



**HAL**  
open science

## A species-dependent bio-optical model of case I waters for global ocean color processing

Séverine Alvain, Cyril Moulin, Yves Dandonneau, Hubert Loisel,  
Francois-Marie Breon

► **To cite this version:**

Séverine Alvain, Cyril Moulin, Yves Dandonneau, Hubert Loisel, Francois-Marie Breon. A species-dependent bio-optical model of case I waters for global ocean color processing. *Deep Sea Research Part I: Oceanographic Research Papers*, 2006, 53 (5), pp.917-925. 10.1016/j.dsr.2006.01.011 . hal-00157049

**HAL Id: hal-00157049**

**<https://hal.science/hal-00157049>**

Submitted on 25 Jun 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A SPECIES-DEPENDENT BIO-OPTICAL MODEL OF CASE I WATERS FOR GLOBAL OCEAN COLOR PROCESSING

S. Alvain<sup>1</sup>, C. Moulin<sup>\*1</sup>, Y. Dandonneau<sup>2</sup>, H. Loisel<sup>3</sup>, F.-M. Bréon<sup>1</sup>

<sup>1</sup>IPSL/LSCE, CNRS-CEA, Gif-sur-Yvette, France

<sup>2</sup>IPSL/LODYC, IRD-CNRS-UPMC, Paris, France

<sup>3</sup>ELICO, ULCO, Wimereux, France

\* Corresponding author, E-mail: Cyril.Moulin@cea.fr

**Abstract.** The PHYSAT method, which enables identification of four different phytoplankton groups from their impact on the normalized water-leaving radiance (nLw) spectra, is applied to coincident *in situ* measurements of both chlorophyll *a* concentration (Chl *a*) and nLw. Observations show that measurements acquired in waters dominated by haptophytes, diatoms and *Synechococcus*-like cyanobacteria have optical properties that deviate significantly from the mean OC4V4 bio-optical model, which is currently used as the standard for global ocean color processing. A specific OC4v4-like relationship, i.e., Chl *a* as a 4<sup>th</sup>-order polynomial of the “maximum band ratio”, was fitted for each dominant phytoplankton group identified by PHYSAT. The resulting OC4-SD bio-optical model thus starts from a classification of the dominant phytoplankton group, followed by a species-dependent estimate of Chl *a*. It has been applied to global daily SEAWIFS data of the year 2001. Monthly mean maps of Chl *a* derived from OC4-SD or OC4v4 show large regional differences that can reach 50% at high latitudes. The new algorithm leads to lower concentrations in regions where the standard model retrievals are known to be too high, such as the Mediterranean Sea. Conversely, higher concentrations are retrieved in regions dominated by diatom blooms, such as the northern North Atlantic in summer, where previous studies have demonstrated a low bias in standard SEAWIFS Chl *a*.

**Keywords:** Ocean color, Remote sensing, Oceanic Primary Production

## 1. Introduction

For about 10 years, satellite missions dedicated to ocean color (OCTS, POLDER, SEAWIFS, MODIS) have been changing our understanding of the biological ocean by providing a quasi-global daily monitoring of the chlorophyll *a* concentration (Chl *a*, in  $\text{mg}\cdot\text{m}^{-3}$ ) in surface waters. These satellite data enable us to address a wide range of scientific issues, from the study of local marine biological processes to the estimate of the global ocean productivity (see McClain *et al.*, 2004). The standard ocean color products are normalized water-leaving radiances (nLw; in  $\text{mW cm}^{-2} \mu\text{m}^{-1} \text{sr}^{-1}$ ) at different wavelengths in the blue and green parts of the solar spectrum (e.g., at 412, 443, 490, 510 and 555 nm for SEAWIFS), and Chl *a*, which is estimated from nLw values and a bio-optical model.

The large variety of phytoplankton assemblages and of oceanographic conditions encountered in the global ocean coupled to the paucity of *in situ* data oblige us to define mean empirical relationships that are assumed to be valid at regional to global scales. A single bio-optical model (OC4V4; O'Reilly *et al.*, 2000) is for instance currently used to generate global maps of Chl *a* from both SEAWIFS and MODIS imagery. This model was derived by fitting a mean relationship from a large dataset of coincident nLw and Chl *a in situ* measurements acquired in various waters. However, the large dispersion of *in situ* data around this mean relationship (O'Reilly *et al.*, 2000) shows that this approach is unlikely to retrieve Chl *a* with an accuracy better than 30%. The observed dispersion in water optical properties for a given Chl *a* is usually attributed to the optical impact of yellow substances or of ecosystem characteristics.

Even though OC4V4 performs satisfactorily at the global scale (McClain *et al.*, 2004), several regional studies have emphasized significant biases between satellite and *in situ* Chl *a* estimates. These studies led to various adaptations of OC4V4, such as those developed for the oligotrophic Mediterranean waters (Bricaud *et al.*, 2002; D'Ortenzio *et al.*, 2002) or for the productive Antarctic waters (Arrigo *et al.*, 1998; Dierssen and Smith, 2000). Such regional approaches are, however, difficult to conciliate with global satellite data processing. Besides, they do not provide any biological or physical interpretation of the observed variability in water optical properties. More recently, bio-

optical models were developed to account for the specific optical properties of species like diatoms (Cota *et al.*, 2003; Sathyendranath *et al.*, 2004), *Synechococcus* (Morel *et al.*, 1997) and *Trichodesmium* (Subramaniam *et al.*, 2002). However, these species-dependent bio-optical models cannot be used for the processing of global satellite data, as they require some prior information on the dominant phytoplankton group.

The PHYSAT method was developed to identify the dominant phytoplankton species from ocean color measurements (Alvain *et al.*, 2005). This classification relies on the difference between the shape of the measured nLw spectrum between 412 and 555 nm and that of a standard. The analysis of coincident nLw spectra and *in situ* pigment inventories performed by Alvain *et al.* (2005) has shown that four phytoplankton groups (*i.e.*, *Prochlorococcus*, *Synechococcus*-like cyanobacteria (SLC), haptophytes and diatoms) can be identified. Here we applied PHYSAT to the NOMAD *in situ* dataset (Werdell and Bailey, 2005) to derive species-dependent OC4V4-like relationships between normalized water-leaving radiances and Chl *a*. An independent *in situ* dataset was used to demonstrate the improvement in terms of Chl *a* accuracy. These relationships were finally applied to global SEAWIFS data for the year 2001.

## 2. Analysis of *in situ* measurements

The NOMAD dataset has recently been made available to the scientific community, in place of the former SEABAM dataset, to develop new bio-optical models. It includes 3466 *in situ* coincident nLw and Chl *a* measurements acquired in various waters around the globe. Before application of the PHYSAT method to this dataset, 942 measurements were discarded because at least one of the required spectral bands was missing. 372 measurements identified as Case 2 waters by the criterion of the SEAWIFS standard processing (*i.e.*, if  $nLw > 0.059 \text{ mW cm}^{-2} \mu\text{m}^{-1} \text{ sr}^{-1}$ ; Patt *et al.*, 2003) were also removed from the database. Finally, 101 Case 1 measurements were discarded because they have a Chl *a* value lower than  $0.04 \text{ mg m}^{-3}$  or larger than  $10 \text{ mg m}^{-3}$ . Note that this latter threshold is higher than that of the original PHYSAT method ( $3 \text{ mg.m}^{-3}$ ; Alvain *et al.*, 2005) to maximize the range of validity of our bio-optical model. The application of PHYSAT to the remaining 2051 *in situ* measurements led to a successful classification

in one of the four groups for 337 of them, as shown in Figure 1. Whereas measurements identified as *Prochlorococcus* and SLC seem to be located approximately in the center of the dataset in Figure 1, those identified as haptophytes and diatoms seem to be respectively in the lower and higher part of the dataset. Figure 2 shows the geographical distribution of the NOMAD measurements successfully classified with PHYSAT. As already emphasized by Werdell and Bailey (2005), most of the measurements performed at high latitudes (above 30°) were performed in coastal waters of North America.

An OC4V4-like polynomial fit (O'Reilly *et al.*, 2000) was applied to the data belonging to the four phytoplankton groups:

$$\log(\text{Chl } a) = a \log^4(r) + b \log^3(r) + c \log^2(r) + d \log(r) + e \quad (1)$$

The  $r$  parameter corresponds to the “maximum band” blue-to-green ratio, defined as:

$$r = \max(\text{Rrs}(443), \text{Rrs}(490), \text{Rrs}(510)) / \text{Rrs}(555) \quad (2)$$

Rrs is the remote sensing reflectance at a given wavelength  $\lambda$ , defined as:

$$\text{Rrs}(\lambda) = nLw(\lambda) / E_0(\lambda) \quad (3)$$

$E_0$  is the top-of-atmosphere solar irradiance. Coefficients  $a$  to  $e$  are given in Table 1 for each group, except for *Prochlorococcus*, because the result of the fitting for this species is not significantly different from OC4V4. Figure 3 confirms that the relationships fitted for diatoms and haptophytes are significantly above and below, respectively, OC4V4. This is in good agreement with the results obtained by Cota *et al.* (2003) in the Labrador Sea for both diatoms and *Prymnesiophytes* (a species that belongs to our haptophytes group). The relationship obtained for SLC data has a different behavior since it yields strongly lower (resp. slightly larger) Chl  $a$  values for Chl  $a$  above (resp. below) 0.3 mg m<sup>-3</sup>. These results suggest that the presence of different phytoplankton species, as identified by PHYSAT, explains a significant part of the overall variability of the NOMAD dataset.

### 3. Application to global SEAWIFS data

From the results obtained from the analysis of the NOMAD dataset, we defined a species-dependent bio-optical model, OC4-SD, as a two-step algorithm. First PHYSAT

is applied to the standard SEAWIFS Chl *a* and nLw spectrum to attempt the identification of a phytoplankton group. If the identification is successful, the suitable bio-optical model (see Table 1) is used to compute an improved Chl *a* value. Otherwise, if the identification is unsuccessful or leads to Prochlorococcus, or if the first guess Chl *a* is outside the validity range (see Table 1), the standard SEAWIFS Chl *a* value is kept.

SeaWiFS daily level-3 binned products at a resolution of 1/12° were processed for the year 2001 with OC4-SD to generate global mean monthly maps of Chl *a*. As described in Alvain *et al.* (2005), the first step of the processing rejects pixels with an aerosol optical thickness greater than 0.15 or with a Chl *a* outside the 0.04-3 mg.m<sup>-3</sup> range. PHYSAT is then applied to all valid pixels of the daily level-3 products, with an average of about 50% of successfully classified pixels. Both OC4v4 and OC4-SD were finally applied to every classified pixel of daily level-3 products, which were then used to generate monthly mean global maps of Chl *a*.

Figure 4 compares maps of Chl *a* from both bio-optical models for four months in 2001. The differences are well organized at the global scale and are generally greater than ±20% at mid and high latitudes. Higher Chl *a* values (up to 60%) are found in regions where diatoms are frequently observed. This result is consistent with previous studies that demonstrate the underestimation of the standard SEAWIFS Chl *a* in diatom-dominated waters (Sathyendranath *et al.*, 2003; Dierssen and Smith, 2000). Conversely, lower Chl *a* are found in some specific areas known for their too high OC4V4 Chl *a*, such as the Mediterranean Sea (Bricaud *et al.*, 2002; D'Ortenzio *et al.*, 2002). OC4-SD Chl *a* values are slightly larger in most oligotrophic tropical waters because of the quasi-permanent presence of *Synechococcus*-like cyanobacteria.

#### **4. Validation of the OC4-SD Chl *a***

We tentatively validate the OC4-SD bio-optical model by using independent datasets of *in situ* Chl *a* measurements acquired in case 1 waters and for which clear-sky SEAWIFS pixels of the same day are available within ±9 km. The validation is based

on a set of 1045 nearly coincident SEAWIFS nLw and *in situ* Chl *a* measurements, acquired during several cruises presented in Table 2. PHYSAT successfully classified 391 measurements in one of the four phytoplankton groups, out of which only 182 as haptophytes, diatoms or SLC. The large number of data identified as *Prochlorococcus* was expected since a lot of *in situ* measurements were performed in oligotrophic waters, mainly during the GeP&Co cruises, where this species dominates.

The validity range criteria of Table 1 lead to further rejection, limiting the dataset of haptophytes, diatoms or SLC observations to 111 cases. Although very reduced from the original validation database, these 111 cases sample several regions and seasons, as shown in Figure 5 and Table 2. Contrary to the NOMAD dataset (see Figure 2), this validation dataset includes mostly remote open-ocean measurements, both in the North Atlantic and in the southern Pacific. The comparison of both OC4V4 and OC4-SD SEAWIFS Chl *a* with *in situ* Chl *a* is shown in Figure 6. The species-dependent algorithm described above leads to both a higher correlation coefficient and a regression slope closer to 1 than the original OC4V4 algorithm. These statistical parameters and the analysis of Figure 6 demonstrate that the OC4-SD bio-optical model yields significantly less biased and dispersed Chl *a* values than OC4V4 for pixels successfully classified, and obviously does not degrade the results for other pixels for which the OC4V4 Chl *a* is kept.

## 5. Discussion

We have seen in sections 2 and 3 that the percentage of successful PHYSAT identification of a dominant phytoplankton group differs markedly according to the dataset: it is of 16% for the NOMAD dataset (see section 2), of 37% for the validation dataset presented in section 3, and reaches about 50% of the SEAWIFS pixels at the global scale. Table 3 shows that this discrepancy is mostly due to differences between the various datasets themselves and not to a bias introduced by PHYSAT. As already shown in Figures 2 and 5, the valid measurements in the NOMAD dataset include many more measurements (35%) performed at high latitudes and high Chl *a*, i.e. in highly dynamical and productive coastal (but not case 2) waters, than in our validation dataset

(16%) or in global SEAWIFS data (4%). In general, the repartition of valid measurements shown in Table 3 clearly demonstrates that our validation dataset, and even more the GeP&CO dataset, is much more representative of the global ocean as observed by SEAWIFS than the NOMAD dataset. Note that the GeP&CO experiment was designed to sample the maximum of water types at different seasons.

Table 3 also shows that the percentage of successful PHYSAT classification is always lower in regions of high Chl a. This suggests that these rich waters, which were poorly sampled during the GeP&CO cruises used to develop PHYSAT, may contain phytoplankton groups that are not identified with our method. It is, however, noteworthy that for all water types, there are significantly fewer successfully classified measurements in the NOMAD dataset, which relies on in situ optical measurements, than in the two other datasets, which both rely on SEAWIFS remote-sensing measurements. It is possible that some slight intrinsic biases exist between the two kinds of optical measurements, for example because of the atmospheric correction of SEAWIFS data. However it is also likely that once again the geographical distributions and the sampled biological conditions from one dataset to the other control the percentage of successful classification.

## 6. Conclusion

Identification of the dominant phytoplankton group from remote sensing measurement is sometimes possible based on the spectral shape of the reflectance spectrum, as demonstrated in Alvain *et al.* (2005). The information on the phytoplankton group can then be used for a better estimate of Chl a concentration. We have used an in-situ dataset of Chl a concentration and water leaving radiance spectra (NOMAD; Werdell and Bailey, 2005) to derive a species-dependent bio-optical model, OC4-SD, suitable for the processing of global ocean color satellite data. The differences between the mean bio-optical relationships for the various species, up to 50%, explain part of the variability found in the whole dataset. When haptophytes, diatoms or SLC are recognized from the remote sensing data provided by ocean color satellites, a specific bio-optical relationship shall be applied instead of the standard OC4V4 model.

The processing of SEAWIFS data results in monthly mean chlorophyll *a* concentrations that have the same general patterns as with the standard algorithm, but with large-scale differences. In particular, the monthly mean Chl *a* is decreased by 20 to 30% in regions dominated by haptophytes while it is increased by 50 to 60% in regions of diatom blooms. Most of these changes are consistent with known deficiencies of the SEAWIFS standard product, i.e. too high Chl *a* in regions such as the Mediterranean Sea, and an underestimate of Chl *a* for diatom blooms. The validation of OC4-SD based on 111 independent data values confirms the general improvement in Chl *a* retrievals compared to OC4V4. This new bio-optical algorithm can be applied as is to other recent sensors that provide measurements in spectral bands similar to those of SEAWIFS. Providing that it is also applicable to the CZCS dataset, OC4-SD would facilitate the interpretation of observed long-term trends in global chlorophyll *a* concentrations (Antoine *et al.*, 2005; Gregg *et al.*, 2005) by showing whether these trends are related to changes in ecosystem structure or in primary productivity.

## **7. Acknowledgements**

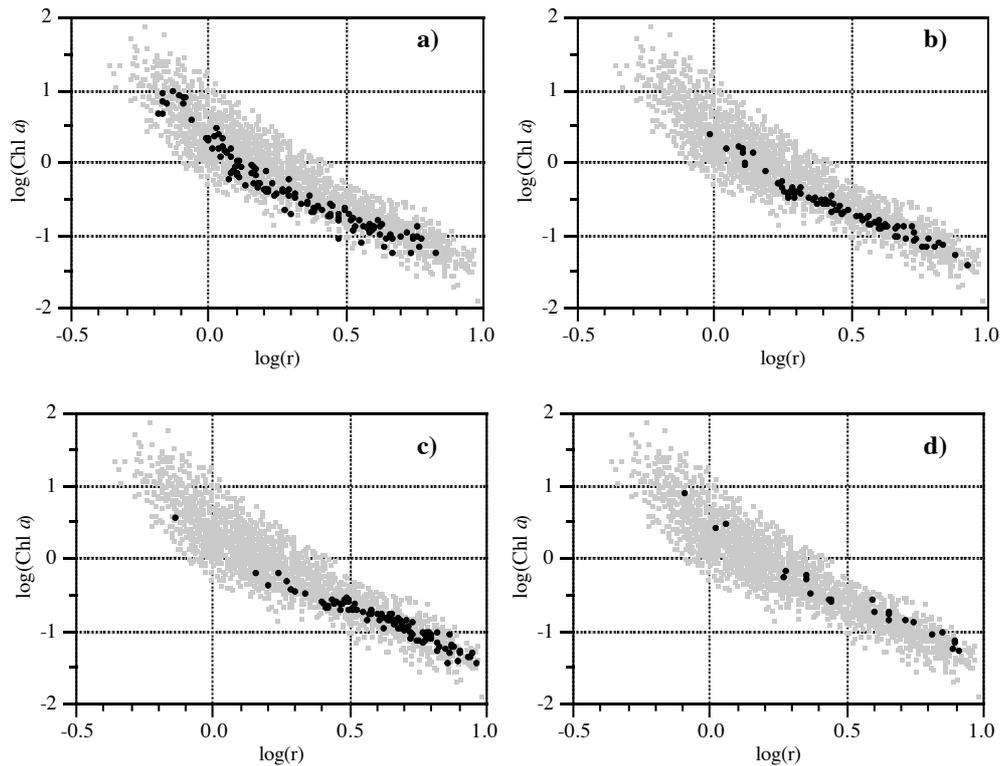
The authors would like to thank all participants in the NASA SIMBIOS program as well as all voluntary contributors for collecting data that have been assembled in the NOMAD dataset (<http://seabass.gsfc.nasa.gov/cgi-bin/nomad.cgi>). The authors also would like to thank the NASA SeaWiFS Project and the NASA/GSFC/DAAC for the production and distribution of SEAWIFS data, respectively (<http://daac.gsfc.nasa.gov/data/dataset/SEAWIFS>). We are also grateful to all organizers and observers of Gep&CO, OISO, Skogafoss, Toucan , Colibri, Kiwi and Dyfamed ocean sampling campaigns used for the validation of OC-SD, supported by CNES - TAOB/GeP&SIMBAD, PNEC, INSU/CNRS and IRD.

## 8. References

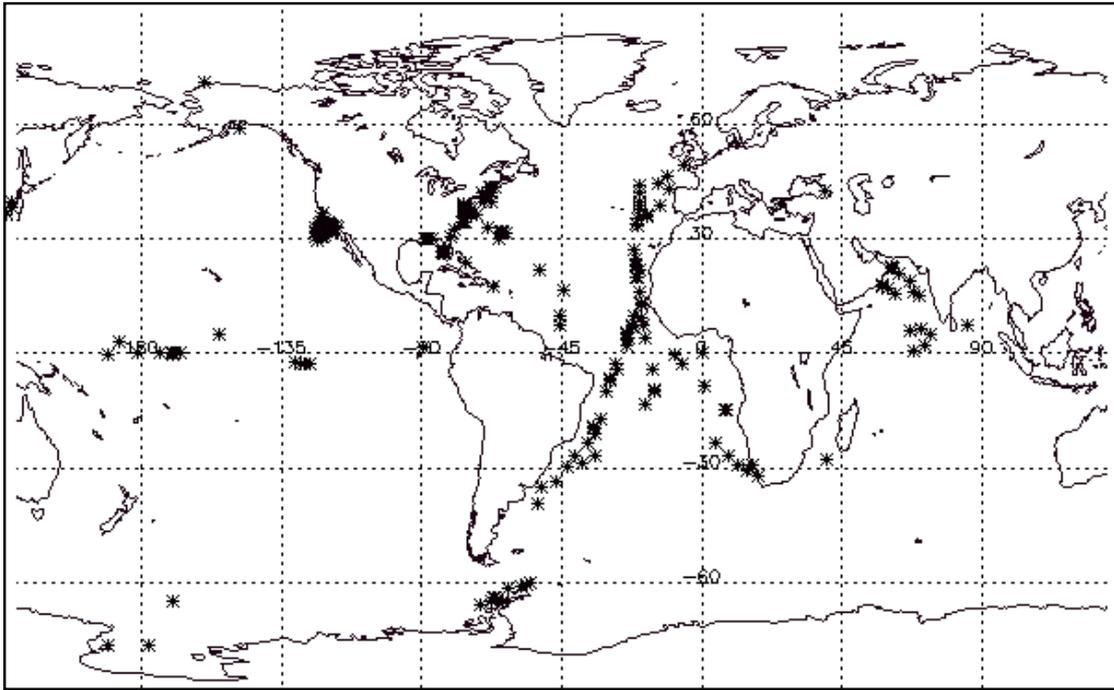
- Alvain. S., Moulin C., Dandonneau Y. and Bréon F.M. ,Remote sensing of phytoplankton groups in case 1 waters from global SeaWiFS imagery, *Deep Sea Res. I*, **52**, 1989-2004, 2005.
- Antoine, D., Morel, A., Gordon, H. R., Banzon, V. F., and Evans, R. H., Bridging ocean color observations of the 1980s and 2000s in search of long-term trends, *J. Geophys. Res.*, **110**, C06009, doi:10.1029/2004JC002620, 2005.
- Arrigo, K. R., Robinson D. H., Worthen D. L., Schieber B. and Lizotte M. P., Bio-optical properties of the southwestern Ross Sea, 1998. , *J. Geophys. Res.*, **103**, 21683-21695, 1998.
- Bricaud, A., Bosc E. and Antoine D., Algal biomass and sea surface temperature in the Mediterranean Basin. Intercomparison of data from various satellite sensors, and implications for primary production estimates. *Remote Sens. Environ.*, **81**, 163-178, 2002.
- Cota, G. F., Harrison W. G., Platt T., Sathyendranath S., and Stuart V., Bio-optical properties of the Labrador Sea, *J. Geophys. Res.*, **108**(C7), 3228, 2003.
- Dandonneau, Y., Deschamps P.-Y., Nicolas J.-M., Loisel H., Blanchot J., Montel Y., Thieuleux F., and Bécu G., Seasonal and interannual variability of ocean color and composition of phytoplankton communities in the North Atlantic, Equatorial Pacific and South Pacific, *Deep-Sea Res. II*, **51**, 303-318, 2004.
- Dierssen, H. M. and Smith R.C., Bio-optical properties and remote sensing ocean color algorithms for Antarctic Peninsula waters, *J. Geophys. Res.*, **105**, 26301-26312, 2000.
- D'Ortenzio, F., Marullo S., Ragni M., Ribera d'Alcala M., Santoleri R.. Validation of empirical SeaWiFS algorithms for chlorophyll-a retrieval in the Mediterranean Sea. A case study for oligotrophic seas, *Remote Sens. Environ.*, **82**, 79-94, 2002.

- Gregg W. W., Casey, N. W., and McClain C. R., Recent trends in global ocean chlorophyll, *Geophys. Res. Lett.*, **32**, L03606, doi:10.1029/2004GL021808, 2005.
- McClain, C. R., Feldman G. C., and Hooker S. B., An overview of the SeaWiFS project and strategies for producing a climate research quality global ocean bio-optical time series, *Deep Sea Res. II*, **51**, 5-42, 2004.
- Morel, A., Consequences of a *Synechococcus* bloom upon the optical properties of oceanic Case 1 waters, *Limnol. Oceanogr.*, **42**, 1746-1754, 1997.
- O'Reilly, J. E. and 21 co-authors, Ocean color chlorophyll *a* algorithms for SeaWiFS, OC2 and OC4: Version 4, in SeaWiFS Postlaunch Calibration and Validation Analyses (Part 3), NASA Tech. Memo. 2000-206892, 10, S.B. Hooker and E.R. Firestone Eds, NASA GSFC, pp 57, 2000.
- Patt, F. S. *et al.*, Algorithm Updates for the fourth SEAWIFS data reprocessing, Seawifs Postlaunch tech. Report, Vol. 22, NASA Tech. Memo. 206892, 2003.
- Sathyendranath, S., Cota G., Stuart V., Maass H., T. Platt, Remote sensing of phytoplankton pigments: a comparison of empirical and theoretical approaches, *Int. J. Remote Sens.*, **22**, 249-273, 2001.
- Sathyendranath, S., Watts L., Devred E., Platt T., Caverhill C. and Maass H., Discrimination of diatoms from other phytoplankton using ocean-colour data, *Marine ecology progress series*, **272**, 59-68, 2004.
- Stuart, V., Sathyendranath S., Platt T., Maass H. and Irwin B. D., Pigments and species composition of natural phytoplankton populations: effect on the absorption spectra, *J. Plankton Res.*, **20**, 187-217, 1998.
- Subramaniam, A., Brown C. W., Hood R. R., Carpenter E. J. and Capone D. G., Detecting *Trichodesmium* blooms in SeaWiFS imagery, *Deep Sea Res. II*, **49**, 107-121, 2002.

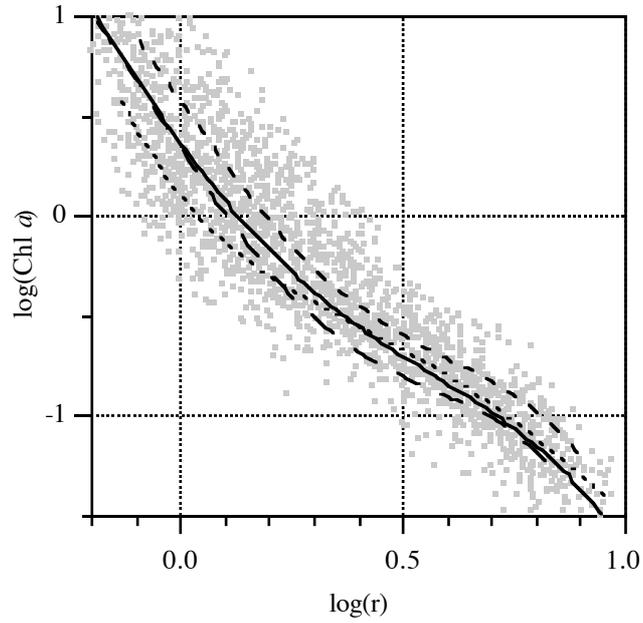
Werdell, P.J. and Bailey, S.W., An improved bio-optical data set for ocean color algorithm development and satellite data product validation, *Remote Sens. Env.*, **98**, 122-140, 2005.



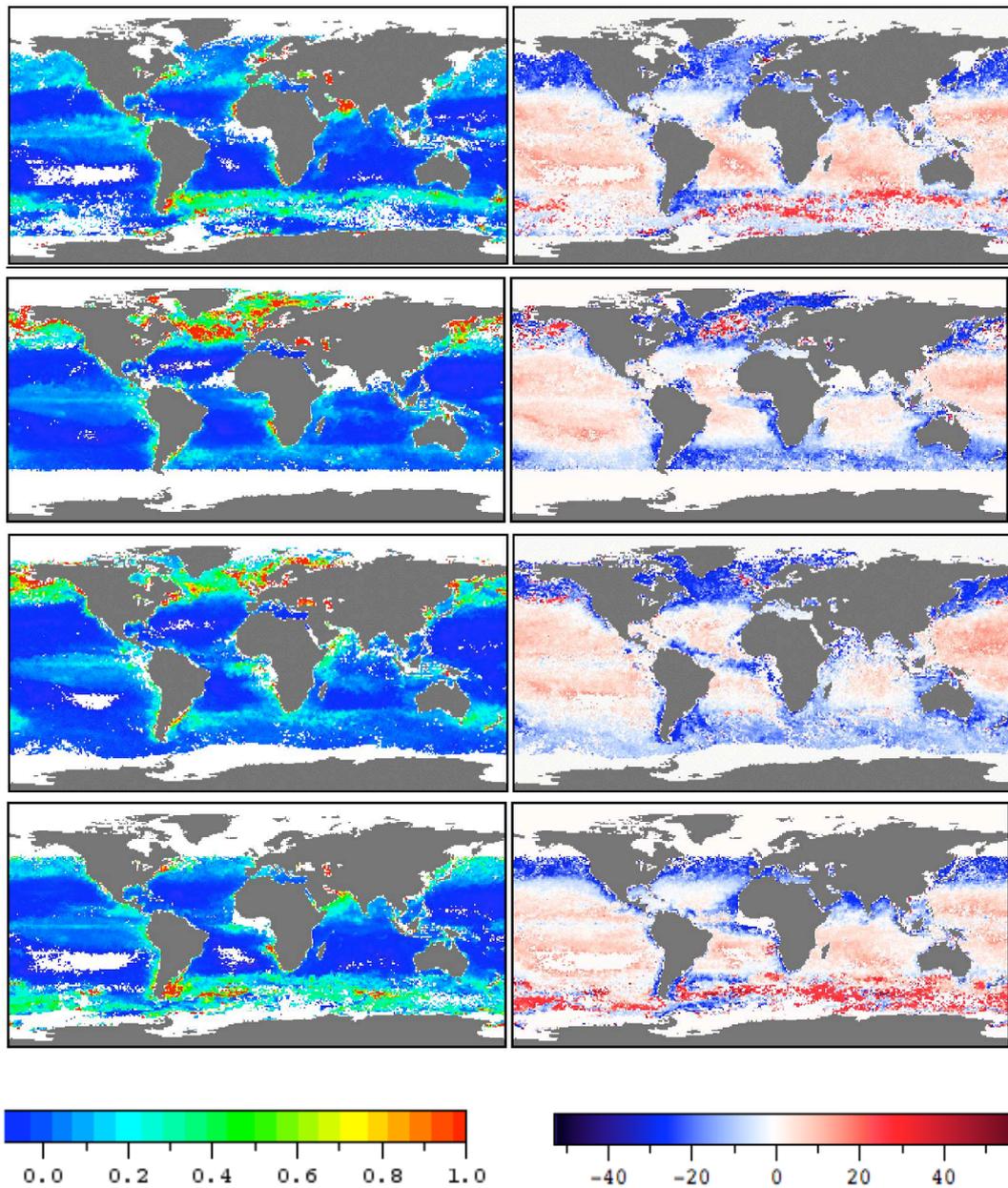
**Figure 1.** Chl *a* as a function of the “maximum band” blue-to-green ratio *r* (see Equation 2) for NOMAD measurements. Black dots represent measurements that have been classified by PHYSAT as: a) haptophytes ( $n = 130$ ), b) *Prochlorococcus* ( $n = 83$ ), c) cyanobacteria ( $n = 100$ ), and d) diatoms ( $n = 24$ ). Light grey dots in the four graphs show the whole NOMAD dataset.



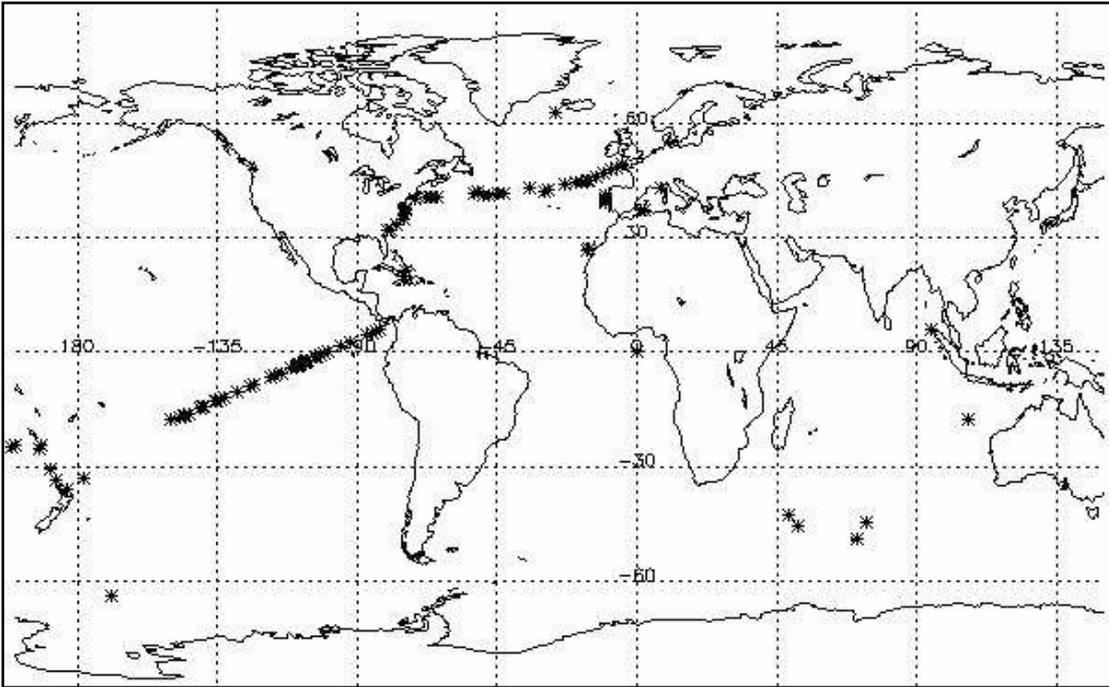
**Figure 2.** Geographical distribution of NOMAD measurements (Werdell and Bailey, 2005) successfully classified with PHYSAT.



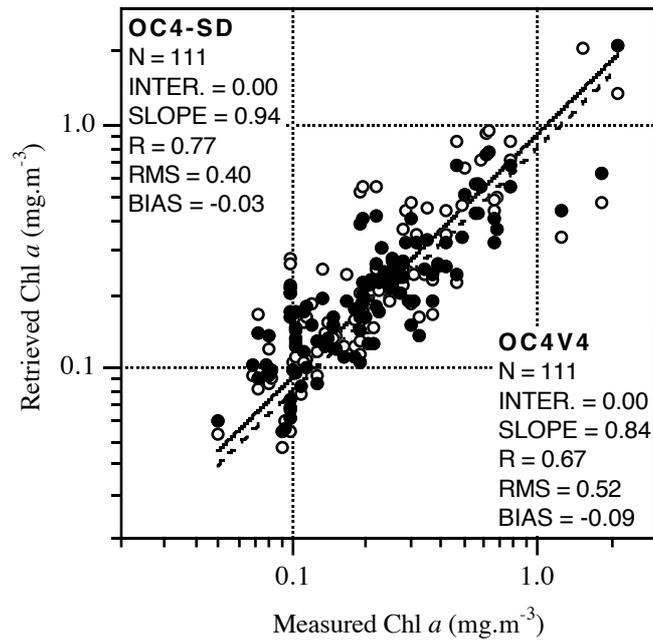
**Figure 3.** OC4V4-like relationships (see Equation 1 and Table 1) for haptophytes (long-dashed line), cyanobacteria (short-dashed line), and diatoms (medium-dashed line). The solid line shows the standard OC4V4 relationship. As in Figure 1, light grey dots show the whole NOMAD dataset.



**Figure 4.** Maps of the monthly mean OC-SD Chl a (left; in  $\text{mg}\cdot\text{m}^{-3}$ ), and of the relative difference (right; in %) between OC4-SD and OC4v4 monthly mean Chl a for March (top), June, September, and December (bottom) 2001.



**Figure 5.** Locations of the 111 *in situ* Chl *a* values used for the validation of the OC4-SD model. Data are from Gep&Co, OISO, Skogafoss, Toucan, Colibri, and Kiwi cruises and Dyfamed station (see Table 2).



**Figure 6.** Comparison of Chl *a* calculated using both OC4v4 (open circles) and OC4-SD (filled circles) with *in situ* Chl *a* for the 111 independent measurements classified by PHYSAT as Haptophytes (n = 50), SLC (n = 55) or Diatoms (n = 6). Linear regressions of the form  $\text{Chl } a(\text{retrieved}) = \alpha \cdot \text{Chl } a(\text{measured})$  are shown as dashed (OC4V4) and solid (OC4-SD) lines, and the corresponding statistics are given in the Figure.

**Table 1.** Coefficients of the various bio-optical models presented in this work (see Equation 1) and range of validity in term of Chl *a* of these models.

	OC4V4	Haptophytes	SLC	Diatoms
a	-1.532	-4.889	2.249	-4.303
b	0.649	5.096	-5.975	5.051
c	1.93	0.972	4.912	-0.333
d	-3.067	-3.430	-2.77	-3.235
e	0.366	0.341	0.104	0.58
Chl <i>a</i> validity range (mg.m <sup>-3</sup> )	0.01-30	0.06-3	0.05-4	0.06-10

**Table 2.** Description of the different datasets used to validate the OC4-SD bio-optical model.

Campaign	Number of data values used	Dominant Group	Period	Region	Principal Investigator
Gep&Co	91	1 diatom 55 SLC 35 haptophytes	08/00-08/02	North Atlantic - South Pacific	Y. Dandonneau
Dyfamed	4	4 haptophytes	09/99-09/00 07/02-08/02	Mediterranean	J-C. Marty
Kiwi 7	1	1 diatom	12/97	Austral ocean	R. Barber
Oiso 6 4	4	4 diatoms	01/01	Austral Ocean	N. Metzl
Skogafoss	1	1 haptophyte	07/02	North Atlantic	Y. Dandonneau
Colibri	8	8 haptophytes	04/03	North Atlantic	Y. Dandonneau

**Table 3.** Compared characteristics of the three datasets used in this study, with particular emphasis on their geographical and biological coverage. The SEAWIFS dataset presented in this table is composed of 20 days (from 9 to 18 June and from 6 to 15 December 2001) of global daily L3-binned data. The valid measurements are those that fulfilled the criteria defined in section 2. Only a fraction of these valid measurements are successfully classified in one of the four PHYSAT phytoplankton groups (Alvain *et al.* 2005).

		SEAWIFS	NOMAD	Validation Dataset	
				GeP&CO only	All
Number of valid measurements		49478	2051	623	1045
Number of classified measurements		31202	328	333	391
Percentage of valid measurements and percentage of successful classification, both in %	Lat > 30° and Chl <i>a</i> < 0.5 mg.m <sup>-3</sup>	38 (69)	30 (22)	28 (55)	44 (30)
	Lat > 30° and Chl <i>a</i> > 0.5 mg.m <sup>-3</sup>	3 (51)	35 (7)	17 (32)	16 (12)
	Lat < 30° and Chl <i>a</i> < 0.5 mg.m <sup>-3</sup>	58 (58)	28 (23)	54 (60)	38 (58)
	Lat < 30° and Chl <i>a</i> > 0.5 mg.m <sup>-3</sup>	1 (38)	7 (7)	1 (43)	2 (32)