sensing of size structure of phytoplankton communities using optical properties of the Chukchi and Bering Sea shelf region," Biogeosciences **8**, 3567–3580 (2011).
16. A. M. Ciotti and A. Bricaud, "Retrievals of a size parameter for phytoplankton and spectral light absorption by colored detrital matter from water leaving radiances at SeaWiFS channels in a continental shelf region off Brazil," Limnol. Oceanogr. Methods **4**, 237–253 (2006).
17. C. B. Mouw and J. A. Yoder, "Optical determination of phytoplankton size composition from global SeaWiFS imagery," J. Geophys. Res. **115**, C12018 (2010).
18. C. B. Mouw, J. A. Yoder, and S. C. Doney, "Impact of phytoplankton community size on a linked global ocean optical and ecosystem model," J. Mar. Syst. **89**, 61–75 (2012).
19. T. S. Kostadinov, D. A. Siegel, and S. Maritorena, "Retrieval of the particle size distribution from satellite ocean color observations," J. Geophys. Res. **114**, C09015 (2009).
20. T. S. Kostadinov, D. A. Siegel, and S. Maritorena, "Global variability of phytoplankton functional types from space: assessment via the particle size distribution," Biogeosciences **7**, 3239–3257 (2010).
21. 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, "Synoptic relationships between surface chlorophyll-*a* and diagnostic pigments specific to phytoplankton functional types," Biogeosciences **8**, 311–327 (2011).
22. J. R. Moisan, T. A. H. Moisan, and M. A. Linkswiler, "An inverse modeling approach to estimating phytoplankton pigment concentrations from phytoplankton absorption spectra," J. Geophys. Res. **116**, C09018 (2011).
23. E. Torrecilla, D. Stramski, R. A. Reynolds, E. Millán-Núñez, and J. Piera, "Cluster analysis of hyperspectral optical data for discriminating phytoplankton pigment assemblages in

the open ocean," Remote Sens. Environ. **115**, 2578–2593 (2011).
24. J. M. Sieburth, V. Smetacek, and J. Lenz, "Pelagic ecosystem structure: heterotrophic compartments of the plankton and their relationship to plankton size fractions," Limnol. Oceanogr. **23**, 1256–1263 (1978).
25. T. Platt, C. Fuentes-Yaco, and K. T. Frank, "Spring algal bloom and larval fish survival," Nature **423**, 398–399 (2003).
26. C. Fuentes-Yaco, P. A. Koeller, S. Sathyendranath, and T. Platt, "Shrimp (*Pandalus borealis*) growth and timing of the spring phytoplankton bloom on the Newfoundland-Labrador shelf," Fish. Ocean. **16**, 116–129 (2007).
27. A. Nair, S. Sathyendranath, T. Platt, J. Morales, V. Stuart, M. H. Forget, E. Devred, and H. Bouman, "Remote sensing of phytoplankton functional types," Remote Sens. Environ. **112**, 3366–3375 (2008).
28. C. S. Reynolds, *The Ecology of Phytoplankton* (Cambridge University, 2006).
29. A. Morel and A. Bricaud, "Theoretical results concerning light absorption in a discrete medium, and application to specific absorption of phytoplankton," Deep Sea Res. Part I **28**, 1375–1393 (1981).
30. A. Bricaud, H. Claustre, J. Ras, and K. Oubelkheir, "Natural variability of phytoplanktonic absorption in oceanic waters: influence of the size structure of algal populations," J. Geophys. Res. **109**, C11010 (2004).
31. 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**, 404–417 (2002).
32. R. J. W. Brewin, N. J. Hardman-Mountford, S. J. Lavender, D. E. Raitsos, T. Hirata, J. Uitz, E. Devred, A. Bricaud, A. Ciotti, and B. Gentili, "An intercomparison of bio-optical techniques for detecting dominant phytoplankton size class from satellite remote sensing," Remote Sens. Environ. **115**, 325–339 (2011).
33. H. Martens and T. Næs, *Multivariate Calibration* (Wiley, 1989).
34. K. H. Esbensen, T. Midtgaard, and S. Schonkopf, *Multivariate Analysis in Practice* (Wennberg, 1994).
35. L. Moberg, B. Karlberg, S. Blomqvist, and U. Larsson, "Comparison between a new application of multivariate regression and current spectroscopy methods for the determination of chlorophylls and their corresponding pheopigments," Anal. Chim. Acta **411**, 137–143 (2000).
36. J. Seppäla and K. Olli, "Multivariate analysis of phytoplankton spectral in vivo fluorescence: estimation of phytoplankton biomass during a mesocosm study in the Baltic Sea," Mar. Ecol. Prog. Ser. **370**, 69–85 (2008).
37. L. Moberg, B. Karlberg, K. Sørensen, and T. Källqvist, "Assessment of phytoplankton class abundance using absorption spectra and chemometrics," Talanta **56**, 153–160 (2002).
38. P. A. Stæhr and J. J. Cullen, "Detection of *Karenia mikomotoi* by spectral absorption signatures," J. Plankton Res. **25**, 1237–1249 (2003).
39. R. Martínez-Guijarro, I. Romero, M. Pachés, J. G. del Río, C. M. Martí, G. Gil, A. Ferrer-Riquielme, and J. Ferrer, "Determination of phytoplankton composition using absorption spectra," Talanta **78**, 814–819 (2009).
40. IOCCG, "Remote sensing of inherent optical properties: fundamentals, tests of algorithms, and applications," Reports of the International Ocean Colour Coordinating Group, No. 5, Z.-P. Lee, ed. (IOCCG, 2006).
41. A. Bricaud, M. Babin, H. Claustre, J. Ras, and F. Tièche, "Light absorption properties and absorption budget of South East Pacific waters," J. Geophys. Res. **115**, C08009 (2010).
42. D. Antoine, M. Chami, H. Claustre, F. D'Ortenzio, A. Morel, G. Bécu, B. Gentili, F. Louis, J. Ras, E. Roussier, A. J. Scott, D. Tailliez, S. B. Hooker, P. Guevel, J. F. Desté, C. Dempsey, and D. Adams, "BOUSSOLE: a joint CNRS-INSU, ESA, CNES and NASA ocean color calibration and validation activity," NASA Technical Memorandum No. 2006-214147 (2006).
43. D. Antoine, F. D'Ortenzio, S. B. Hooker, G. Bécu, B. Gentili, D. Tailliez, and A. J. Scott, "Assessment of uncertainty in the ocean reflectance determined by three satellite ocean color sensors (MERIS, SeaWiFS, and MODIS-A) at an offshore site in the Mediterranean Sea (BOUSSOLE project)," J. Geophys. Res. **113**, C07013 (2008).
44. A. Morel and L. Prieur, "Analysis of variations in ocean color," Limnol. Oceanogr. **22**, 709–722 (1977).
45. H. R. Gordon and W. R. Mc Cluney, "Estimation of the depth of sunlight penetration in the sea for remote sensing," Appl. Opt. **14**, 413–416 (1975).
46. A. Morel and S. Maritorena, "Bio-optical properties of oceanic waters: a reappraisal," J. Geophys. Res. **106**, 7163–7180 (2001).
47. H. Claustre and J. C. Marty, "Specific phytoplankton biomasses and their relation to primary production in the tropical North Atlantic," Deep Sea Res. Part I **42**, 1475–1493 (1995).
48. J. Ras, J. Uitz, and H. Claustre, "Spatial variability of phytoplankton pigment distributions in the subtropical South Pacific Ocean: comparison between in situ and modeled data," Biogeosciences **5**, 353–369 (2008).
49. F. Vidussi, H. Claustre, J. Bustillos-Guzmàn, C. Cailliau, and J. C. Marty, "Determination of chlorophylls and carotenoids of marine phytoplankton: separation of chlorophyll *a* from divinyl-chlorophyll *a* and zeaxanthin from lutein," J. Plankton Res. **18**, 2377–2382 (1996).
50. F. Vidussi, H. Claustre, B. B. Manca, A. Luchetta, and J. C. Marty, "Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter," J. Geophys. Res. **106**, 19939–19956 (2001).
51. J. Aiken, Y. Pradhan, R. Barlow, S. Lavender, A. Poulton, P. Holligan, and N. Hardman-Mountford, "Phytoplankton pigments and functional types in the Atlantic Ocean: a decadal assessment, 1995–2005," Deep Sea Res. Part II **56**, 899–917 (2009).
52. J. Uitz, Y. Huot, F. Bruyant, M. Babin, and H. Claustre, "Relating phytoplankton photophysiological properties to community structure on large scale," Limnol. Oceanogr. **53**, 614–630 (2008).
53. J. Uitz, H. Claustre, N. Garcia, F. B. Griffiths, J. Ras, and V. Sandroni, "A phytoplankton class-specific primary production model applied to the Kerguelen islands region (Southern Ocean)," Deep Sea Res. Part I **56**, 541–560 (2009).
54. E. Organelli, C. Nuccio, C. Melillo, and L. Massi, "Relationships between phytoplankton light absorption, pigment composition and size structure in offshore areas of the Mediterranean Sea," Adv. Oceanogr. Limnol. **2**, 107–123 (2011).
55. K. Allali, A. Bricaud, M. Babin, A. Morel, and P. Chang, "A new method for measuring spectral absorption coefficients of marine particles," Limnol. Oceanogr. **40**, 1526–1532 (1995).
56. A. Bricaud, A. Morel, M. Babin, K. Allali, and H. Claustre, "Variations of light absorption by suspended particles with chlorophyll *a* concentration in oceanic (case 1) waters: analysis and implications for bio-optical models," J. Geophys. Res. **103**, 31033–31044 (1998).
57. K. Allali, A. Bricaud, and H. Claustre, "Spatial variations in the chlorophyll-specific absorption coefficients of phytoplankton and photosynthetically active pigments in the equatorial Pacific," J. Geophys. Res. **102**, 12413–12423 (1997).
58. A. Bricaud and D. Stramski, "Spectral absorption coefficients of living phytoplankton and nonalgal biogenous matter: a comparison between Peru upwelling area and Sargasso Sea," Limnol. Oceanogr. **35**, 562–582 (1990).
59. M. Kishino, M. Takahashi, N. Okami, and S. Ichimura, "Estimation of the spectral absorption coefficients of phytoplankton in the sea," Bull. Mar. Sci. **37**, 634–642 (1985).
60. R. Bidigare, J. Morrow, and D. Kiefer, "Derivative analysis of spectra absorption by photosynthetic pigments in the western Sargasso Sea," J. Mar. Res. **47**, 323–341 (1989).
61. F. Tsai and W. Philpot, "Derivative analysis of hyperspectral data," Remote Sens. Environ. **66**, 41–51 (1998).
62. P. Geladi and B. R. Kowalski, "Partial least squares regression: a tutorial," Anal. Chim. Acta **185**, 1–17 (1986).
63. B. H. Mevik and R. Wehrens, "The pls package: principal component and partial least squares regression in *R*," J. Stat. Softw. **18**, 1–24 (2007).

64. J. C. Marty, J. Chiavérini, M. D. Pizay, and B. Avril, "Seasonal and interannual dynamics of nutrients and phytoplankton pigments in the western Mediterranean Sea at the DYFAMED time-series station (1991–1999)," Deep Sea Res. Part II **49**, 1965–1985 (2002).
65. J. C. Marty and J. Chiavérini, "Hydrological changes in the Ligurian Sea (NW Mediterranean, DYFAMED site) during 1995–2007 and biogeochemical consequences," Biogeosciences **7**, 2117–2128 (2010).
66. N. Hoepffner and S. Sathyendranath, "Effect of pigment composition on absorption properties of phytoplankton," Mar. Ecol. Prog. Ser. **73**, 11–23 (1991).
67. E. Bosc, A. Bricaud, and D. Antoine, "Seasonal and interannual variability in algal biomass and primary production in the Mediterranean Sea, as derived from four years of SeaWiFS observations," Glob. Biogeochem. Cycles **18**, GB1005 (2004).
68. V. Barale, J. M. Jacquet, and M. Ndiaye, "Algal blooming patterns and anomalies in the Mediterranean Sea as derived from the SeaWifs data set (1998–2003)," Remote Sens. Environ. **112**, 3300–3313 (2008).
69. M. R. Corson, D. R. Korwan, R. L. Lucke, W. A. Snyder, and C. O. Davis, "The hyperspectral imager for the coastal ocean (HICO) on the international space station," in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium* (IEEE, 2008), pp. 101–104.
70. R. L. Lucke, M. R. Corson, N. McGlothlin, S. Butcher, D. Wood, D. R. Korwan, R. Li, W. A. Snyder, C. O. Davis, and D. Chen, "Hyperspectral imager for the coastal ocean: instrument description and first images," Appl. Opt. **50**, 1501–1516 (2011).
71. D. F. Millie, O. M. Schofield, G. J. Kirkpatrick, G. Johnsen, P. A. Tester, and B. T. Vinyard, "Detection of harmful algal blooms using photopigments and absorption signatures: a case study of the Florida red tide dinoflagellate, *Gymnodinium breve*," Limnol. Oceanogr. **42**, 1240–1251 (1997).
72. S. E. Craig, S. E. Lohrenz, Z. Lee, K. L. Mahoney, G. J. Kirkpatrick, O. M. Schofield, and R. G. Steward, "Use of hyperspectral remote sensing reflectance for detection and assessment of the harmful alga, *Karenia brevis*," Appl. Opt. **45**, 5414–5425 (2006).
73. B. Lubac, H. Loisel, N. Guiselin, R. Astoreca, L. F. Artigas, and X. Mériaux, "Hyperspectral and multispectral ocean color inversions to detect *Phaeocystis globosa* blooms in coastal waters," J. Geophys. Res. **113**, C06026 (2008).
74. S. Sathyendranath, L. Lazzara, and L. Prieur, "Variations in the spectral values of specific absorption of phytoplankton," Limnol. Oceanogr. **32**, 403–415 (1987).