



# Improvement to the PhytoDOAS method for identification of coccolithophores using hyper-spectral satellite data

A. Sadeghi<sup>1</sup>, T. Dinter<sup>1,2</sup>, M. Vountas<sup>1</sup>, B. B. Taylor<sup>2</sup>, M. Altenburg-Soppa<sup>2</sup>, I. Peeken<sup>2,3</sup>, and A. Bracher<sup>1,2</sup>

<sup>1</sup>Institute of Environmental Physics, University of Bremen, Bremen, Germany

<sup>2</sup>Alfred-Wegener-Institute for Polar and Marine Research, Bremerhaven, Germany

<sup>3</sup>MARUM (Center for Marine Environmental Sciences), Bremen, Germany

Correspondence to: A. Sadeghi (sadeghi@iup.physik.uni-bremen.de)

Received: 19 October 2011 – Published in Ocean Sci. Discuss.: 23 November 2011

Revised: 23 October 2012 – Accepted: 27 October 2012 – Published: 30 November 2012

**Abstract.** The goal of this study was to improve PhytoDOAS, which is a new retrieval method for quantitative identification of major phytoplankton functional types (PFTs) using hyper-spectral satellite data. PhytoDOAS is an extension of the Differential Optical Absorption Spectroscopy (DOAS, a method for detection of atmospheric trace gases), developed for remote identification of oceanic phytoplankton groups. Thus far, PhytoDOAS has been successfully exploited to identify cyanobacteria and diatoms over the global ocean from SCIAMACHY (SCanning Imaging Absorption spectroMeter for Atmospheric CHartography) hyper-spectral data. This study aimed to improve PhytoDOAS for remote identification of coccolithophores, another functional group of phytoplankton. The main challenge for retrieving more PFTs by PhytoDOAS is to overcome the correlation effects between different PFT absorption spectra. Different PFTs are composed of different types and amounts of pigments, but also have pigments in common, e.g. chl *a*, causing correlation effects in the usual performance of the PhytoDOAS retrieval. Two ideas have been implemented to improve PhytoDOAS for the PFT retrieval of more phytoplankton groups. Firstly, using the fourth-derivative spectroscopy, the peak positions of the main pigment components in each absorption spectrum have been derived. After comparing the corresponding results of major PFTs, the optimized fit-window for the PhytoDOAS retrieval of each PFT was determined. Secondly, based on the results from derivative spectroscopy, a simultaneous fit of PhytoDOAS has been proposed and tested for a selected set of PFTs (coccolithophores, diatoms and dinoflagellates) within an optimized fit-window, proven by spectral orthog-

onality tests. The method was then applied to the processing of SCIAMACHY data over the year 2005. Comparisons of the PhytoDOAS coccolithophore retrievals in 2005 with other coccolithophore-related data showed similar patterns in their seasonal distributions, especially in the North Atlantic and the Arctic Sea. The seasonal patterns of the PhytoDOAS coccolithophores indicated very good agreement with the coccolithophore modeled data from the NASA Ocean Biochemical Model (NOBM), as well as with the global distributions of particulate inorganic carbon (PIC), provided by MODIS (MODerate resolution Imaging Spectroradiometer)-Aqua level-3 products. Moreover, regarding the fact that coccolithophores belong to the group of haptophytes, the PhytoDOAS seasonal coccolithophores showed good agreement with the global distribution of haptophytes, derived from synoptic pigment relationships applied to SeaWiFS chl *a*. As a case study, the simultaneous mode of PhytoDOAS has been applied to SCIAMACHY data for detecting a coccolithophore bloom which was consistent with the MODIS RGB image and the MODIS PIC map of the bloom, indicating the functionality of the method also in short-term retrievals.

## 1 Introduction

Phytoplankton play an important role in the marine ecosystems as the basis of the ocean food chain. Phytoplankton are the main oceanic primary producers and contribute to the global carbon cycle by acting as the biological pump for carbon in the ocean (Raven and Falkowski, 1999). Dissolved

carbon dioxide is fixed via photosynthesis and released as organic carbon to ocean environment, where it can sink down directly or indirectly through other trophic levels to the ocean floor, or just be recycled in the upper ocean. The most feasible approach to monitor the global distribution of marine phytoplankton and to estimate their total biomass is the use of satellite data (e.g. Platt and Sathyendranath, 1988), which corresponds to the field of ocean-color remote sensing. Using ocean-color sensors, long-term records of aquatic parameters are provided remotely on a global scale, which have different applications (McClain, 2009), e.g. to improve the understanding of ocean biogeochemistry and marine ecosystem dynamics, to assess fishery productivity, to be used as input data for ocean modeling, etc. In retrieving phytoplankton biomass, most bio-optical ocean-color algorithms (e.g. O'Reilly et al., 1998) derive the concentration of chlorophyll *a* (chl *a*). Chl *a*, as a common pigment among all phytoplankton species, is generally used as a proxy for phytoplankton biomass (Falkowski et al., 1998) and is a classic index of ocean health or primary production (Kirk, 1994). However, aside from chl *a* estimation, several distinct attempts have been made for remote identification of phytoplankton groups. In particular, remote identification of phytoplankton functional types (PFTs) has been of interest to Earth system modeling, due to the specific biogeochemical impacts of different phytoplankton groups (see summary by Nair et al., 2008). These attempts had several purposes, from space-borne detection of harmful algal blooms (Kahru et al., 2004) and global-scale mapping of dominant phytoplankton groups (Alvain et al., 2005) to the improvement of the accuracy of satellite-derived chl *a* in waters with multiple dominating phytoplankton populations (Morel, 1997; Sathyendranath et al., 2004; Aiken et al., 2008). Even though some taxonomic groups (e.g. diatoms and coccolithophores) at the same time belong to different PFTs, the detection of the most important taxonomic groups is a necessary step towards the understanding of the PFT distribution in the global ocean (because PFT is a concept, while a taxonomic group is the biological entity).

Most ocean color methods (see IOCCG report 5, 2006) rely on empirical algorithms, based on relationships between apparent optical properties, inherent optical properties and geophysical parameters (e.g. chl *a*) inferred from a set of in-situ data, which are regionally biased. Many studies to derive PFTs from space are also connected to these algorithm principles. Using additional sets of in-situ data, including regional distributions of PFTs and optical parameters, these studies connect the changes observed in optical parameters to the variations measured in pigment compositions, cell size and phytoplankton populations (e.g. Alvain et al., 2005, 2008; Aiken et al., 2007). Some studies have suggested the application of a regionally parameterized algorithm instead of the generic retrieval algorithm (Sathyendranath et al., 2004).

Nevertheless, the fact that all these approaches are more or less dependent on large sets of a-priori in-situ measurements has motivated alternative methods to be developed in this field. PhytoDOAS (Bracher et al., 2009) was also established in this context, relying on an essentially different algorithm than the multispectral ocean-color algorithms. For instance, PhytoDOAS uses the whole spectral information on a large wavelength range, instead of using just a few wavelength bands reflectance data. Using this method, global distribution of two major PFTs, diatoms and cyanobacteria, has been quantitatively derived by Bracher et al. (2009) from the data provided by SCIAMACHY, a hyper-spectral satellite sensor on-board Envisat. The study presented here is dedicated to improving the PhytoDOAS method in order to discriminate another PFT (coccolithophores) from SCIAMACHY data. PhytoDOAS, as an extension of Differential Optical Absorption Spectroscopy, DOAS (Perner and Platt, 1979), to the aquatic medium, is based on using differences in spectral features of absorption spectra of major PFTs in order to discriminate between them. Distinguishing fine spectral differences in absorption effects of different PFTs over a wide wavelength range requires utilizing a hyper-spectral sensor. However, contrary to the distinctive spectral behaviors in absorption spectra of atmospheric trace gases – which make them fairly straightforward targets to be discriminated by DOAS – absorption spectra of phytoplankton species contain strong spectral correlation in any operating wavelength window. This fact refers to a basic challenge in the retrieval of more PFTs by this method. This study proposes some solutions to overcome this limitation, which are discussed in detail in Sect. 2.4. Corresponding results, including the monthly and seasonal distributions of coccolithophores, are presented as comparisons with dinoflagellates- and the coccolithophore-related data.

## 2 Material and method

### 2.1 From DOAS to PhytoDOAS

PhytoDOAS, as a retrieval method of phytoplankton groups, is an extension of DOAS (Perner and Platt, 1979) from the atmospheric domain into the aquatic media. DOAS expands the *Beer-Lambert* law to all possible interactions between light and all atmospheric optical components. The DOAS retrieval can be expressed by the following equation:

$$\left\| \tau(\lambda) - \sum_{i=1}^N \sigma'_i(\lambda) S C_i - \rho'(\lambda) S_r - \sum_{p=0}^M b_p \lambda^p \right\|^2 \rightarrow \min, \quad (1)$$

where  $\tau(\lambda) = \ln \frac{I_0(\lambda)}{I(\lambda)}$  is the *optical thickness* of the atmosphere, with  $I_0(\lambda)$  and  $I(\lambda)$  being the measured radiations at the top of the atmosphere for the solar irradiance and the Earth's backscattered radiance, respectively;  $\sigma'_i(\lambda)$ , so-called *differential cross-section*, is the subtraction of the absorption

cross-section of a given trace gas from a low-order polynomial fitted to that;  $SC_i$ , so-called *slant column density*, is a coefficient demonstrating the total amount of a given absorber per unit area integrated along the atmospheric light-path [molecules  $\text{cm}^{-2}$ ];  $S_r\rho'(\lambda)$  contains the spectral impact of the Ring effect (rotational Raman scattering of air molecules), needed to be accounted for in DOAS retrieval (Vountas et al., 1998, 2003); and  $\sum b_p\lambda^p$  is a low-order fitted polynomial, covering the slowly varying parts of all spectra, including the spectral component associated with Rayleigh and Mie scattering. Here, the squared notation of the whole equation refers to the least square optimization, which is implemented in DOAS by minimizing the residuals (for details about DOAS equation, see e.g. Vountas et al., 2007).

PhytoDOAS was born when the DOAS method was applied for retrieving oceanic phytoplankton, as the living light-absorbing particles of *case-I* waters (Vountas et al., 2007; Bracher et al., 2009). The underlying idea is that the backscattered light from the ocean into the atmosphere (and hence to the satellite sensor) carries some information from water and its optical constituents. Regarding the standard DOAS equation (Eq. 1), the PhytoDOAS equation (Eq. 2) contains two additional terms, as explained by Bracher et al. (2009):

$$\left\| \tau(\lambda) - \sum_{i=1}^N \sigma'_i(\lambda) SC_i - \rho'(\lambda) S_r - a'(\lambda) S_a - v'(\lambda) S_v - \sum_{p=0}^M b_p \lambda^p \right\|^2 \rightarrow \min, \quad (2)$$

where  $a'(\lambda)S_a$  and  $v'(\lambda)S_v$  are associated with phytoplankton absorption and vibrational Raman scattering (VRS) of water molecules, respectively. More precisely,  $a'(\lambda)$  is the differential part of *specific absorption spectrum* of the target phytoplankton group (absorption spectrum of the phytoplankton sample normalized by its measured chl *a* concentration) in [ $\text{m}^2 (\text{mg chl } a)^{-1}$ ], and  $v'(\lambda)$  is the differential part of the VRS spectrum, obtained from a reflectance model developed by Vasilkov et al. (2002) and Vountas et al. (2003).

The main outputs of the PhytoDOAS equation (Eq. 2) are  $S_a$  [ $\text{mg chl } a \text{ m}^{-2}$ ] and  $S_v$  (unitless), being the *fit factor* coefficients attributed to the PFT absorption spectrum and the spectral signature of VRS, respectively. These fit factors are retrieved independently through two separate fit processes. Since spectral features of VRS are weak in the visible, whereas the absorption features of PFTs are fitted in visible, a second fit is done for determining the *VRS fit factor* ( $S_v$ ) in the UV range, which is then extrapolated to the visible (Bartlett et al., 1998; Vountas et al., 2003, 2007) for estimation of the average light-path in the observed water body (Vountas et al., 2007).

It must be noted that, in addition to phytoplankton, there are other important optical components in *case-I* waters: CDOM (colored dissolved organic matter) and non-algal par-

ticulates. However, spectrally smooth behaviors of CDOM (Bricaud et al., 1981; Carder et al., 1989) and non-algal particulates (Allali et al., 1995; Mitchell et al., 2000), in the wavelength range considered in this retrieval, lead their absorption to be covered in the PhytoDOAS equation by the fitted polynomial, as well as the absorption and scattering of water molecules and particle scattering (Kirk, 1994; Gordon et al., 1975).

It was shown by Vountas et al. (2007) that there is a strong relation between VRS and average light-path in water, suggesting the former as a proxy for the quantitative estimation of the latter. Hence, as explained in Vountas et al. (2007) and Bracher et al. (2009), the chl *a* concentration of the target PFT for each satellite ground pixel is estimated through dividing the PFT absorption fit factor ( $S_a$ ) by the respective average light-path in the water ( $\delta$ ) as follows:

$$C = \frac{S_a}{\delta}, \quad (3)$$

where  $C$  is the chl *a* concentration of the target PFT, [ $\text{mg m}^{-3}$ ].

PhytoDOAS was initially used to identify diatoms and cyanobacteria in *case-I* waters using SCIAMACHY hyperspectral data (Bracher et al., 2009).

Regarding the PhytoDOAS equation (Eq. 2), three following sets of input data are needed to perform this method: (1) satellite measurements, i.e. extraterrestrial solar irradiance,  $I_0(\lambda)$ , and Earth's backscattered radiation,  $I(\lambda)$ , both measured by the satellite sensor at the top of the atmosphere, to be embedded into the optical depth, through  $\tau(\lambda) = \ln \frac{I_0(\lambda)}{I(\lambda)}$ ; (2) atmospheric spectra, i.e. absorption cross-sections of water vapor, trace gases and spectral signature of the Ring effect; and (3) aquatic spectra, i.e. specific absorption spectrum of the PFT and the spectral signature of VRS.

As an improvement to PhytoDOAS, presented in this study, instead of just a single PFT target, three selected PFT targets are fitted simultaneously. This new approach is referred to as *multi-target fit* (see Sect. 2.4). Accordingly, to incorporate the *multi-target fit* in PhytoDOAS, the term of PFT absorption in Eq. (2), i.e.  $a'(\lambda)S_a$ , must be replaced by a multiple term containing the absorption spectra of the selected PFTs; within this multiple absorption term,  $\sum_{j=1}^3 a'_j(\lambda)S_{aj}$ , a certain absorption fit factor is assigned to each PFT target. Therefore, the improved PhytoDOAS can be introduced by this expression:

$$\left\| \tau(\lambda) - \sum_{i=1}^N \sigma'_i(\lambda) SC_i - \rho'(\lambda) S_r - \sum_{j=1}^3 a'_j(\lambda) S_{aj} - v'(\lambda) S_v - \sum_{p=0}^M b_p \lambda^p \right\|^2 \rightarrow \min. \quad (4)$$

In PhytoDOAS (as in DOAS), the overall *chi-square* value,  $\chi^2$ , is used as a scalar indicator of the total fit quality.

Furthermore, the fit spectrum of the retrieval target is compared with its original spectrum for each oceanic pixel to check the fit quality.

## 2.2 Satellite data

Due to spectral correlation of phytoplankton absorption features, satellite data used in PhytoDOAS must be highly spectrally resolved. To meet this requirement, data measured by the satellite sensor SCIAMACHY (SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY) are used in this study. SCIAMACHY, originally designed for atmospheric measurements, covers a wide wavelength range (240 nm to 2400 nm) with a spectral resolution between 0.2 nm and 1.5 nm (Bovensmann et al., 1999). In this study, the SCIAMACHY data in UV and visible regions, from nadir-viewing geometry with a spectral resolution of 0.24 nm to 0.48 nm, have been used. These data correspond to backscatter radiation from the Earth's surface, with the spatial resolution of about 30 km × 60 km, which defines the pixel size in this wavelength region. Each ground pixel is associated with a direct measurement of solar irradiance at the top of the atmosphere in the same wavelength region, to be used later in the retrieval as the unattenuated radiation,  $I_0$ . Within PhytoDOAS, SCIAMACHY data are used in the two following steps: first, SCIAMACHY visible data are used to fit the absorption spectrum of target PFT within the wavelength range of 429 nm to 495 nm in the study by Bracher et al. (2009) and extended up to 521 nm in this study, leading to PFT absorption fit factors; secondly, SCIAMACHY data from 340 nm to 385 nm are exploited to fit the VRS spectral signature of water molecules, leading to the VRS fit factors, necessary for the estimation of the average light-path in water for each oceanic pixel (see previous section).

Moreover, Moderate resolution Imaging Spectroradiometer (MODIS)-Aqua level-3 products were used to obtain the seasonal composites of particulate inorganic carbon (PIC) in a 9 km<sup>2</sup> grid during spring and summer 2005 (the MODIS algorithm for PIC has been described in Balch et al., 2005). Since coccolithophores are the main planktonic producer of calcium carbonate, suspended PIC in open ocean is an indicator of coccolithophores (Balch et al., 2005). Hence, the PhytoDOAS retrievals of coccolithophores were compared to the global distribution of PIC. In addition, the PIC and PhytoDOAS monthly mean data from a selected region in the North Atlantic, used in Sadeghi et al. (2012), were statistically compared.

From pigment-derived synoptic relationships, based on the method of Hirata et al. (2011), the SeaWiFS level-3 chl *a* products were used to calculate the seasonal distribution of haptophytes in 2005 on global scale. Haptophytes are a phytoplankton taxonomic group including coccolithophores, among others. This abundance-based PFT algorithm links the variations in phytoplankton community structure to the variations in total chl *a*, extracted from the SeaWiFS level-3 prod-

ucts. The variation in phytoplankton community structure itself is derived from analyzing and quantifying the information on biomarker pigments through a large in-situ HPLC dataset.

As an initial case study, the PhytoDOAS detection of a certain coccolithophore bloom (in December 2009) was compared to a weekly composite of MODIS-Aqua PIC for the respective period (see Sect. 3.3).

## 2.3 Spectral data

In addition to hyper-spectral satellite radiation measurements, the PhytoDOAS retrieval requires reference spectra of atmospheric and oceanic species. For atmospheric spectra, absorption cross-sections of ozone, NO<sub>2</sub>, glyoxal or OCHCHO, iodine oxide, O<sub>4</sub>, water vapor and the Ring effect (as a *pseudo-absorber* spectrum) are fitted, using the same spectra as utilized in Bracher et al. (2009). As required by the PhytoDOAS *triple-target fit*, three phytoplankton absorption spectra were used in this study. The absorption spectrum of coccolithophores was acquired from an *E. huxleyi* culture (*E. huxleyi* is the dominant species of coccolithophores; see e.g. Tyrrell and Merico, 2004). A dinoflagellate-dominated natural sample was used to obtain the respective absorption spectrum. In both cases, samples were measured with a point-source integrating-cavity absorption meter, PSICAM (Roettgers et al., 2007). Total chl *a* concentrations of the *E. huxleyi* culture and of all pigments for the natural dinoflagellate sample were obtained from high-performance liquid chromatography (HPLC) following the method described by Hoffmann et al. (2006). The dinoflagellate-dominated sample was taken during the OOMPH field experiment (Organics over the Ocean Modifying Particles in both Hemispheres) with RV *Marion Dufresne* on 2 February 2007, at 59.88° W and 46.01° S, within a dinoflagellate bloom. Performing the CHEMTAX analysis (Mackey et al., 1996) on the HPLC data of the natural sample, the chl *a* concentrations of all containing phytoplankton groups were calculated, indicating a domination of 92 % for dinoflagellates. To derive the specific absorption spectra, each absorption spectrum was normalized by the respective chl *a* concentration. The third phytoplankton reference spectrum, i.e. diatoms, was acquired from in-situ measurements conducted during the EIFEX Atlantic Southern Ocean cruise (ANTXXI/3; 14 March 2004 at 49.4° S and 2.1° E) and is the same as used in Bracher et al. (2009).

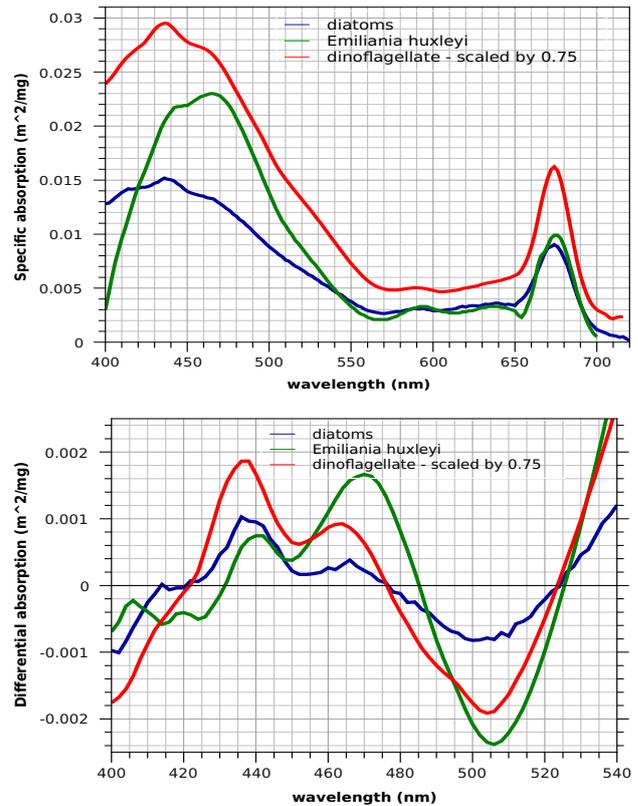
It must be noted that no absorption spectrum of coccolithophores dominating the phytoplankton biomass by over 50 % in a natural sample was available. Nevertheless, the specific absorption spectrum of coccolithophores (obtained from *E. huxleyi* cultures) used in this study is spectrally very similar to the specific absorption of natural samples measured by Siegel et al. (2007) in a coccolithophore bloom off the Namibia coast (Benguela Upwelling). The latter spectrum was well comparable to the absorption spectra obtained from

the coccolithophore-dominated natural samples in the Kattegat. However, these natural samples of Siegel et al. (2007) do not have detailed HPLC analysis to prove the domination via chl *a* concentration of coccolithophores among the total phytoplankton biomass.

Figure 1 (upper panel) shows the specific absorption spectra of the three PFTs used as the retrieval targets in this study, i.e. for *E. huxleyi*, dinoflagellates and diatoms. From these measured spectra, the corresponding differential absorption spectra have been derived (lower panel in Fig. 1), based on the separation approach described in the method section (see Sect. 2.1).

## 2.4 Improvement to PhytoDOAS: challenges and approaches

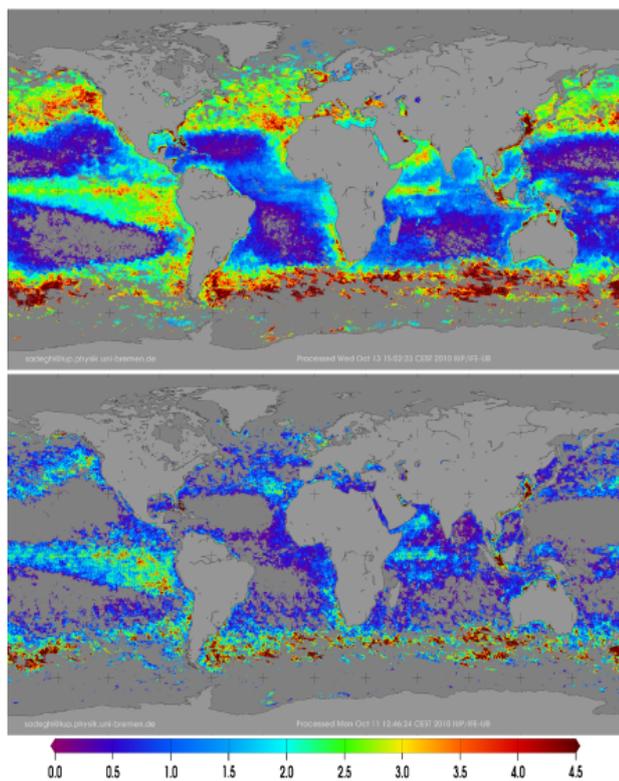
Following the method by Bracher et al. (2009), besides cyanobacteria and diatoms, more major PFTs (or dominant species of a PFT) are expected to be retrieved. Nevertheless, there are some challenges to be overcome in order to improve PhytoDOAS to be a reliable retrieval tool for other PFTs (or dominant species). The main challenge is the spectral correlation between absorption spectra of different phytoplankton targets, which arises from their common photosynthetic pigments and causes in turn difficulties to distinguish different groups remotely. In this study, several approaches have been combined to overcome the correlation found among typical PFT absorption spectra. Firstly, investigations proved that *multi-target fitting* leads to higher fit quality as compared to fitting only one PFT spectrum at the time. This approach results in significantly lower values for the absorption fit factors of each target, compared to the previous approach of the *single-target fitting* (see Fig. 2). This is of importance, because high fit factors in PhytoDOAS lead to an overestimation of the PFT concentration. This can be explained by the fact that when three (or more) PFT targets are fitted simultaneously, the phytoplankton spectral input of the PhytoDOAS equation becomes bio-optically more realistic; i.e. in this case PhytoDOAS accounts for more optical components of the ocean water, which usually contains simultaneously several types of phytoplankton species. Comparably, when we omit some trace gases from our DOAS retrieval, the retrieval results of the others are affected. Figure 2 shows different fit factor results obtained for *E. huxleyi* by following the *single-target fit* and *triple-target fit* modes of PhytoDOAS (the *triple-target fit* includes *E. huxleyi* together with diatoms and dinoflagellates). The *triple-target fit* results (lower panel in Fig. 2) are characterized by lower values of fit factors, almost over the whole global ocean, as compared to the *single-target fit* results (Fig. 2 upper panel). This is in a better agreement with the coccolithophore chl *a* data achieved by the NASA Ocean Biochemical Model, NOBM (Gregg et al., 2003; Gregg and Casey, 2007). Moreover, the results are also verified by looking at the fit residuals, as a measure of the fit goodness, represented by the average



**Fig. 1.** Upper panel: specific absorption spectra of coccolithophores (*E. huxleyi*) (green), dinoflagellates (red) and diatoms (blue). The first spectrum was obtained from a culture, while the two latter spectra were taken from in-situ samples. All spectra were measured using a point-source integrating-cavity absorption meter. Lower panel: differential absorption spectra of three phytoplankton targets. Each of them was derived by subtracting a second-order polynomial from the corresponding specific absorption spectrum, which was shown in upper panel.

value of *chi-square* ( $\chi^2$ ) for all accounting pixels. Since a reliable fit, associated with a low fit residual, corresponds to a low value of averaged  $\chi^2$ , the latter quantity can be used as an indicator of the fit quality. In this sense, the fact that the PhytoDOAS *triple-target fit* represents lower values of averaged  $\chi^2$ , compared to the PhytoDOAS *single-target fit*, implies the privilege of the *triple-target fit*. In the example fit factor maps shown in Fig. 2, the averaged  $\chi^2$  values for the *triple-target* and *single-target* fits are 0.00039 and 0.00052, respectively (see also the comparison of overall residuals in Fig. 4).

However, due to the limitations imposed by the spectral correlation, it is necessary to determine and optimize some factors when running a *multi-target fit*, in order to receive an acceptable fit quality: the retrieval should be optimized by identifying how many PFT targets, in which combination and within which wavelength window are fitted simultaneously. The following methods have been used to investigate



**Fig. 2.** Global fit factor maps [ $\text{mg chl m}^{-2}$ ] for *E. huxleyi* in March 2005, obtained by single-target fit (upper panel) and triple-target fit (lower panel) modes of PhytoDOAS using SCIAMACHY data.

the optimized fit-window, the set of PFTs for *multi-target fit* and the fit quality in each tested option, which can be listed as follows:

- comparison of the overall mean of *chi-square* values for different fit results (explained above);
- the *fourth-derivative analysis* for analyzing the fine spectral features of the different PFT absorption spectra (to be explained below);
- the orthogonality survey of the different PFT absorption spectra (to be explained in the Appendix);
- comparison of the fit absorption spectra with the input spectra for selective pixels.

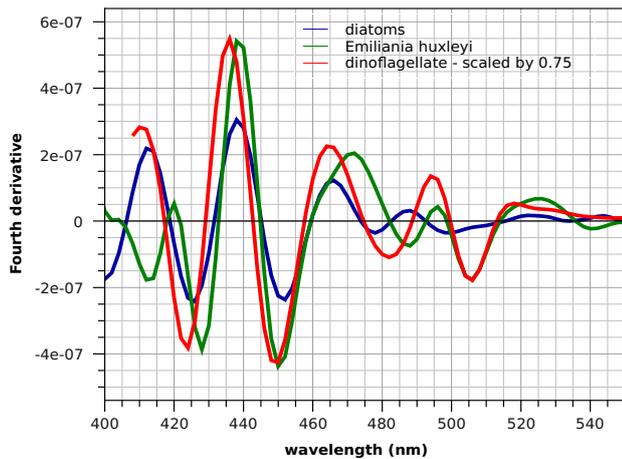
Although the most reliable option for the quality test of retrieval will be, of course, validating the results by comparing them with the available high-quality in-situ measurements, however, it cannot always be fulfilled. The general reason for this limitation is the very low availability of in-situ data with respect to the global distributions of major PFTs. In particular, in the case of coccolithophores there is a complexity pertaining to the in-situ measuring of the whole group: via HPLC method and flow-cytometry, the pigments of other

haptophyte species (e.g. *Phaeocystis*) are spoiling the measurements of coccolithophores; by microscopic techniques (as they cannot detect cells less than  $5\ \mu\text{m}$ ) part of the coccolithophore cells are not accounted for in the measurement. Hence, there are uncertainties in the determination of coccolithophore concentrations, which limit the validation of the respective satellite retrievals with in-situ measurements. Moreover, there is a specific difficulty associated with the collocation (matching) of SCIAMACHY ground pixels to the existing in-situ data due to their large pixel size ( $30 \times 60\ \text{km}^2$ ), which limits strongly the available match-up points.

Therefore, to validate our coccolithophore retrievals (see Sect. 3.2), we compared our results with the available products related to global distribution of coccolithophores, including the following:

- NOBM modeled (assimilated) data of PFT global distribution (see Sect. 3.2);
- MODIS-Aqua global PIC concentration, as a proxy of coccolithophores (Balch et al., 2005);
- haptophytes global biomass derived from SeaWiFS chl *a* products according to Hirata et al. (2011).

The *fourth-derivative spectroscopy* (Aguirre-Gomez et al., 2001) was an important approach used in this study to analyze the spectral behavior of PFT absorption spectra. The core concept here is as follows: in the fourth-derivative curve of a given absorption spectrum, each peak corresponds to the maximum absorption for a specific pigment at the same wavelength position. Therefore, the distribution of peak positions in a fourth-derivative curve is an indicator of pigment composition for that PFT. Figure 3 shows the fourth-derivative curves for the specific absorption spectra of three phytoplankton targets, which have been fitted simultaneously via PhytoDOAS. The fourth-derivative method can be used to identify tiny differences in PFT spectral behavior. This is helpful to avoid spectral correlations between different phytoplankton targets and to find the appropriate wavelength window to fit them simultaneously. As shown in Fig. 3, there is a spectral difference between target spectra in the interval from 495 nm to 521 nm, especially between diatoms and *E. huxleyi*, for which the spectral behaviors are more alike in the wavelengths below 495 nm. Practically, in the simultaneous PhytoDOAS fit, the spectral differences seen in the fourth-derivative curves have been used (as one criterion) to select the set of PFT targets, i.e. the proper combination of PFTs, and also to specify the wavelength range of the actual fit-window. This explains why in this study a wider fit-window (429 nm to 521 nm) has been used than in Bracher et al. (2009), which was from 429 nm to 495 nm. The optimized set of PFT targets and respective wavelength window were also investigated through comparing the level of linear-independence of absorption spectra for different sets of PFTs within various ranges of fit-windows. This approach, referred



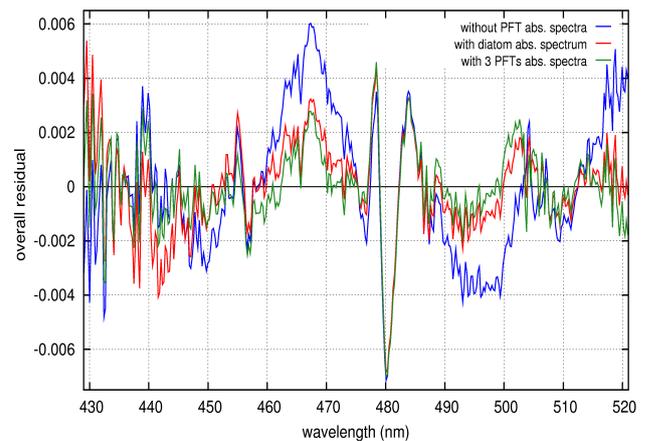
**Fig. 3.** Fourth-derivative curves of the specific absorption spectra of the three PFTs shown in Fig. 1: diatoms (blue), coccolithophores (green) and dinoflagellates (red). The latter curve was scaled to 0.75.

to as the *spectral orthogonality test*, is explained in the Appendix.

Nevertheless, to decide about the best set of PFTs for a simultaneous fit (for a given wavelength window), two restrictions should be considered: firstly, when the values are comparable, the preference would be toward the set of PFTs who are more abundant in the global ocean and also existing together or succeeding each other in the same regions; secondly, as cyanobacteria have a spectral signature distinguishable from the others, they can be excluded from the final set to be retrieved via the PhytoDOAS *single-target* mode; this can be also justified because cyanobacteria are often favored in the oceanic habitats where the other species are not so abundant.

Overall, the set of PFTs used in the simultaneous retrieval process (via the PhytoDOAS *triple-target* fit) has been selected to include diatoms, coccolithophores (*E. huxleyi*) and dinoflagellates. The appropriate fit-windows for retrieving these absorption targets were chosen in conjunction with the results of the derivative spectral analysis (introduced below), leading to the wavelength range of 429 nm to 521 nm. To reach the final set of PFTs and the optimized fit-window, above considerations (and derivative analysis) were associated along with testing several other possible options, controlled by the  $\chi^2$  threshold. As a consequence, regarding the results of the *spectral orthogonality test* (Table A1 in Appendix), the factor introduced for the spectral independence shows significantly lower value for the selected fit-window, compared to the fit-window used in Bracher et al. (2009) (i.e. 429 nm to 495 nm).

Figure 4 compares the fit quality of PhytoDOAS retrieval when being performed in the *single-target fit* mode and in the *multi-target* mode, using the overall fit residuals. It can be seen that the overall residual for the single-target fit (di-



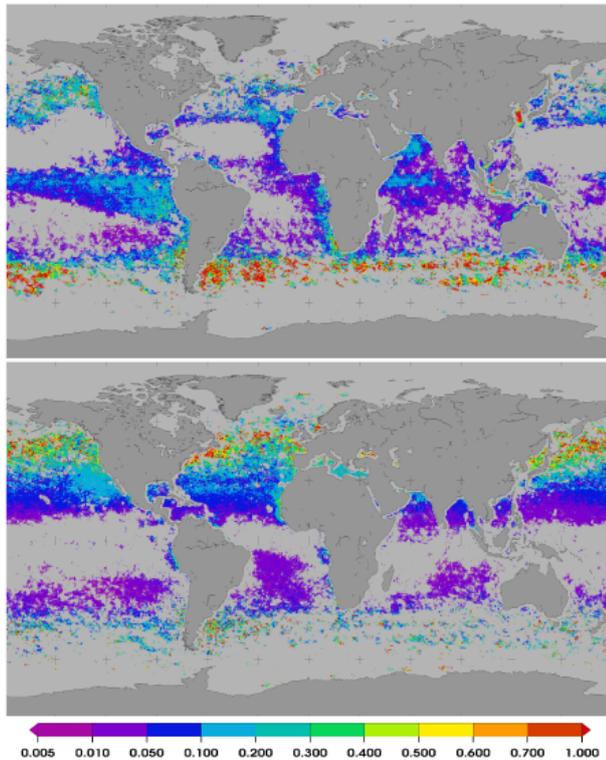
**Fig. 4.** The improvement to PhytoDOAS was performed by a simultaneous fit of the absorption spectra of selected phytoplankton targets in an appropriate wavelength window. This figure compares the overall residuals of the PhytoDOAS retrievals in three different fit-modes: without any PFT target (blue), only with the diatoms absorption spectrum (red) and with the absorption spectra of three selected PFTs (green). All three residual spectra have been obtained via consecutive runs of DOAS for the same SCIAMACHY orbit, passing over North Atlantic (the first of July 2005). A sample ground-pixel of this orbit has been taken to plot the residuals, which corresponds to the pixel-center located at 54.51° N and 21.47° W.

atoms; in red), almost over the whole fit-window, is higher than for the *triple-target fit* (diatoms, dinoflagellates and coccolithophores; in green). Furthermore, regarding the DOAS method, this figure also shows that when the spectral contribution of phytoplankton absorption is taken into account by the DOAS retrieval (over the ocean), the fit quality will be clearly better; this can be inferred from the obviously higher residual of the retrieval when running without any PFT target (blue curve).

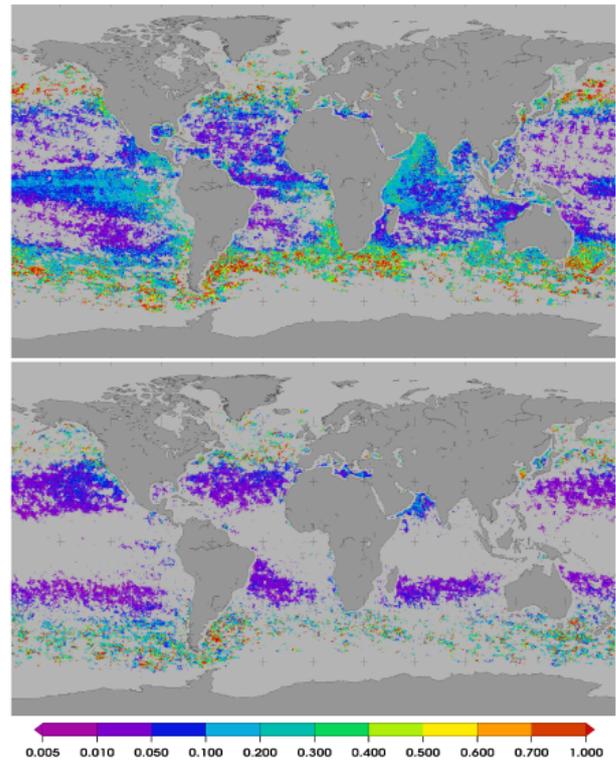
### 3 Results and discussion

#### 3.1 Monthly averages by triple-target fit

In this section the results of using PhytoDOAS to retrieve coccolithophores and dinoflagellates from SCIAMACHY data 2005 are presented as monthly averages of global chl *a* distributions. The results were obtained by conducting the *triple-target fit* mode of PhytoDOAS, with diatoms, *E. huxleyi* and dinoflagellates absorption spectra as the input PFT targets over the wavelength range of 429–521 nm. In this configuration the average value of the overall *chi-square* was minimal and fit spectra in selected oceanic pixels were in good agreement with the original PFT absorption spectra. Figures 5 and 6 show the monthly averaged global distribution of chl *a* for coccolithophores and dinoflagellates using SCIAMACHY data from March and October 2005.



**Fig. 5.** Global distribution of chl  $a$  [ $\text{mg m}^{-3}$ ] for coccolithophores (upper panel) and dinoflagellates (lower panel), retrieved via PhytoDOAS method in the *triple-target* mode from SCIAMACHY data; monthly averages over March 2005.



**Fig. 6.** Global distribution of chl  $a$  [ $\text{mg m}^{-3}$ ] for coccolithophores (upper panel) and dinoflagellates (lower panel), retrieved via the PhytoDOAS method in the *triple-target* mode from SCIAMACHY data; monthly averages over October 2005.

In each set of these figures, there are clear differences in the distribution of chl  $a$  for these two PFT targets. For example, in Fig. 5 (March 2005) coccolithophores show high chl  $a$  in the north of the polar front and elevated chl  $a$  in parts of the tropics and subtropics, while dinoflagellates show elevated chl  $a$  in the North Atlantic and the North Pacific, where coccolithophore chl  $a$  is much lower. Furthermore, the temporal variation of patterns for each PFT over the year can be seen by comparing these two sets of chl  $a$  maps. For instance, in October, compared to March 2005, averaged chl  $a$  of coccolithophores is still high in the north of the polar front, while the North Atlantic and the North Pacific now show high values, which might be explained by the pronounced seasonal cycle of this species in the north. The chl  $a$  contents of dinoflagellates in the northern mid- and high latitudes are lower in October 2005, compared to March 2005, whereas in the southern mid- and high latitudes higher chl  $a$  contents are observed in March.

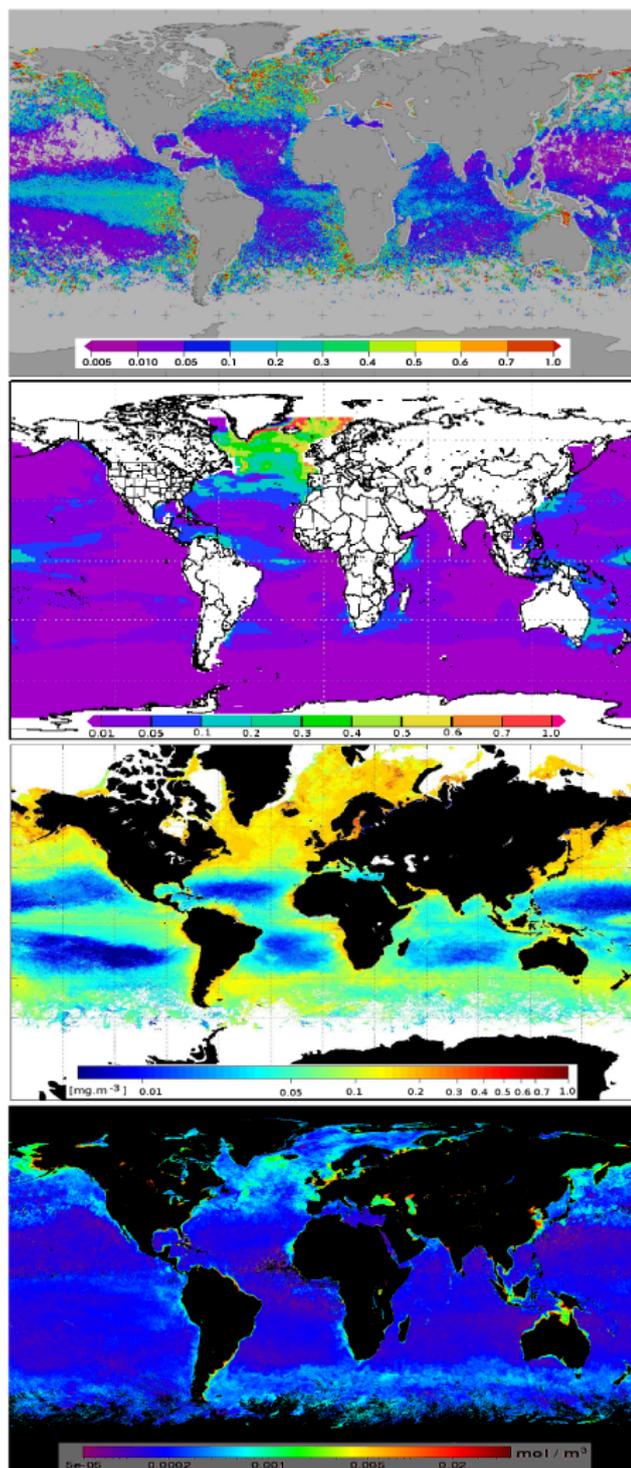
### 3.2 Comparison with coccolithophore-related data: seasonal averages

In Figs. 7 and 8 the seasonal averages of global distributions of the PhytoDOAS coccolithophores (upper panels) are com-

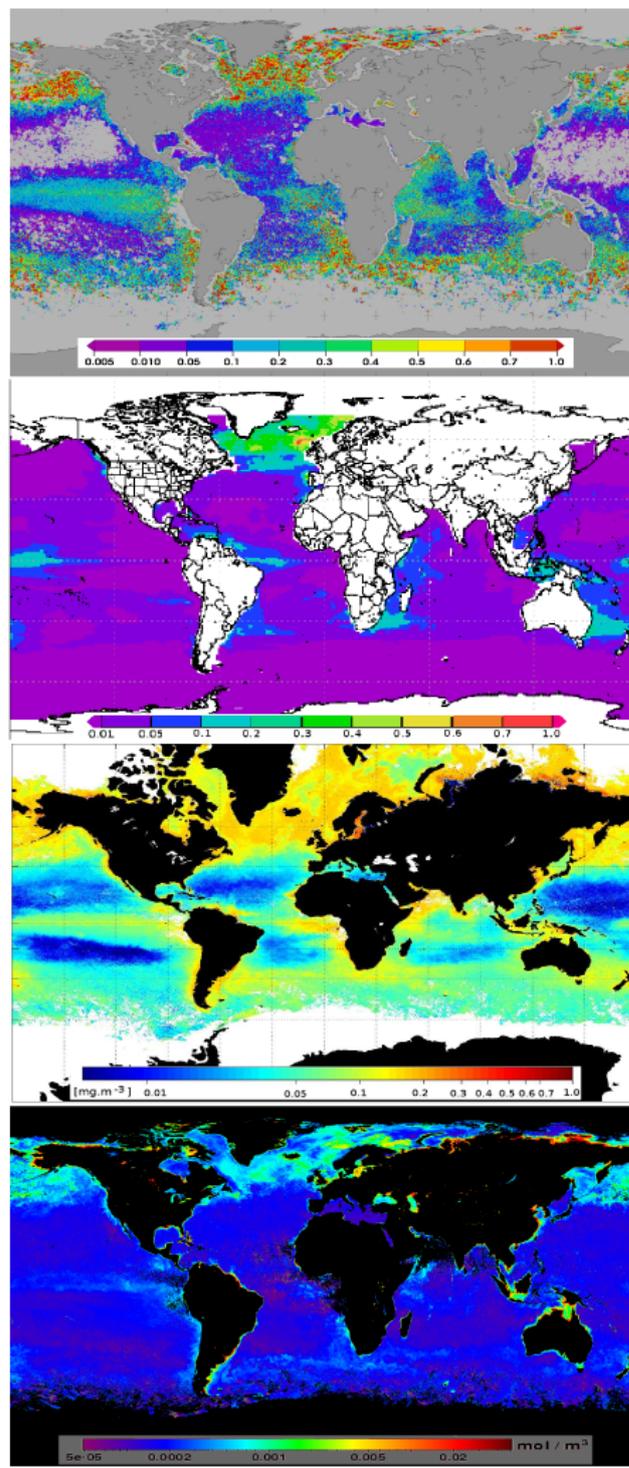
pared with three available global products related to coccolithophores: the NOBM coccolithophore modeled data (2nd upper panels); global distribution of haptophytes based on Hirata et al. (2011) (3rd panels, downwards); and the PIC concentration from MODIS-Aqua (lowest panels).

The distributions of coccolithophores from PhytoDOAS and NOBM (2nd upper panels in Figs. 7 and 8) indicate similar patterns in the North Atlantic, east of Australia (northern parts of the Tasman Sea), the mid-Pacific (partly), the tropical regions of the South Atlantic near South America (partly), the south-east waters of Africa (more pronounced in spring) and also partly in the western parts of the mid-Indian Ocean (eastern coast of Africa). Nevertheless, there are also some regions, where the retrieved and modeled results look totally different. For example, in both sets of figures (spring and summer) the retrieval results suggest high chl  $a$  over the North Pacific, while model results show almost nothing in these regions. The same feature can be seen partially in the high latitudes of the Southern Hemisphere and also in the western part of the South America in the Pacific Ocean. In overall, the coccolithophore chl  $a$  is lower in NOBM than from PhytoDOAS.

The global distribution of haptophytes (third panels in Figs. 7 and 8) is very well reproduced by the PhytoDOAS



**Fig. 7.** Comparison of the PhytoDOAS coccolithophore chl *a* [ $\text{mg m}^{-3}$ ] (upper panel) with the NOBM coccolithophore chl *a* [ $\text{mg m}^{-3}$ ] (second upper panel), the haptophytes chl *a* [ $\text{mg m}^{-3}$ ] (third panel) and the MODIS PIC concentration [ $\text{mol CaCO}_3 \text{ m}^{-3}$ ] (lower panel). All results represent the seasonal-mean values over the months of April, May and June 2005.



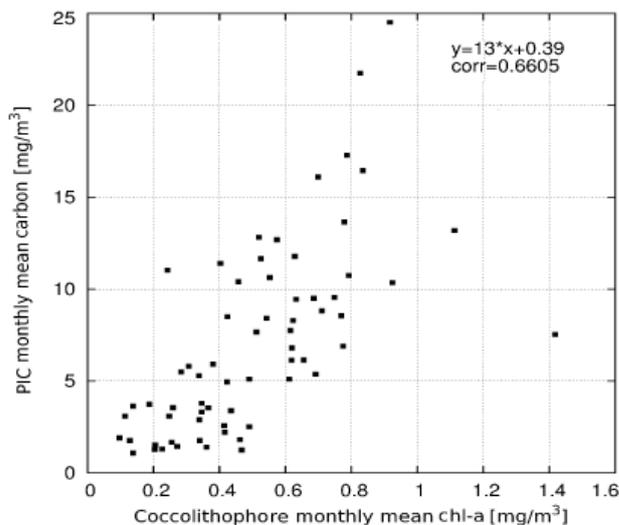
**Fig. 8.** Comparison of the PhytoDOAS coccolithophore chl *a* [ $\text{mg m}^{-3}$ ] (upper panel) with the NOBM coccolithophore chl *a* [ $\text{mg m}^{-3}$ ] (second upper panel), the haptophytes chl *a* [ $\text{mg m}^{-3}$ ] (third panel) and the MODIS PIC concentration [ $\text{mol CaCO}_3 \text{ m}^{-3}$ ] (lower panel). All results represent the seasonal-mean values over the over the months of July, August and September 2005.

coccolithophore distributions. Deviations are seen only in very few regions, e.g. around Japan, which might be explained by the contribution of other haptophyte species than coccolithophores. A second reason might be that the global pigment-based synoptic relationship is not appropriate for the specific regions due to a regional bias in the in-situ database used in Hirata et al. (2011).

The seasonal distribution patterns of the PhytoDOAS coccolithophores (upper panels in Figs. 7 and 8) show very good agreement with the PIC distributions (lowest panels) in both spring and summer 2005. The similar patterns for enhanced coccolithophores and PIC cannot only be seen on large scales, e.g. in the North Atlantic, the North Pacific, the belt-like area between the subtropical front and the northern parts of Sub-Antarctic front (in spring 2005), but also are visible on regional scales, e.g. in the Bering Sea, the Labrador Sea, northern part of the Arabian Sea (in spring 2005), the Arafura Sea (as well as in the Gulf of Carpentaria) and even in the Black Sea and the Caspian Sea. Nevertheless, the elevated values of the PhytoDOAS coccolithophores in the mid-Pacific cannot be seen well in the PIC maps, both in spring and summer 2005. Moreover, the elevated values of the PhytoDOAS coccolithophores in the southern parts of the subtropical front in summer 2005 (upper panel in Fig. 8) are not as pronounced in the respective PIC map (lowest panel in Fig. 8). Since PIC is known to be a reliable proxy of the abundance of coccolithophores in *case I* waters (Balch et al., 2005), the similar patterns mentioned above imply the functionality of the improved PhytoDOAS in retrieving coccolithophores.

The scatter plot given in Fig. 9 shows the correlation between PhytoDOAS coccolithophores and MODIS PIC in the North Atlantic, as monthly mean data over an 8-yr period (2008–2010). Time series of PhytoDOAS coccolithophores for this region, as a coccolithophore-rich area, have been studied by Sadeghi et al. (2012) along with corresponding variations in total chl *a*, PIC and three geophysical parameters. The main complexity, preventing a higher correlation of coccolithophores and PIC, as shown in Fig. 9, is arising from the simultaneous existence of the suspended coccoliths (calcite plates detached from coccolithophores), associated with other factors, e.g. different amounts of CaCO<sub>3</sub> existing in each specific type of coccolith plate (Balch et al., 2005); different numbers of coccoliths are attributed to different species of coccolithophores, etc.

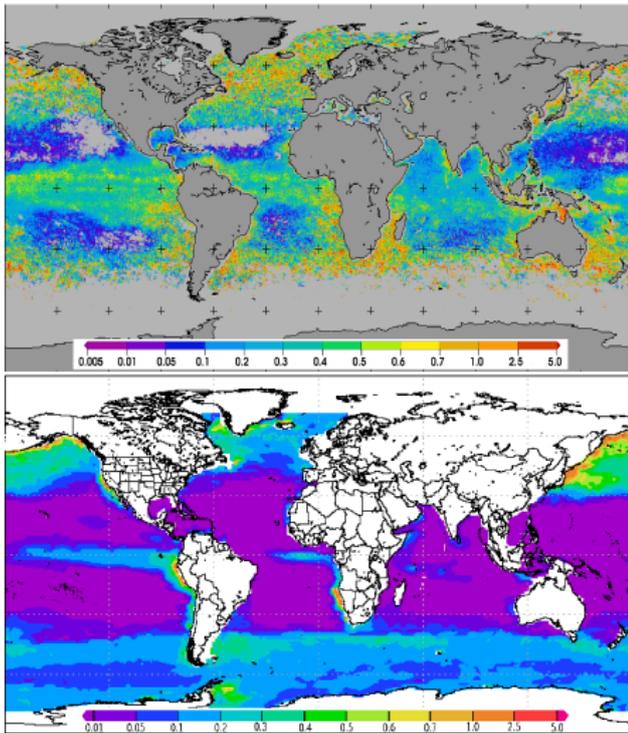
Regarding the uncertainties and roughnesses of the NOBM coccolithophore model (Gregg and Casey, 2007), the agreements shown above between the PhytoDOAS retrieval and the NOBM modeled data are regarded only as an initial approval of the method functionality. Nevertheless, as mentioned before, this primary source of comparison has been used due to the shortage (and uncertainties) of the coccolithophore in-situ data for a global comparison, as well as the limitations associated with collocating the available in-situ data with SCIAMACHY's large ground pixels. How-



**Fig. 9.** PhytoDOAS coccolithophore chl *a* [ $\text{mg m}^{-3}$ ] versus MODIS PIC [ $\text{mg C m}^{-3}$ ] for a region in the North Atlantic, confined to these geographical boundaries:  $14^{\circ}\text{W}$ – $24^{\circ}\text{W}$  and  $53^{\circ}\text{N}$ – $63^{\circ}\text{N}$ . Each data point corresponds to the respective monthly mean value from January 2003 to December 2010, according to Sadeghi et al. (2012).

ever, since both comparisons with the global distributions of haptophytes (obtained by Hirata et al., 2011) and PIC (from MODIS level-3 products) also show very similar patterns with the PhytoDOAS results, it proves the capability of the PhytoDOAS improved algorithm for the detection of coccolithophores. According to Balch et al. (2005), the two-band algorithm used in MODIS PIC products was validated using ship-derived and satellite-derived results from a variety of marine environments, showing a high overall accuracy with the standard error of  $0.08 \mu\text{g PIC L}^{-1}$ . On the other hand, the PFT algorithm used for haptophytes has also a reasonable accuracy, with the mean uncertainty of haptophytes–chl *a* relationships being 10.0 % over the entire range of the observed in-situ chl *a* (Hirata et al., 2011).

The retrieval results of diatoms via *triple-target fit* have also very similar distributions as compared to the ones, using the PhytoDOAS *single-target* mode within the fit-window of 429–495 nm, which had been achieved before and validated with in-situ data by Bracher et al. (2009). The minor differences can be attributed to the widening of the fit-window up to 521 nm and also to the fact that two more PFT spectral targets have been added into the fitting. It seems that the increasing effect of the wider fit-window is compensated by the decreasing effect of the simultaneous fit. As an example of the PhytoDOAS diatoms via the *triple-target* mode, the seasonal mean chl *a* of the PhytoDOAS diatoms is compared with the NOBM diatoms for the spring 2005 (Fig. 10, upper and lower panel, respectively). In most regions (e.g. the North Atlantic, the North Pacific, the South Atlantic and the



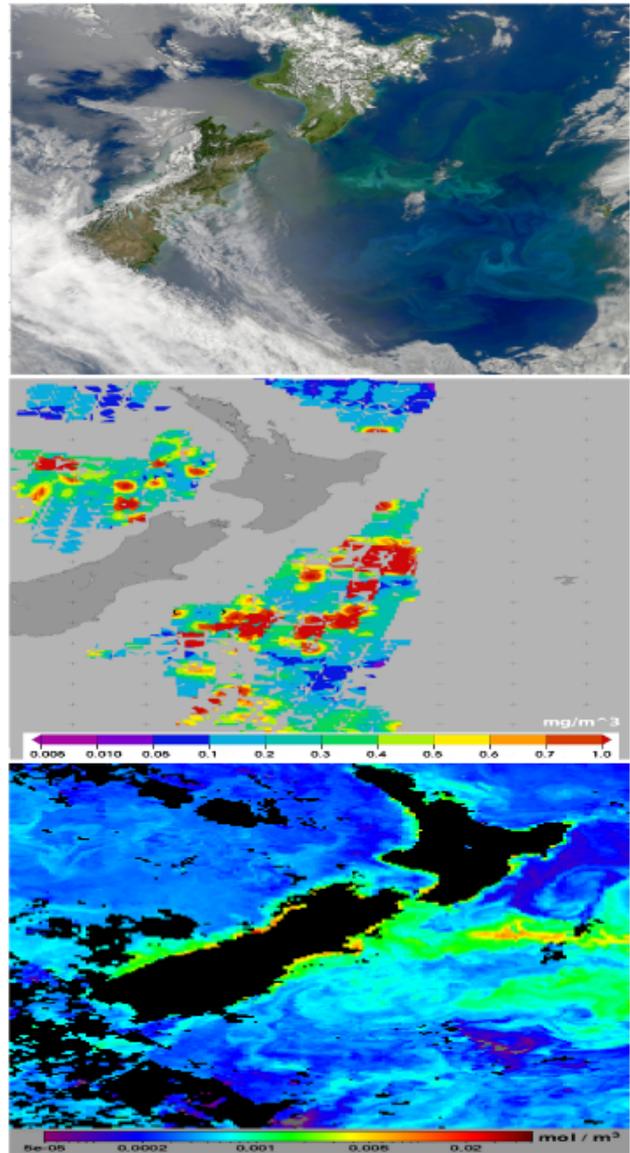
**Fig. 10.** Comparison of the PhytoDOAS retrieved chl  $a$  of diatoms [ $\text{mg m}^{-3}$ ] (upper panel) with the NOBM modeled result of diatom chl  $a$  [ $\text{mg m}^{-3}$ ] (lower panel), both in the northern spring (April/May/June) 2005.

south Pacific) there is good agreement between the retrieval and the modeled data, though the comparison is not possible in very high latitudes in the South, where SCIAMACHY has no coverage.

### 3.3 Bloom detection by PhytoDOAS

As mentioned above, since the availability of in-situ data of the PFT distributions at global scale is low (or their collocation adaptability is difficult), phytoplankton blooms can provide us the opportunity to test our retrieval method under realistic conditions.

Figure 11 shows a case study of detecting a phytoplankton bloom using PhytoDOAS. It corresponds to a coccolithophore bloom over Chatham Rise (South Pacific, east side of New Zealand), reported by NASA on 23 December 2009 (upper panel in Fig. 11). The PhytoDOAS retrieval of this bloom is shown in the middle panel of Fig. 11. It was obtained from two weeks data of SCIAMACHY (centered at 23 December 2009) via *triple-target fit* within the fit-window of 429–521 nm. The retrieval result of the bloom has also been compared to the PIC distribution over that region (lower panel in Fig. 11), which was prepared as an 8-day composite (19–26 December 2009) from MODIS-Aqua level-3 products. As shown in Fig. 11 the bloom is also visible in the



**Fig. 11.** A case study of phytoplankton bloom detection by PhytoDOAS: the upper panel is a true-color image of MODIS sensor (reported by NASA on 23rd December 2009) showing a coccolithophore bloom near Chatham island, on the eastern side of New Zealand. The middle panel depicts the PhytoDOAS retrieval of the coccolithophore chl  $a$  for the same region over a period of two weeks (centered at 23 December 2009). The lower panel illustrates the distribution of MODIS-Aqua PIC for the bloom region as an 8-date composite (19 to 26 December 2009).

PhytoDOAS retrieval as well as in the MODIS-Aqua PIC result. However, the pattern of the bloom in PhytoDOAS is not completely the same as in the MODIS RGB image and the MODIS PIC map. The main reason for this difference is the different time-frames operated for detection of this bloom: while the MODIS true-color image is an instantaneous picture, the PhytoDOAS result is a map of averaged chl  $a$

retrieved from two weeks data. Only few SCIAMACHY orbits cross this small region over one day, among which a large fraction of attaining pixels are flagged out due to the sensitivity of the retrieval to cloud contamination. The cloud contamination is also a reason for missing a lot of pixels in the eastern part. However, the choice of using a two-week time frame for the retrieval has been relying on the fact that a typical coccolithophore bloom lasts a few days, not just one day; therefore, both in the PhytoDOAS map and in the PIC map (with an 8-day time frame) the bloom is observed pronouncedly. On the other hand by using a wider time-frame, the slight motion and spreading of the bloom over the time (caused by wind, for example) will also affect the retrieval output pattern.

#### 4 Conclusions

PhytoDOAS is a new method to retrieve the biomass of phytoplankton functional types (PFTs) using hyper-spectral satellite data. This method is potentially an alternative method to retrieve total phytoplankton biomass from satellite data with higher accuracy by summing up all major PFT chl *a*, accounting for the total chl *a*. However, to retrieve total phytoplankton biomass, major PFT distribution should be first determined; hence, this study aimed to improve PhytoDOAS by retrieving another functional group, coccolithophores, than the two PFTs identified before by Bracher et al. (2009), i.e. diatoms and cyanobacteria. The main challenge to fulfill the improvement of PhytoDOAS was to overcome the spectral correlation between absorption spectra of target PFTs. The spectral correlation arises from common pigments among different PFTs (e.g. chl *a*), as well as from similar absorption regions of most uncommon pigments. The approaches used in this study to improve PhytoDOAS beyond this limitation can be summarized as follows: the fourth-derivative analysis for the recognition of tiny spectral differences between PFT spectra, which in combination with spectral orthogonality tests led to an extended fit-window for retrieving selected PFTs; simultaneous fitting of several PFTs in each retrieval process, associated with chi-square tests, which in combination with some biological a-priori knowledge led to a small modification of the selected PFT set. As a consequence, the optimized PFT set was determined to include diatoms, coccolithophores and dinoflagellates. Using these approaches, global distributions of coccolithophores have been obtained with PhytoDOAS from SCIAMACHY data for the year 2005, shown as chl *a* monthly and seasonal means. The seasonal averages of PhytoDOAS coccolithophore chl *a* for spring and winter 2005 have been compared with the MODIS-Aqua particulate inorganic carbon (PIC) global distributions, the NOBM assimilated product of coccolithophores, and also with the distribution of haptophytes obtained from another PFT algorithm (Hirata et al., 2011). The seasonal patterns of the PhytoDOAS coc-

colithophores showed (almost overall) very good agreement with the distributions PIC and haptophytes, whereas the agreement with the NOBM modeled data was good only regionally (especially in the northern latitudes). As an example of diatom retrieval by the improved PhytoDOAS, the seasonal average of diatom chl *a* for summer 2005 was shown, indicating a very good agreement with the respective result of NOBM modeled data. Unfortunately, no data source for the global distribution of dinoflagellates was available for comparing with the PhytoDOAS dinoflagellates. As a case study, the PhytoDOAS *triple-target* approach has been applied to detect a coccolithophore bloom reported by NASA in December 2009. Comparisons with NASA RGB image and also with the MODIS-Aqua PIC results over the bloom region confirm strongly the functionality of the PhytoDOAS method in short time-frames, as in detection of coccolithophore blooms.

Surely, a better spatial resolution than 30 km × 60 km will resolve better phytoplankton dynamics. Still on the global scale, as stated by Aiken et al. (2007) and already discussed in Bracher et al. (2009), phytoplankton blooms in the open ocean are often larger than 50 km × 100 km and persist over a few days to several weeks. As shown in the recently published papers by Sadeghi et al. (2012) and Ye et al. (2012), the SCIAMACHY PhytoDOAS PFT dataset can be used to study phytoplankton dynamics in specific regions over longer time scale and is useful for parameterizing and evaluating biogeochemical models. However, opposed to other PFT satellite datasets, with PhytoDOAS not only the dominant groups within a pixel are identified from their optical imprint on the satellite data (e.g. Alvain et al., 2005, 2008), but also several PFTs can be quantified with their specific chl *a* concentration. The abundance-based PFT satellite methods (e.g. Hirata et al., 2011) also give chl *a* concentration for various PFTs, but these methods are purely based on empirical relationships within regionally biased in-situ datasets and are not using the optical satellite information to infer the optical signatures of specific PFTs; so with these methods the unexpected cannot be detected.

#### Outlook

There are further steps to proceed the improvement of PhytoDOAS, as well as to expand its applications, which can be classified as follows: The retrieval quality of PhytoDOAS has to be improved by doing further tests with current PFT targets followed by introducing more PFT reference spectra into the retrieval. The global distribution of dinoflagellates retrieved by PhytoDOAS must be compared with an appropriate dataset of this taxonomic group. The PhytoDOAS method will be applied to the whole available SCIAMACHY data (2002–2012) and will be validated with collocated HPLC-based in-situ PFT data. Since absorption spectra of PFTs and also of specific species show some spatial variability over the global ocean (Bricaud et al., 1995),

**Table A1.** Summation of scalar products of specific absorption spectra for all 3-element sets of major phytoplankton species for different wavelength-windows.

WL (nm)	Diat. Dino. Emil.	Diat. Dino. Cyan.	Diat. Dino. Phae.	Diat. Emil. Cyan.	Diat. Emil. Phae.	Diat. Cyan. Phae.	Dino. Emil. Cyan.	Dino. Emil. Phae.	Dino. Cyan. Phae.	Emil. Cyan. Phae.
428–560	0.0317	0.0331	0.0315	0.0320	0.0305	0.0319	0.0323	0.0308	0.0322	0.0311
428–550	0.0340	0.0354	0.0338	0.0342	0.0327	0.0341	0.0346	0.0331	0.0344	0.0333
428–540	0.0366	0.0378	0.0363	0.0367	0.0353	0.0364	0.0371	0.0356	0.0368	0.0357
428–530	0.0395	0.0406	0.0391	0.0395	0.0381	0.0391	0.0399	0.0384	0.0395	0.0384
428–522	0.0420	0.0431	0.0415	0.0420	0.0405	0.0415	0.0425	0.0409	0.0419	0.0409
428–510	0.0462	0.0471	0.0455	0.0462	0.0446	0.0455	0.0467	0.0451	0.0460	0.0451
428–500	0.0500	0.0507	0.0491	0.0499	0.0484	0.0490	0.0504	0.0489	0.0495	0.0488
428–496	0.0516	0.0521	0.0506	0.0514	0.0499	0.0505	0.0519	0.0505	0.0510	0.0503
428–490	0.0538	0.0542	0.0528	0.0535	0.0522	0.0525	0.0541	0.0528	0.0531	0.0525
428–486	0.0572	0.0574	0.0561	0.0569	0.0556	0.0558	0.0574	0.0561	0.0563	0.0557

one of the main future tasks is to obtain and represent PhytoDOAS retrievals on a regional basis. This demands applying PhytoDOAS separately on different oceanic biogeochemical provinces (e.g. based on Longhurst, 1998), for each of which the input phytoplankton absorption spectra must be selected beforehand from the in-situ measurements in that province. Furthermore, a climatology on the PFT distribution can be developed based on the PhytoDOAS SCIAMACHY results. The resulting dataset will be useful to various marine biogeochemical and ecosystem studies and models. Also the global spatially and temporally resolved PFT dataset can be used as a basis for a phytoplankton-specific absorption climatology and then be applied to improve the common ocean color chl *a* retrievals. The PhytoDOAS dataset may also help to improve global estimates of biogenic gas emissions resulting from oceanic phytoplankton, e.g. dimethylsulphide (DMS) production, which is mostly connected to the abundance of coccolithophores (Keller, 1989).

## Appendix A

### Spectral orthogonality test

Due to the limitations imposed by the spectral correlation, it is necessary to determine and optimize some factors when running a *multi-target fit*, in order to receive an acceptable fit quality: as mentioned in Sect. 2.4, along with the *fourth-derivative* analysis, the optimized set of PFT targets and respective wavelength window were investigated via an orthogonality test, which is explained in the following. The specific absorption spectra of different sets of PFTs were analyzed based on the orthogonality condition for a vector space or the concept of *linear independence* in linear algebra: a set of vectors of the same dimensions (e.g. denoted by  $V_i$ ,  $V_j$ , etc.) is linearly independent if for each pair of them the *scalar product* is equal to zero, which means that they are all mutu-

ally orthogonal and could be regarded as a basis for the vector space. This condition can be mathematically demonstrated as follows:

$$V_i \cdot V_j = \sum_{k=1}^n V_{ik} V_{jk} = 0, \quad (\text{A1})$$

where  $V_{ik}$  (with  $k = 1, 2, \dots, n$ ) represents elements of the vector  $V_i$ . However, what happens if two vectors are not linearly independent? Simply, the vectors are not orthogonal – rather they are correlated; the more they are correlated, the higher will be the amount of their *scalar product*. Regarding the absorption spectra (over the same wavelength range with the same gridding intervals) as different vectors of the same dimension, the concept of *linear dependence* can be used for quantifying the existing correlations. Let us denote a given absorption vector as  $A_i$ , for which the vector elements are introduced by  $A_{ik}$  (with  $k = 1, 2, \dots, n$ ). Knowing that the correlations do exist, it can be argued that, over a specific wavelength window, a set of vectors (among several vectors) is less correlated if the summation of their mutual linear products is less than the respective summations of the other sets. Therefore, taking each specific absorption spectrum as a vector, the following quantity should be computed and compared for all available sets of PFTs, with a certain number of elements:

$$\sum_{i,j=1}^3 A_i \cdot A_j = \sum_{i,j=1}^3 \sum_{k=1}^n A_{ik} A_{jk}, \quad (\text{A2})$$

where  $i \neq j$  and  $A_i$  and  $A_j$  are normalized absorption vectors, which have been normalized to the length of vectors. Assuming  $a_i$  as the vector form of a given specific absorption spectrum, being normalized to the vector's length, means that  $A_i = \frac{a_i}{a_i}$ . Additionally, with respect to the efficiency of the *triple-target* (achieved by performing  $\chi^2$  tests), here the total number of absorption vectors has been set to three; i.e. the

above quantity is calculated for 3-element sets among major PFTs.

The quantity introduced by Eq. (A2) has been calculated for five major phytoplankton groups: diatoms, dinoflagellates, coccolithophores (*E. huxleyi*), cyanobacteria and *Phaeocystis* sp. The resulting values, computed for different wavelength windows, are presented in Table A1. According to results presented in Table A1, following information can be extracted: the linear independence, as expected, is different for various sets of PFT absorption spectra for a given wavelength range; it also varies with the operational wavelength fit-windows, indicating a decrease with widening the fit-window. However, this does not imply that the wider fit-window (implying lower dependency) would provide a better fit quality. In practice by enlarging the fit-window, the spectral overlaps of other optical components of seawater (e.g. water molecules) will also be increased. Additionally, due to the technical limitations, it is much easier to preserve the fit-window not too wide as it is confined within a single wavelength cluster of the SCIAMACHY sensor. For instance, the sensor has different integration times (for collecting data) for its different wavelength bands, leading to quite different ground-pixel sizes. That is also the reason that all fit-windows start from 428 nm and cannot go beyond 530 nm (the upper limit of the cluster 15 of SCIAMACHY data).

This leads to finding a compromise between the mentioned effects and the factor of spectral independence, as an optimized fit-window. The final range of fit-window is also dependent on spectral behavior of the selected set of PFTs, which needs a more precise analysis of the respective spectra (see below). Conversely, narrowing the width of fit-window reduces the amount of specific absorption features, which in turn causes the loss of spectral information needed for the well-functioning of the retrieval process.

As a consequence, when comparing the quantities of Eq. (A2) to decide about the best set of PFTs for a simultaneous fit (within a given wavelength window), according to Table A1, the introduced factor for the spectral independence shows significantly lower value for the selected fit-window of 429–521 nm, compared to the fit-window used in Bracher et al. (2009) (i.e. 429–495 nm).

**Supplementary material related to this article is available online at: <http://www.ocean-sci.net/8/1055/2012/os-8-1055-2012-supplement.zip>.**

*Acknowledgements.* We are thankful to ESA, DLR, and the SCIAMACHY Quality Working Group (SQWG) for providing us with SCIAMACHY level-1 data. We thank NASA-GSFC for NOBM modeled data (Giovanni project) and the MODIS RGB picture. We thank Erika Allhusen for her support during analysis of in-situ data. The authors are grateful to the IPEV/Aerotraces program for logistical support during the *Marion Dufresne* cruise and also to Ruediger Roettgers for his valuable comments on phytoplankton ecology and ocean optics. Funding was provided by the HGF Innovative Network Funds (Phytooptics). Funding for TD was supplied via the EU project SHIVA-226224-FP7-ENV-2008-1. This work is a contribution to the “Earth System Science Research School (ESSReS)”, an initiative of the Helmholtz Association of German research centers (HGF) at the Alfred Wegener Institute for Polar and Marine Research.

Edited by: O. Zielinski

## References

- Aguirre-Gomez, R., Weeks, A. R., and Boxall, S. R.: The identification of phytoplankton pigments from absorption spectra, *Int. J. Remote Sens.*, 22, 315–338, 2001.
- Aiken, J., Fishwick, J. R., Lavender, S. J., Barlow, R., Moore, G., and Sessions, H.: Validation of MERIS reflectance and chlorophyll during the BENCAL cruise October, 2002: Preliminary validation of new products for phytoplankton functional types and photosynthetic parameters, *Int. J. Remote Sens.*, 28, 497–516, 2007.
- Aiken, J., Hardman-Mountford, N. J., Barlow, R., Fishwick, J., Hirata, T., and Smyth, T.: Functional links between bioenergetics and bio-optical traits of phytoplankton taxonomic groups: an overarching hypothesis with applications for ocean colour remote sensing, *J. Plankton Res.*, 30, 165–181, 2008.
- Allali, K., Bricaud, A., Babin, M., Morel, A., and Chang, P.: A new method for measuring spectral absorption coefficients of marine particles, *Limnol. Oceanogr.*, 40, 1526–1532, 1995.
- Alvain, S., Moulin, C., Dandonneau, Y., and Brèona, F. M.: Remote sensing of phytoplankton groups in case I waters from global SeaWiFS imagery, *ELSEVIER, Deep-Sea Res. Pt. I*, 52, 1989–2004, 2005.
- Alvain, S., Moulin, C., Dandonneau, Y., and Loisel, H.: Seasonal distribution and succession of dominant phytoplankton groups in the global ocean: A satellite view, *Global Biogeochem. Cy.*, 22, GB3001, doi:10.1029/2007GB003154, 2008.
- Balch, W. M., Gordon, H. R., Bowler, B. C., Drapeau, D. T., and Booth, E. S.: Calcium carbonate measurements in the surface global ocean based on Moderate-Resolution Imaging Spectroradiometer data, *J. Geophys. Res.*, 110, C07001, doi:10.1029/2004JC002560, 2005.
- Bartlett, J., Voss, K., Sathyendranath, S., and Vodacek, A.: Raman scattering by pure water and seawater, *Appl. Opt.*, 37, 3324–3332, doi:10.1364/AO.37.003324, 1998.
- Bovensmann, H., Burrows, J. P., Buchwitz, M., Frerick, J., Noël, S., Rozanov, V. V., Chance, K. V., and Goede, A. P. H.: SCIAMACHY – Mission objectives and measurement modes, *J. Atmos. Sci.*, 56, 127–150, 1999.
- Bracher, A., Vountas, M., Dinter, T., Burrows, J. P., Röttgers, R., and Peeken, I.: Quantitative observation of cyanobacteria and

- diatoms from space using PhytoDOAS on SCIAMACHY data, *Biogeosciences*, 6, 751–764, doi:10.5194/bg-6-751-2009, 2009.
- Bricaud, A., Morel, A., and Prieur, L.: Absorption by dissolved organic matter of the sea (yellow substance) in the UV and visible domains, *Limnol. Oceanogr.*, 26, 43–53, 1981.
- Bricaud, A., Babin, M., Morel, A., and Claustre, H.: Variability in the chlorophyll-specific absorption coefficients of natural phytoplankton: Analysis and parameterization, *J. Geophys. Res.*, 100, 321–332, 1995.
- Carder, K. L., Steward, R. G., Harvey, G. R., and Ortner, P. B.: Marine humic and fulvic acids: their effects on remote sensing of ocean chlorophyll, *Limnol. Oceanogr.*, 34, 68–81, 1989.
- Falkowski, P. G., Barber, R. T., and Smetacek, V.: Biogeochemical Controls and Feedbacks on Ocean Primary Production, *Science*, 281, 200–206, 1998.
- Gordon, H. R., Brown, O. B., and Jacobs, M. M.: Computed Relationships Between the Inherent and Apparent Optical Properties of a Flat Homogeneous Ocean, *Appl. Opt.*, 14, 417–427, 1975.
- Gregg, W. W. and Casey, N. W.: Modeling coccolithophores in the global oceans, *Deep Sea Res. Pt. II*, 54, 447–477, 2007.
- Gregg, W. W., Ginoux, P., Schopf, P. S., and Casey, N. W.: Phytoplankton and iron: validation of a global three-dimensional ocean biogeochemical model, *Deep-Sea Res. Pt. II*, 50, 3147–3169, 2003.
- Hirata, T., Hardman-Mountford, N. J., Brewin, R. J. W., Aiken, J., Barlow, R., Suzuki, K., Isada, T., Howell, E., Hashioka, T., Noguchi-Aita, M., and Yamanaka, Y.: Synoptic relationships between surface Chlorophyll-a and diagnostic pigments specific to phytoplankton functional types, *Biogeosciences*, 8, 311–327, doi:10.5194/bg-8-311-2011, 2011.
- Hoffmann, L. J., Peeken, I., Lochte, K., Assmy, P., and Veldhuis, M.: Different reactions of Southern Ocean phytoplankton size-classes to iron fertilization, *Limnol. Oceanogr.*, 51, 1217–1229, 2006.
- IOCCG report 5. Remote Sensing of Inherent Optical Properties: Fundamentals, Tests of Algorithms, and Applications. Lee, Z.-P. (ed.), Reports of the International Ocean-Colour Coordinating Group, No. 5, IOCCG, Dartmouth, Canada, 2006.
- Kahru, M., Mitchell, B. G., Diaz, A., Miura, M.: MODIS Detects a Devastating Algal Bloom in Paracas Bay, Peru, *EOS Trans. AGU*, 85, 465–472, 2004.
- Keller, M. D.: Dimethyl sulfide production and marine phytoplankton: the importance of species composition and cell size, *Biol. Oceanogr.*, 6, 375–382, 1989.
- Kirk, J. T. O.: *Light and Photosynthesis in Aquatic Ecosystems*, 2nd Edition, Cambridge University Press, ISBN 0 521 45353 4, Cambridge, UK, 1994.
- Longhurst A. R.: *Ecological Geography of the Sea*, Academic Press, San Diego, California, 1998.
- Mackey, M. D., Mackey, D. J., Higgins, H. W., and Wright, S. W.: CHEMTAX – a program for estimating class abundances from chemical markers: Application to HPLC measurements of phytoplankton, *Mar. Ecol.-Prog. Ser.*, 14, 265–283, 1996.
- McClain, C. R.: A Decade of Satellite Ocean Color Observations, *Annu. Rev. Mar. Sci.*, 1, 19–42, 2009.
- Mitchell, B. G., Bricaud, A., Carder, K., Cleveland, J., Ferrari, G. M., Gould, R., Kahru, M., Kishino, M., Maske, H., Moisan, T., Moore, L., Nelson, N., Phinney, D., Reynolds, R. A., Sosik, H., Stramski, D., Tassan, S., Trees, C., Weidemann, A., Wieland, J. D., and Vodacek, A.: Determination of spectral absorption coefficients of particles, dissolved material and phytoplankton for discrete water samples, in: *Ocean Optics Protocols For Satellite Ocean Color Sensor Validation, Revision 2*, edited by: Fargion, G. S., Mueller, J. L., and McClain, C. R., NASA Technical Memorandum 2000-209966, Chapter 12, 125–153, 2000.
- Morel, A.: Consequences of a *Synechococcus* bloom upon the optical properties of oceanic Case 1 waters, *Limnol. Oceanogr.*, 42, 1746–1754, 1997.
- Nair, A., Sathyendranath, S., Platt, T., Morales, J., Stuart, V., Forget, M., Devred, E., and Bouman, H.: Remote sensing of phytoplankton functional types, *Remote Sens. Environ.*, 112, 3366–3375, 2008.
- O'Reilly, J. E., Maritorena, S., Mitchell, B. G., Siegel, D. A., Carder, K. L., Garver, S. A., Kahru, M., and McClain, C.: Ocean color chlorophyll algorithms for SeaWiFS, *J. Geophys. Res.*, 103, 24937–24953, 1998.
- Perner, D. and Platt, U.: Detection of nitrous acid in the atmosphere by differential optical absorption, *Geophys. Res. Lett.*, 93, 917–920, 1979.
- Platt, T. and Sathyendranath, S.: Oceanic primary production: Estimation by remote sensing at local and regional scales, *Science*, 241, 1613–1620, 1988.
- Raven, J. A. and Falkowski, P. G.: Oceanic sinks for atmospheric CO<sub>2</sub>, *Plant Cell Environ.*, 22, 741–755, 1999.
- Roettgers, R., Haese, C., and Doerffer, R.: Determination of the particulate absorption of microalgae using a point-source integrating-cavity absorption meter: verification with a photometric technique, improvements for pigment bleaching, and correction for chlorophyll fluorescence, *Limnol. Oceanogr.-Meth.*, 5, 1–12, 2007.
- Sadeghi, A., Dinter, T., Vountas, M., Taylor, B., Altenburg-Soppa, M., and Bracher, A.: Remote sensing of coccolithophore blooms in selected oceanic regions using the PhytoDOAS method applied to hyper-spectral satellite data, *Biogeosciences*, 9, 2127–2143, doi:10.5194/bg-9-2127-2012, 2012.
- Sathyendranath, S., Watts, L., Devred, E., Platt, T., Caverhill, C., and Maass, H.: Discrimination of diatoms from other phytoplankton using ocean-colour data, *Mar. Ecol.-Prog. Ser.*, 272, 59–68, 2004.
- Siegel, H., Ohde, T., Gerth, M., Lavik, G., and Leipe, T.: Identification of coccolithophore blooms in the SE Atlantic Ocean off Namibia by satellites and in-situ methods, *Cont. Shelf Res.*, 27, 258–274, 2007.
- Toby Tyrrell, T. and Merico, A.: *Emiliania huxleyi*: bloom observations and the conditions that induce them, in: *Coccolithophores. From Molecular Processes to Global Impact*, edited by: Thierstein, H. R. and Young, J. R., Springer-Verlag Berlin Heidelberg, 75–97, 2004.
- Vasilkov, A. P., Joiner, J., Gleason, J., and Bhartia, P.: Ocean Raman scattering in satellite backscatter UV measurements, *Geophys. Res. Lett.*, 29, 1837, doi:10.1029/2002GL014955, 2002.
- Vountas, M., Rozanov, V., and Burrows, J. P.: Ring effect: Impact of rotational Raman scattering on radiative transfer in earth's atmosphere, *J. Quant. Spectrosc. Ra.*, 60, 943–961, 1998.
- Vountas, M., Richter, A., Wittrock, F., and Burrows, J. P.: Inelastic scattering in ocean water and its impact on trace gas retrievals from satellite data, *Atmos. Chem. Phys.*, 3, 1365–1375, doi:10.5194/acp-3-1365-2003, 2003.

Vountas, M., Dinter, T., Bracher, A., Burrows, J. P., and Sierk, B.: Spectral studies of ocean water with space-borne sensor SCIAMACHY using Differential Optical Absorption Spectroscopy (DOAS), *Ocean Sci.*, 3, 429–440, doi:10.5194/os-3-429-2007, 2007.

Ye, Y., Voelker, C., Bracher, A., Taylor, B. B., and Wolf-Gladrow, D. A.: Environmental controls on  $N_2$  fixation by *Trichodesmium* in the tropical eastern North Atlantic Ocean –A model-based study, *Deep-Sea Res. Pt. I*, 64, 104–117, 2012.