Vertical distribution of chlorophyll $a$ concentration and phytoplankton community composition from in situ fluorescence profiles: a first database for the global ocean

R. Sauzède$^{1,2}$, H. Lavigne$^3$, H. Claustre$^{1,2}$, J. Uitz$^{1,2}$, C. Schmechtig$^{1,2}$, F. D’Ortenzio$^{1,2}$, C. Guinet$^4$, and S. Pesant$^{5,6}$

$^1$Laboratoire d’Océanographie de Villefranche, CNRS, UMR7093, Villefranche-Sur-Mer, France
$^2$Université Pierre et Marie Curie-Paris 6, UMR7093, Laboratoire d’océanographie de Villefranche, Villefranche-Sur-Mer, France
$^3$Istituto Nazionale di Oceanografia e di Geofisica Sperimentale, Sgonico (OGS), Italy
$^4$Centre d’Etudes Biologiques de Chizé, CNRS, Villiers en Bois, France
$^5$MARUM, Center for Marine Environmental Sciences, Universität Bremen, Bremen, Germany
$^6$PANGAEA, Data Publisher for Earth and Environmental Science, Bremen, Germany
Received: 29 March 2015 – Accepted: 6 April 2015 – Published: 21 April 2015
Correspondence to: R. Sauzède (sauzede@obs-vlfr.fr)
Published by Copernicus Publications.
Abstract

In vivo chlorophyll a fluorescence is a proxy of chlorophyll a concentration, and is one of the most frequently measured biogeochemical properties in the ocean. Thousands of profiles are available from historical databases and the integration of fluorescence sensors to autonomous platforms led to a significant increase of chlorophyll fluorescence profile acquisition. To our knowledge, this important source of environmental data has not yet been included in global analyses. A total of 268 127 chlorophyll fluorescence profiles from several databases as well as published and unpublished individual sources were compiled. Following a robust quality control procedure detailed in the present paper, about 49 000 chlorophyll fluorescence profiles were converted in phytoplankton biomass (i.e. chlorophyll a concentration) and size-based community composition (i.e. microphytoplankton, nanophytoplankton and picophytoplankton), using a method specifically developed to harmonize fluorescence profiles from diverse sources. The data span over five decades from 1958 to 2015, including observations from all major oceanic basins and all seasons, and depths ranging from surface to a median maximum sampling depth of around 700 m. Global maps of chlorophyll a concentration and phytoplankton community composition are presented here for the first time. Monthly climatologies were computed for three of Longhurst’s ecological provinces in order to exemplify the potential use of the data product. Original data sets (raw fluorescence profiles) as well as calibrated profiles of phytoplankton biomass and community composition are available in open access at PANGAEA, Data Publisher for Earth and Environmental Science.

1 Introduction

Phytoplankton biomass is generally recognized to play a key role in the global carbon cycle, stressing the need for a better understanding of its spatio-temporal distribution and variability in the global ocean. Chlorophyll \( a \) concentration is widely used as a proxy to estimate phytoplankton biomass. The geographic and temporal distribution of this proxy is already well documented at global scale thanks to synoptic remote sensing observations by Ocean Color Radiometry (OCR, McClain, 2009; Siegel et al., 2013). Nevertheless, OCR observations are restricted to the ocean surface layer, “sensing” only one fifth of the so-called euphotic layer where phytoplankton photosynthesis is realized and which can sometimes extend well below 100 m (Gordon and McCluney, 1975; Morel and Berthon, 1989). It is therefore essential to better resolve the global distribution of phytoplankton biomass on the vertical.

The vertical distribution of chlorophyll \( a \) can be estimated with greatest accuracy from the analysis of water samples by High Performance Liquid Chromatography (HPLC, Claustre et al., 2004; Peloquin et al., 2013). However, these in situ measurements are relatively scarce because their acquisition requires ship-based sampling and their analysis is costly. Moreover, because these measurements are made on water samples, the vertical resolution is generally weak (e.g. around one measurement every 10 m). The measurement of in vivo chlorophyll \( a \) fluorescence is widely used as a proxy for chlorophyll \( a \) concentration (Lorenzen, 1966). Besides dissolved oxygen concentration, fluorescence is the most measured biogeochemical property in the global ocean. The advantages of this method are as follows: (1) it can be easily measured in situ using reliable sensors, (2) the vertical resolution is high, yielding several values per meter; and (3) data are available in digital format immediately after their acquisition. The integration of fluorescence sensors on autonomous platforms (e.g. profiling floats, animals, gliders) has recently led to a burst in the acquisition of in vivo chlorophyll \( a \) fluorescence data (Claustre et al., 2010a). However, the relationship between chlorophyll \( a \) fluorescence and phytoplankton biomass is highly variable and depends on several factors,
including phytoplankton physiological state and community composition (Cunningham, 1996; Falkowski et al., 1985; Kiefer, 1973). The conversion of in situ chlorophyll a fluorescence measurements into phytoplankton biomass must therefore be done with great care.

FLAVOR (Fluorescence to Algal communities Vertical distribution in the Oceanic Realm) is a method developed to transform and combine large numbers of fluorescence profiles from various sampling sensors and platforms (Sauzède et al., 2015a). This neural network-based method generates vertical distributions of (1) chlorophyll a concentration and (2) phytoplankton community size indices (i.e. microphytoplankton, nanophytoplankton and picophytoplankton) based on the shape of in situ fluorescence profiles (i.e. normalized profiles) and the day and location of acquisition. In addition to chlorophyll a concentration, community composition is an essential variable determining the possible impact of phytoplankton on oceanic carbon fluxes and climate change scenarios (e.g., Le Quere et al., 2005). Global data compilations of phytoplankton community composition from discrete water samples have recently been published in ESSD (Peloquin et al., 2013) but data remain rather sparse. It could be an invaluable source of information to have a database of phytoplankton community size indices with the same spatio-temporal resolution as the fluorescence datasets. It now becomes possible using the FLAVOR method to transform and combine all available in situ fluorescence data into a single reference database that comprises essential information on chlorophyll a concentration and phytoplankton community size indices vertical distributions.

Presently, the widely used climatology of the global vertical distribution of chlorophyll a concentration is published in the World Ocean Atlas 2001 (Conkright et al., 2002). The latter climatology is based on estimates from analyzed water samples available in the World Ocean Database (WOD, Levitus et al., 2013) and the World Data Center (WDC, http://gcmd.gsfc.nasa.gov/). This climatology, based on seven discrete depths (0–10–20–30–50–75–100 m), is mainly limited by the lack of in situ estimations of chlorophyll a concentration, which leads to a strong spatial interpolation of data. Moreover, the discrete depths used to compute the climatology fail to finely reproduce
the vertical distribution of the phytoplankton biomass, especially in areas characterized by very deep (> 100 m) Deep Chlorophyll Maxima (DCM) such as the core of subtropical oligotrophic gyres. Using FLAVOR, the potential of the high vertical (around one data point per meter) and spatial resolution of chlorophyll fluorescence measurements would improve significantly the 3-D climatologies of chlorophyll $a$ concentration. Moreover, climatologies of phytoplankton community size indices could be created with a similar spatio-temporal resolution.

This paper presents a global compilation of chlorophyll fluorescence profiles obtained from online databases and from published and unpublished individual sources. These were converted into a global compilation of phytoplankton biomass (i.e. chlorophyll $a$ concentration) and community composition using the FLAVOR method. Prior to the application of FLAVOR, a ten-step quality control procedure was specifically developed. The remaining profiles were then analyzed. As examples of application, we present the first maps of global mean chlorophyll $a$ concentration for several oceanic layers as well as global maps phytoplankton community size indices. To further assess the quality of the resulting database, the climatological chlorophyll $a$ concentration here computed for the surface layer is compared to the climatological remotely sensed chlorophyll $a$ concentration available from Modis Aqua. Moreover, monthly 3-D climatologies of chlorophyll $a$ concentration and associated phytoplankton community size indices are analyzed for several ecological provinces defined by Longhurst (2010). Overall, the dataset presented here can be readily exploited to deepen our understanding of the spatio-temporal distribution and variability of phytoplankton biomass and associated community composition in the global ocean. It is obviously a first steps towards a database that will regularly be improved thanks to the ongoing intensification of chlorophyll $a$ fluorescence profile acquisition by Bio-Argo profiling floats, gliders and instrumented mammals.
2 Data and methods

2.1 Origins of in situ chlorophyll fluorescence measurements

The database presented in this study is available from PANGAEA, Data Publisher for Earth and Environmental Science in two formats: (1) the database containing all compiled raw fluorescence profiles (the raw database, http://doi.pangaea.de/10.1594/PANGAEA.844212, Sauzède et al., 2015b) and (2) the database containing the fluorescence profiles which are calibrated in chlorophyll a concentration and associated phytoplankton community size indices (the calibrated database, http://doi.pangaea.de/10.1594/PANGAEA.844485, Sauzède et al., 2015c). The data of in situ vertical fluorescence profiles compiled for creating the raw database were obtained from several available online databases as well as published and unpublished individual sources. The duplicates and single-surface values, which are not vertical profiles, were automatically removed (not integrated in the raw database). Finally, the raw database contains 268,127 fluorescence profiles. Following a robust quality control procedure detailed hereafter (Sect. 2.2), about 49,000 chlorophyll fluorescence profiles were converted in phytoplankton biomass (i.e. chlorophyll a concentration) and size-based community composition (i.e. microphytoplankton, nanophytoplankton and picophytoplankton). The origin of this calibrated database is summarized in Table 1. The majority of the data comes from the National Oceanographic Data Center (NODC) and the fluorescence profiles acquired by Bio-Argo floats available on the Oceanographic Autonomous Observations (OAO) web platform (63.7 and 12.5 % respectively, see percentages of data in the database depending on their origin in Table 1).

Different modes of acquisition were used to collect the data presented in this study: (1) the CTD profiles are acquired using a fluorometer mounted on a CTD-rosette, (2) the OSD (Ocean Station Data) profiles are derived from water samples analyzed by fluorometry and are defined as “low” resolution profiles (Boyer et al., 2009), (3) the UOR (Undulating Oceanographic Recorder) profiles are acquired by a “fish” equipped with fluorometer and towed by a research vessel, (4) AP profiles are acquired by au-
tonomous platforms (here Bio-Argo profiling floats or elephant seals equipped with a fluorometer, Claustre et al., 2010b; Guinet et al., 2012). Table 2 lists the number of raw fluorescence profiles in the raw database according to these four modes of acquisition.

It is worth to note that the data acquired from gliders were not included in the database. Although glider data are extremely numerous, they are restricted to a very small spatio-temporal window. As a consequence, a database including gliders data would likely be spatially and temporally biased, in contradiction with our first aim of building a global climatological database.

2.2 Quality control

In order to use the FLAVOR method (see details in Sect. 2.3), a specific and adapted data quality control procedure was developed and applied to each in situ chlorophyll fluorescence profile. This procedure was schematically implemented according to four main steps of data control (Fig. 1), each step being developed for discarding most, if not all, spurious fluorescence profiles that would deteriorate the quality of the database. Firstly, several basic tests were applied: (1) duplicates and single-surface values, which are not vertical profiles, were removed (these profiles were removed from the beginning of the process so they are not included in the so-called raw database), (2) coastal profiles are removed using a bathymetric mask of 500 m depth, (3) the uppermost measurement has to be located within the 0–10 m layer while the deepest measurement has to be at or below 100 m. Secondly, tests on the profile vertical resolution are applied: (4) a minimum of ten values per profile is required (i.e. condition on the vertical resolution acquisition), (5) a minimum of five different values per profile is required (i.e. condition on the sensor resolution). Then, several tests on the fluorescence profile shape are applied. These conditions are based on the parameter used for the development of the FLAVOR method, $Z_0$, which is the depth at which the fluorescence profile returns to a constant background value (see details in Sect. 2.3 and examples in Fig. 1b and c): (6) the median of the fluorescence values from the surface down to
20 m has to be greater than the median of the values of the last 10 % of the deepest samples of the profile (see Fig. 1a), (7) the depth $Z_0$ has to be within the last 10 % of the deepest samples of the profile (see Fig. 1b). Finally, a test on the noise of the profiles was developed and applied: (8) profiles with aberrant data caused by electronic noise are removed (i.e. variability greater than 20 % of the total profile range, see Fig. 1c). To finish, a visual check allowed to verify all the remaining fluorescence profiles. The number of fluorescence profiles rejected at each step of the quality control procedure is presented in Table 3. Around 80 % of the raw fluorescence profiles were thus removed by this procedure. This step is an essential prerequisite for the development of a “clean” database of vertical distributions of phytoplankton biomass and community composition in the global ocean. The quality control procedure removed 77, 71, 28 and 25 % of the OSD, UOR, AP and CTD profiles, respectively, with profiles removed by the test on the bathymetry not taken into account.

2.3 Conversion of chlorophyll fluorescence into chlorophyll a concentration and phytoplankton community composition

In order to assess the vertical distribution of the total chlorophyll a concentration (hereafter, [TChl]) and the chlorophyll a concentration associated to each phytoplankton size index (hereafter, [microChl], [nanoChl] and [picoChl] for microphytoplankton, nanophytoplankton and picophytoplankton respectively), the FLAVOR method (Sauzède et al., 2015a) is applied to each chlorophyll fluorescence satisfying the quality control procedure (see Sect. 2.2). In summary, FLAVOR is a neural network-based method which uses as inputs (1) the shape of the chlorophyll fluorescence profile (ten values from the normalized profile with values range between 0 and 1), (2) the depth $Z_0$ which is the depth at which the fluorescence profile returns to a constant background value (see examples of $Z_0$ depths represented by the horizontal red line for two profiles on Fig. 1b and c); and (3) the location (latitude and longitude) and the day of acquisition of the fluorescence profile. The outputs of FLAVOR are the vertical distributions of (1) [TChl] and (2) [microChl], [nanoChl] and [picoChl]. Those vertical distributions have the
same vertical resolution as the input raw fluorescence profile. The global absolute errors of FLAVOR retrievals are 40, 46, 35 and 40 % for the [TChl], [microChl], [nanoChl] and [picoChl] respectively (Sauzède et al., 2015a). It is important to note that one of the main failures of FLAVOR is that the impact of the daytime Non-Photochemical Quenching (NPQ; see, e.g., Cullen and Lewis, 1995), which is responsible for a decrease of chlorophyll fluorescence values at high irradiance, is not accounted for by the method. If density profiles are available with fluorescence profiles, the NPQ could be corrected using the method of Xing et al. (2012) which involves substituting the fluorescence values acquired within the mixed layer by the maximum value within this layer.

An additional step of the quality control is further applied once the FLAVOR method has been operated. It is based on the Chauvenet’s criterion which is used to identify statistical outliers in the retrieved biomass data (Buitenhuis et al., 2013; Glover et al., 2011; O’Brien et al., 2013). The criterion was applied to the surface data of each profile (median of values from the surface down to 20 m). As the Chauvenet’s criterion is based on the assumption that the data follows a normal distribution, the analysis was performed on the log-normalized [TChl] surface values. Such a criterion removes aberrant data partially caused by the failure of the FLAVOR method (see number of profiles removed by the Chauvenet’s criterion in Table 3).

The present database corresponds to model outputs, once the FLAVOR method has been applied. It has been mentioned previously that this method is not adapted for the retrieval of chlorophyll \( a \) concentration on a fluorescence profile-by-profile basis (Sauzède et al., 2015a). Rather, FLAVOR and, hence, the resulting database are relevant for large scale investigations, e.g. development of climatologies of the vertical distribution of chlorophyll \( a \) from which regional anomalies or temporal trends might be evidenced.
3 Results and discussion

3.1 Spatial and temporal coverage of the database

A total of 268,127 raw chlorophyll fluorescence profiles were collected (duplicates and single-surface samples automatically removed) and controlled with a specifically developed quality control procedure as explained in Sect. 2.2. The remaining 48,600 chlorophyll fluorescence profiles which successfully pass all the step of the quality control were calibrated in total chlorophyll $a$ concentration and associated phytoplankton community size indices (i.e. microphytoplankton, nanophytoplankton and picophytoplankton) using FLAVOR (see details in Sect. 2.3). The resulting database covers all ocean basins with more profiles in the Northern Hemisphere (75%) than in the Southern Hemisphere (25%, see Figs. 2 and 3). However, the Southern Hemisphere remains relatively well represented with the profiles acquired by autonomous platforms and especially by elephant seals equipped with a fluorometer. Few data were acquired in the Indian Ocean and in the tropical South Atlantic and South Pacific (see Fig. 2).

The highest numbers of fluorescence profiles are found at the BATS (the Bermuda Atlantic Time-series Study) and HOT (the Hawaii Ocean Time-series) time-series stations which are located at $31.67^\circ$N–$64.17^\circ$W and $22.75^\circ$N–$158.00^\circ$W respectively, and where data acquisition started in 1988. On the annual scale, the data acquisition appears evenly distributed, with a slight under representation of autumn months (April to June) in the Southern Hemisphere (Fig. 4). The temporal distribution of fluorescence profiles in the database covers 56 years from 1958 to nowadays (Fig. 5a) and most of the observations were collected after the late 1980s. There are fewer observations from 2010 to 2012 because all data generally acquired by ship-based platforms have not been archived yet in the online databases. A significant increase in data density observed between 2013 and 2015 (in 2015, 124 profiles were acquired in half a month) mainly results from data acquired by Bio-Argo profiling floats. Around a sixth of this global database has been sampled in only two years by the Bio-Argo platforms. This
illustrates the potential of this new type of acquisition which is expected to dramatically increase the number of collected fluorescence profiles in the future.

Vertically, the database includes values of total chlorophyll \(a\) concentration and associated phytoplankton community composition from the surface down to a mean sampling depth of 743 m (with a maximum sampling depth ranging from 100 to 6000 m; Fig. 5b).

### 3.2 Vertical distribution of the chlorophyll biomass

We present the database with respect to the vertical distribution of the total chlorophyll \(a\) concentration ([TChl]). Figure 6 displays the median [TChl] gridded within squares of 3° latitude by 3° longitude and over six vertical layers (0–25, 25–50, 50–75, 75–100, 100–150 and 150–200 m). In the surface layer (0–25 m, see Fig. 6a), the [TChl] median is the highest in the North Atlantic and the lowest in the South Pacific subtropical gyre. The median [TChl] decreases with depth for all the data, except for data acquired in South Pacific and Atlantic subtropical gyres where the median [TChl] increases with depth. This increase is associated with the so-called Deep Chlorophyll Maximum (DCM) that is typical of these oligotrophic regions (e.g., Cullen, 1982; Mignot et al., 2011, 2014).

The global distribution of the phytoplankton community composition, given in terms of fraction of chlorophyll \(a\) concentration associated to micro-, nano- and pico-phytoplankton, is presented for the 0–1.5 \(Z_e\) layer (Fig. 7a–c respectively). Here \(Z_e\), the euphotic depth is defined as the depth at which the irradiance is reduced to 1 % of its surface value. It was estimated according to the method of Morel and Berthon (1989), using the [TChl] profiles derived from FLAVOR. Figure 7 reveals general geographic patterns which are consistent with the knowledge about the ecological domains and biogeochemical provinces (e.g., Longhurst, 2010). On average microphytoplankton are dominant in the subarctic zone with a relative contribution to the chlorophyll biomass reaching more than 70 % in these areas (Fig. 7a). Picophytoplankton are dominant in the subtropical gyres (South and North Pacific as well as South and North Atlantic) with
a contribution reaching 45–55 % (Fig. 7c). Nanophytoplankton appear to be ubiquitous with a relatively stable contribution to biomass of 40–50 % (Fig. 7b).

To further assess the quality, range and representation of the FLAVOR-retrieved [TChl] database presented in this study, the retrieved surface [TChl] is compared to the remotely sensed [TChl]. In this context, the climatological [TChl] mean was extracted at a 9 km spatial resolution from NASA Modis Aqua archive for the time period covering 2002 to 2014. The extracted satellite [TChl] data were re-gridded to a 3° × 3° spatial resolution. Similarly the FLAVOR-retrieved [TChl] values for the upper layer of the database (i.e., mean value calculated between the surface and 20 m) for the same period were re-gridded to 3° × 3° squares. Figures 8 and 9 show that climatological averaged [TChl] from Modis Aqua and from the present database are generally consistent (Fig. 8a and b). The log-transformed ratio of the Modis Aqua to the database [TChl] estimates reveals a rather good agreement with a median value of −0.16 and a standard deviation of 0.58 (see histogram in Fig. 8c). Figure 9 displays the geographic distribution of the log-transformed ratio between the Modis Aqua and the database estimates of climatological surface [TChl]. The ratio shows no specific spatial bias. The FLAVOR method was validated in all ocean basins individually. The retrievals obtained for the Southern and the Arctic Ocean were slightly less accurate than for the other basins. It is therefore possible that the estimation errors are greater in these areas and particularly due to an insufficient data density. This observation has also to be nuanced considering the difficulties in retrieving accurate ocean color satellite [TChl] in these high latitude environments (Gregg and Casey, 2004; Guinet et al., 2012; Johnson et al., 2013; Peloquin et al., 2013; Siegel et al., 2005; Szeto et al., 2011).

3.3 Example of application: climatological time series of the vertical distribution of chlorophyll a concentration and phytoplankton community composition

As an example of application, monthly climatologies were computed for three ecological provinces defined by Longhurst (2010) and well represented in the current dataset
(Fig. 10a): (1) the North Atlantic Subtropical Gyral Province West (NASW, Fig. 10b), (2) the Atlantic Subarctic Province (SARC, Fig. 10c) and (3) the North Pacific Subtropical Gyre Province (NPTG, Fig. 10d). Overall the time series of the vertical distribution in [TChl] are consistent with expectations as detailed by Longhurst (2010). For the NASW province (Fig. 10b), the [TChl] is relatively homogeneous from the surface to around 140 m from January to March; then the stratification of the water column leads to the establishment of a Deep Chlorophyll Maximum (DCM) from April to November. Over the year, [TChl] varies in a restricted range of values (0.35–0.55 mgm$^{-3}$). The dominant phytoplankton groups are the nano- and the pico-phytoplankton with relative chlorophyll contribution reaching 40–45 % for both size-based groups. The contribution of microphytoplankton remains low throughout the year (10 %). For the SARC province, the phytoplankton bloom starts in May (as indicated by a significant increase in [TChl], Fig. 10c). The bloom continues for 4 to 5 months with [TChl] within the 1.5–2 mgm$^{-3}$ range (with maximum values in July). The microphytoplankton contribution increases during the bloom and reaches a maximum (60 %) in August, whereas the nanophytoplankton relative contribution decreases from April to August. The contribution of picophytoplankton increases slightly all along the year to reach a maximum of about 40 % in December. For the NPTG province (Fig. 10d), a DCM (0.15–0.25 mgm$^{-3}$) is established at a depth of 100–125 m and persists all year long. This DCM deepens in summer, consistently with a deeper light penetration in the water column at this period. The [TChl] at DCM reaches a maximum value in June and July. The dominant phytoplankton groups are the nano- and the pico-phytoplankton with relative contribution reaching 45–50 % for both size-based groups and slight opposite temporal evolutions. The contribution of microphytoplankton remains low all over the year (< 10 %).

4 Conclusions and recommendations for use

The phytoplankton biomass (i.e. chlorophyll $a$ concentration) and phytoplankton community size indices were derived from chlorophyll fluorescence profiles using a dedi-
cated calibration method (FLAVOR, Sauzède et al., 2015a). For the first time, in situ chlorophyll fluorescence profiles from various data centers have been collected and synthesized in a global dataset to create unified and interoperable products related to chlorophyll a concentration and phytoplankton communities. This work can thus be considered as a first step towards the development of a 3-D climatological representation of chlorophyll a concentration and phytoplankton community composition. As mentioned before, we recall here that this database should not be used on a profile-by-profile basis. Instead, this database has rather to be used to derive climatologies from which regional or temporal trends might possibly be extracted. To date, and because of the lack of in situ vertical data, the identification of such trends has been based exclusively on surface remotely-sensed data (Beaulieu et al., 2013; Boyce et al., 2010; Gregg, 2005; Gregg and Conkright, 2002). Obviously, the present dataset offers a potential refinement to improve open-ocean climatologies of chlorophyll a with respect to the vertical dimension.

Finally, this database has to be considered as a reference that has the potential to evolve. It is now clear that numerous fluorescence profiles will be acquired through robotic observations (e.g. Claustre et al., 2010b; Johnson et al., 2009). In fact, about a sixth of the profiles of the present database has been sampled by Bio-Argo profiling floats in only two years. Therefore the database proposed here represents a first step towards a global single reference database reconciling the oldest datasets of chlorophyll fluorescence with the future ones mostly acquired remotely by autonomous platforms.

Acknowledgements. This paper is a contribution to the Remotely Sensed Biogeochemical Cycles in the Ocean (remOcean) project, funded by the European Research Council (grant agreement 246777), to the French Bio-Argo project funded by CNES-TOSCA. The French PROOF and CYBER programs are acknowledged for their support of cruises where in situ chlorophyll fluorescence profiles were acquired. We are grateful to all the project PIs who contributed data as well as to the anonymous staff who took part in the data acquisition during the cruises.
References


Table 1. Summary of the contributions of the chlorophyll fluorescence profiles in the database presented in this study.

<table>
<thead>
<tr>
<th>Data source/Institute/Investigator</th>
<th>Period</th>
<th>Number of fluorescence profiles</th>
<th>Percentage of data in the database</th>
<th>Website if available or contact for requests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laboratoire d’Océanographie de Villefranche (LOV) cruises</td>
<td>May 1991–Jan 2012</td>
<td>3,320</td>
<td>6.8%</td>
<td><a href="mailto:clautre@obs-vlfr.fr">clautre@obs-vlfr.fr</a>, <a href="mailto:sauzede@obs-vlfr.fr">sauzede@obs-vlfr.fr</a></td>
</tr>
<tr>
<td>C. Guinet (data acquired by elephants seals, Guinet et al., 2012)</td>
<td>Dec 2007–Jan 2011</td>
<td>1,908</td>
<td>3.9%</td>
<td><a href="mailto:christophe.guinet@cebc.cnrs.fr">christophe.guinet@cebc.cnrs.fr</a></td>
</tr>
<tr>
<td>British Oceanographic Data Center (BODC)</td>
<td>Sep 1996–Nov 2008</td>
<td>1,219</td>
<td>2.5%</td>
<td><a href="http://www.bodc.ac.uk/">http://www.bodc.ac.uk/</a></td>
</tr>
<tr>
<td>Systèmes d’Informations Scientifiques pour la MER (SISMER)</td>
<td>Sep 1999–May 2008</td>
<td>237</td>
<td>0.5%</td>
<td><a href="http://www.ifremer.fr/sismer/">http://www.ifremer.fr/sismer/</a></td>
</tr>
<tr>
<td>Australian Antarctic Data Center (AADC)</td>
<td>Jan 2001–Feb 2006</td>
<td>234</td>
<td>0.5%</td>
<td><a href="http://data.aad.gov.au/">http://data.aad.gov.au/</a></td>
</tr>
<tr>
<td>Southern Ocean Iron RElease Experiment (SOIREE)</td>
<td>Feb 1999</td>
<td>57</td>
<td>0.1%</td>
<td><a href="http://www.uea.ac.uk/~e610/soiree/index.html">http://www.uea.ac.uk/~e610/soiree/index.html</a></td>
</tr>
</tbody>
</table>
Table 2. Summary for the chlorophyll fluorescence profiles in the database presented in the study depending on the different modes of data acquisition.

<table>
<thead>
<tr>
<th>Acquisition</th>
<th>Number of fluorescence profiles</th>
<th>Percentage of data in the database</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTD</td>
<td>27 433</td>
<td>56.4 %</td>
</tr>
<tr>
<td>OSD</td>
<td>10 831</td>
<td>22.3 %</td>
</tr>
<tr>
<td>UOR</td>
<td>2952</td>
<td>6 %</td>
</tr>
<tr>
<td>AP</td>
<td>7384</td>
<td>15.2 %</td>
</tr>
<tr>
<td>Total</td>
<td>48 600</td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Summary of the number of fluorescence profiles rejected at each step of the quality control.

<table>
<thead>
<tr>
<th>QC step number (see Fig. 1)</th>
<th>Number of fluorescence profiles deleted</th>
<th>% of data deleted</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>162 609</td>
<td>74 %</td>
</tr>
<tr>
<td>3</td>
<td>31 904</td>
<td>14.5 %</td>
</tr>
<tr>
<td>4</td>
<td>15 396</td>
<td>7 %</td>
</tr>
<tr>
<td>5</td>
<td>286</td>
<td>0.1 %</td>
</tr>
<tr>
<td>6</td>
<td>3569</td>
<td>1.6 %</td>
</tr>
<tr>
<td>7</td>
<td>2891</td>
<td>1.3 %</td>
</tr>
<tr>
<td>8</td>
<td>1597</td>
<td>0.7 %</td>
</tr>
<tr>
<td>9 – Visual check</td>
<td>244</td>
<td>0.1 %</td>
</tr>
<tr>
<td>Chauvenet’s criterion and</td>
<td>1031</td>
<td>0.5 %</td>
</tr>
<tr>
<td>range criterion after</td>
<td></td>
<td></td>
</tr>
<tr>
<td>calibration (see Sect. 2.3)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1. Schematic overview of the quality control procedure specifically developed for the database presented in this study. The fluorescence profiles represented in the (a), (b) and (c) panels are examples of profiles which are rejected by the quality control steps (6), (7) and (8) respectively.
Figure 2. Geographic distribution of the 48 600 chlorophyll fluorescence profiles in the database that passed through all the steps of the quality control procedure. The colorscale indicates the number of fluorescence profiles in boxes of 3° per 3°.
Figure 3. Frequency distribution of the 48,600 profiles of chlorophyll a concentration and associated phytoplankton community composition in the database as a function of latitude.
Figure 4. Temporal distribution of the 48 600 profiles of chlorophyll a concentration and associated phytoplankton community composition in the database as a function of months with black and grey colors indicating the hemispheres of data acquisition.
Figure 5. Frequency distribution of the 48,600 profiles of chlorophyll a concentration and associated phytoplankton community composition in the database as a function of: (a) years of acquisition and (b) the maximum depth of acquisition. Colors refer to the different modes of data acquisition.
Figure 6. Median total chlorophyll $a$ concentration (mg m$^{-3}$) scaled to a 3° spatial resolution for six vertical layers: (a) 0–25 m, (b) 25–50 m (c) 50–75 m (d) 75–100 m (e) 100–150 m and (f) 150–200 m.
Figure 7. Mean relative contribution to the total chlorophyll a biomass (%) for the three phytoplankton size-based groups gridded and scaled to a 3° resolution within the 0–1.5 $Z_e$ layer: (a) microphytoplankton, (b) nanophytoplankton and (c) picophytoplankton.
Database of global ocean chlorophyll a profiles

R. Sauzède et al.

a) Modis Aqua mean

b) Database mean

c) Histogram

mean=0.14
median=0.12
sd=0.57
**Figure 8.** (a) Climatological mean (2002–2014) chlorophyll a concentration (mg m⁻³) from Modis Aqua (scaled to a 3° resolution); (b) climatological mean (1958–2015) surface chlorophyll a concentration (mg m⁻³) from the present database (averaged over the upper 20 m and scaled to a 3° resolution); (c) histogram of the log10 ratio of the chlorophyll a concentration from the database to the chlorophyll a concentration from Modis Aqua. The mean, median and standard deviation of the ratio are indicated in the figure. The color scale applies to panels (a) and (b).
Figure 9. Geographic distribution of the log10-transformed ratio of the climatological mean surface [TChl] of the database over the upper 20 m of the water column ([TChl]DB) and the climatological mean satellite [TChl] from Modis aqua ([TChl]sat). Both [TChl] were scaled to a 3° resolution.
a) Ecological provinces

b) NASW

- Number of profiles
- Depth (m)
- Phytoplankton community composition (%)
- [Chl] (mg m^{-3})

<table>
<thead>
<tr>
<th>Depth (m)</th>
<th>Phytoplankton community composition (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>50</td>
<td>25</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
</tr>
<tr>
<td>150</td>
<td>75</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
</tr>
<tr>
<td>250</td>
<td>100</td>
</tr>
<tr>
<td>300</td>
<td>100</td>
</tr>
<tr>
<td>[Chl] (mg m^{-3})</td>
<td></td>
</tr>
<tr>
<td>0.00</td>
<td>0.0</td>
</tr>
<tr>
<td>0.50</td>
<td>1.0</td>
</tr>
<tr>
<td>1.00</td>
<td>1.5</td>
</tr>
<tr>
<td>1.50</td>
<td>2.0</td>
</tr>
</tbody>
</table>

- [microChl], [nanoChl], [picoChl]

398
**Figure 10.** Monthly climatologies of the vertical distribution of the total chlorophyll $a$ concentration and associated phytoplankton size-based groups for three ecological provinces defined by Longhurst (2010). *(a)* Geographic distribution of the considered provinces: North Atlantic Subtropical Gyral Province West (NASW), Atlantic Subarctic Province (SARC) and North Pacific Subtropical Gyre Province (NPTG). Climatologies obtained for the *(b)* NASW, *(c)* SARC and *(d)* NPTG. The color scale indicates the total chlorophyll $a$ concentration ($\text{mg m}^{-3}$); the data points superimposed onto the colored monthly vertical profiles show the percentages of integrated chlorophyll $a$ concentration associated with the micro-, nano- and pico-phytoplankton within the water column.