Open Access

Earth System
Science
Data

A synthetic satellite dataset of the spatio-temporal distributions of *Emiliania huxleyi* blooms and their impacts on Arctic and sub-Arctic marine environments (1998–2016)

Dmitry Kondrik, Eduard Kazakov, and Dmitry Pozdnyakov

Nansen International Environmental and Remote Sensing Centre, Saint Petersburg, 199034, Russian Federation

Correspondence: Eduard Kazakov (ekazakov@niersc.spb.ru)

Abstract. A 19-year (1998–2016) continuous dataset is presented of coccolithophore *Emiliania huxleyi* distributions and activity, i.e. the release of $CaCO_3$ in water and the decrease of uptake of dissolved CO_2 by *Emiliania huxleyi* cells (e.g. Kondrik et al., 2018a), in Arctic and sub-Arctic seas. The dataset is based on optical remote-sensing data (mostly OC CCI data) with assimilation of different relevant in situ observations, preprocessed with authorial algorithms. Alongside bloom locations, we provide both detailed information on *E. huxleyi* impacts on carbon balance and the sub-datasets of quantified coccolith concentrations, particulate inorganic carbon content and CO_2 partial pressure in water driven by coccolithophores. All data are presented on a regular 4×4 km grid at a temporal resolution of 8 days. The paper describes the theoretical and methodological basis for all processing and modelling steps. The data are available on Zenodo: https://doi.org/10.5281/zenodo.1402033.

1 Introduction

Among the topics related to ongoing climate change, there are alterations in both biodiversity in marine environments and the carbon balance in the atmosphere–ocean system (Rost et al., 2008). In some specific cases both processes are interrelated, being spurred up by one and the same agent(s). Along with other marine inhabitants, coccolithophores are such entities. Specifically, we look at the algal species named *Emiliania huxleyi* – a unicellular planktonic organism that is the most widespread coccolithophore in the world's oceans. Being simultaneously a calcifying and photosynthetic primary producer of, respectively, inorganic and organic carbon, *Emiliania huxleyi*, in the course of its life cycle, enhances both the concentration of calcite and carbon dioxide partial pressure in ocean surface water. At least within *Emiliania huxleyi* bloom areas, both processes are capable of changing the carbon balance and hence affect both CO_2 fluxes between the atmosphere and surface ocean and the aquatic biogeochemistry. Being a spatially huge phenomenon invariably

occurring in both hemispheres, and steadily propagating in the poleward direction (Winter et al., 2014) due to CO_2 accumulation in the atmosphere and ensuing climate warming (Johannessen, 2008), *Emiliania huxleyi* blooms are believed to be highly relevant to understanding the comprehensive nature of the changes unfolding on our planet.

Historically, the initial build-up of knowledge on coccolithophores in general, and *Emiliania huxleyi* specifically, was broadly based on in situ approaches effected in the course of both shipborne and laboratory activities. Extensive data were obtained on *Emiliania huxleyi* cell morphometry, internal structure, intracellular dark reactions and photoreactions, factors controlling and affecting the cell growth, as well as intrinsic optical properties, such as total sunlight and spectral absorption, scattering and backscattering (Balch et al., 1996a). In addition, regression relationships were established between *Emiliania huxleyi*-driven changes in both inherent hydro-optical parameters and CO_2 partial pressure in surface water within the bloom area (Holligan et al., 1993).

However, as this phenomenon extends over marine areas in excess of hundreds of thousands of square kilometres (Balch et al., 2016; Kondrik et al., 2018a), and is spatially and temporally highly dynamic, a solely satellite remote-sensing approach is able to comply with the challenge of studying it.

Until recently, only a few satellite studies were performed and published on the typical locations of *Emiliania huxleyi* blooms and associated concentrations of particulate inorganic carbon in surface ocean within the bloom area (e.g. Gordon et al., 2001; Balch et al., 2016).

Prior to the publication by Kondrik et al. (2018a), to the best of our knowledge, only a couple of studies (Shutler et al., 2010, 2013) were undertaken to retrieve from space-borne data either both the total content of inorganic carbon produced by a *Emiliania huxleyi* bloom (PIC) and an increase in CO_2 partial pressure (ΔpCO_2) in surface water within the bloom area or else reveal intraannual and interannual variations over long time periods in the location and intensity of *Emiliania huxleyi* blooms. No concatenated time series data of a nearly 20-year duration are available to date on the associated quantifications of bloom surface, bloom intensity or ΔpCO_2 for all *Emiliania huxleyi* blooms occurring within extensive latitudinal belts and encompassing waters of different oceans, i.e. marine tracts significantly distanced longitudinally.

Meanwhile, the above-specified information is an indispensable step towards a further pan-global inventory of the effects produced by *E. huxleyi* blooms on both marine chemistry and ecology, and CO_2 exchange fluxes between the atmosphere and ocean, as such fluxes condition the status of the world's oceans as a sink of CO_2.

In addition to the studies cited above, it is also worth mentioning a few sources of multi-year satellite data on coccolithophore blooms that may be useful for potential users who wish to broaden their multifaceted databases in their studies.

The NASA OCEANCOLOR portal https://oceancolor.gsfc.nasa.gov/atbd/pic/ (last access: 10 January 2019) offers extensive data on particulate inorganic carbon retrieved from MODIS with the Balch et al. (2005) methodology. Downloadable from https://oceandata.sci.gsfc.nasa.gov/ (last access: 10 January 2019), these data have a 4 km and 1 day spatial and temporal resolution and cover the time period starting from 2000.

Also, reported by Loveday and Smyth (2018) a 40-year time series (1978–2018) from AVHRR observation data on coccolithophore blooms is available at https://doi.org/10.1594/PANGAEA.892175. By employing specially developed coccolithophore bloom area masks that were developed from remote-sensing reflectance spectra, these data are monthly worldwide and available at a spatial resolution of 0.1° ($\sim$ 10 km). Although these data do not encompass any additional parameters such as particulate inorganic carbon or CO_2 partial pressure in surface water within the bloom, they can be valuable due to an exceptionally long observation period.

Unlike the publications mentioned above, the present paper reports on extensive concatenated original datasets generated for subpolar and polar seas of the Northern Hemisphere, viz. the North, Labrador (with adjacent North Atlantic open waters), Norwegian, Barents, Greenland and Bering seas. Based on the employed space-borne ocean colour information, the obtained datasets are processed into a nearly two decadal (1998–2016) time series for each of the target seas and marine areas. They encapsulate information about PIC and ΔpCO_2 values in surface water within the bloom area together with intraannual and interannual variations in the location and intensity of *Emiliania huxleyi* blooms over a variety of seas and across a nearly 20-year time period.

Conjoined with a wealth of presently available supplementary data from satellite and shipborne missions on the environmental conditions under which target *Emiliania huxleyi* blooms emerged and developed, the synthetic dataset we are reporting on opens the way to detailed analyses of forward and feedback mechanisms governing the temporal and spatial dynamics of this phenomenon. Further utilisation of the results of such analyses in regional and global climatic models promises to predict future directions of development of the phenomenon in question (Rost et al., 2008).

2   Methodology and dataset content

Based on the facility of available satellite OC CCI (Ocean Colour Climate Change Initiative) and SeaWiFS data in the visible part of the spectrum, the following products have been generated to achieve the goals specified in the previous section, viz. (1) *Emiliania huxleyi* bloom extent, (2) concentration of coccoliths within the bloom, (3) total content of particulate inorganic carbon (PIC) produced by the bloom, and (4) an increase in CO_2 partial pressure in marine surface waters due to the blooming phenomenon.

2.1   Bloom area quantification

Quantification of *Emiliania huxleyi* bloom areas was performed in two stages. Firstly, RGB (red, green, blue) images were generated based on the weighted remote-sensing reflectance, R_{rs}, which is the upwelling spectral radiance just above the water–air interface normalised to the downwelling spectral irradiance at the same level (Bukata et al., 1995). R_{rs} values in the channels centred at 670, 555 and 443 nm were employed. Analysis of the space-borne radiometric data collected by Kondrik et al. (2017a, b) from the five target seas, yielded statistically robust specific ranges of $R_{rs}(\lambda)$ highlighting *Emiliania huxleyi* blooms as turquoise areas; the areas of blooms of other (non-calcifying) algae were reflected in the images as green. Areas with scarce non-calcifying algae abundance showed up as blue or dark blue. The land mask was overlaid so that land areas were coloured light yellow.

In the second stage of quantification of the *Emiliania huxleyi* bloom extent, an additional criterion was imposed on the revealed turquoise areas: R_{rs} values should be maximal at 490 and/or 510 nm, while at other wavelengths they need to be in excess of 0.001 (412 nm), 0.008 (443 nm), 0.01 (490 nm), 0.008 (510 nm), 0.008 (555 nm) and ~ 0 (670 nm). Such a selection provided the highest accuracy of bloom delineation. With the known pixel size, the bloom area can be confidently quantified. An example of *Emiliania huxleyi* bloom extent masking is shown in Fig. 1.

2.2 Determination of the coccolith concentration

Determination of the coccolith concentration within the bloom was performed with the BOREALI algorithm (Bio-Optical REtrieval ALgorIthm, Korosov et al., 2009), based on the Levenberg–Marquardt (L-M) finite difference technique (Press et al., 1992). The L-M technique solves the inverse problem, i.e. in our case it allows us to retrieve the concentrations of water constituents from spectral subsurface remote-sensing reflectance, $R_{rsw}(\lambda)$, which is the upwelling spectral radiance just beneath the water–air interface normalised to the downwelling spectral irradiance at the same level (Jerome et al., 1996). A hydro-optical model accommodating spectral specific absorption and backscattering coefficients of *Emiliania huxleyi* cells and coccoliths. As well as pure water, non-calcifying alga and dissolved organic matter was developed and employed to run the BOREALI (Kondrik et al., 2017a).

The results of validation of coccolith concentration retrievals with BOREALI were assessed through the following statistical measures: coefficient of correlation, r; linear regression equation, $f(x)$; coefficient of determination, R^2; root mean square error, RMSE; and systematic error, BIAS and MAE. BIAS and MAE were then also normalised to the absolute values of coccoliths concentrations determined by using each model: $r = 0.88$, $f(x) = 0.6159x + 6.9197$, $R^2 = 0.77$; RMSE $= 3.55 \times 10^9$ coccoliths m^{-3}, BIAS $= 25.30\%$ and MAE $= 32.30\%$.

In addition, ascertained by both RGB and R_{rs} approaches, *Emiliania huxleyi* bloom areas were further checked using the results of coccolith concentration retrievals. This was done through the application of a threshold. A threshold of 90×10^9 coccoliths m^{-3} was chosen because, firstly, it ensures the best correspondence between the bloom surfaces, determined by our radiometric and BOREALI algorithms. Secondly, this threshold is very close to the average value of coccolith concentrations in developed *Emiliania huxleyi* blooms reported from the world's oceans (for references, see Balch et al., 1996b, 2005). The numerical assessments of bloom surfaces delineated and quantified by the above independent methods converged precisely.
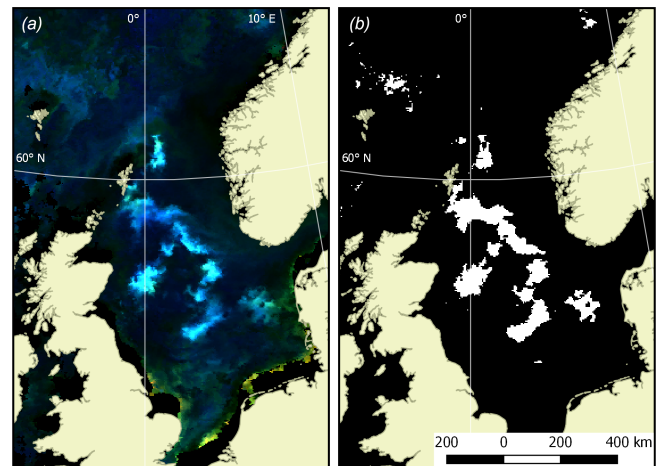

Figure 1. Example of the bloom masking algorithm performance. **(a)** Source of the OC CCI RGB imagery for the North Sea (9 June 2016, with land mask); **(b)** calculated bloom mask (white pixels stand for bloom detected, black pixels are areas void of bloom).

2.3 Coccolith content, particulate inorganic carbon and CO$_2$ partial pressure increment determination

Determination of the coccolith content (CC) was performed through establishing mixed-layer depth (MLD) within the bloom area. The climatology of Montegut et al. (2004) was applied. The identified areas of *Emiliania huxleyi* blooms with retrieved concentrations of coccoliths were overlapped by the respective climatological MLD fields, and for each pixel, the value of MLD was further used to calculate CC. Further, CC values were used to quantify the total content of particulate inorganic carbon (PIC). It was done for each 8-day time period (corresponding to the temporal resolution of the space-borne radiometric data employed) through multiplying the carbon mass per coccolith, m and CC followed by summarising the results of multiplication within all pixels of respective bloom extent. The value of m was equal to 0.2 pg (Balch et al., 2005). The moment at which the PIC assessment could be ideally performed in each bloom corresponded to the situation when two conditions were fulfilled: (a) the bloom attained its largest surface and (b) the spectral curvature of remote-sensing reflectance, $R_{rs}(\lambda)$, exhibited a maximum at about 490 nm as the location of the R_{rs} maximum at about 490 nm is an indication that the bloom is prevalently composed of coccoliths (Kondrik et al., 2017a).

Remote determinations of *Emiliania huxleyi*-driven pCO_2 increment (ΔpCO_2) consisted of establishing a relationship between *Emiliania huxleyi*-driven changes in pCO_2, that is, ΔpCO_2, in bloom pixels, and the respective values of R_{rs} (490). Such a relationship (Kondrik et al., 2018a) with the following statistical characteristics, coefficient of determination, $r^2 = 0.54$, $p \ll 0.001$, and RMSE $= 23.4$ µatm, was used to quantify the spatial variations of ΔpCO_2 in the target seas followed by recalculating ΔpCO_2 for the water tem-

peratures (retrieved from space-borne data) that actually occurred during the respective *Emiliania huxleyi* bloom events (Copin-Montegut, 1988).

2.4   Additional technical workflow

During satellite data processing, several procedures were performed.

1. Satellite images were reprojected. Given the high latitudinal location of the target seas, it was relevant to use an equal-area polar projection. Therefore, the NASA Ease-Grid was employed. The system of coordinates of the WGS-84 (World Geodetic System 1984) is based on Ease-Grid.

2. Automatic cloud masking was corrected in the images from SeaWiFS in 1998–2001. In all images of the OC CCI product obtained in 1998–2001 (when only the SeaWiFS sensor was operational), putative bloom areas proved to be masked. The errors of automatic cloud masking most probably resulted from very high values of brightness stemming from bloom areas (comparable with cloud-produced signals), which may have led to possible mistakes in the masking algorithm. The problem was overcome via manual processing of the data of a lower level, i.e. directly from the SeaWiFS level 2 product (http://oceancolor.gsfc.nasa.gov/cgi/browse.pl? sen=am, last access: 10 January 2019) for the period of 1998–2001 in all studied areas. As a result, in the RGB images the areas masked as clouds in OC CCI images proved to exhibit large bloom areas with the brightness of signals typical of *Emiliania huxleyi*. This approach was legitimate as OC CCI data obtained by different sensors have been brought to the SeaWiFS standard channels, and the entire data time series (1998–2016) was radiometrically uniform.

3. Missing pixels masked as ragged clouds were filled. In the case of ragged clouds, some pixels of RGB images are not informative. A special algorithm for filling such gaps included averaging of $R_{rs}(\lambda)$ values from neighbouring pixels and from temporarily previous and following images of the same pixel. The use of this algorithm in each of the cloud-masked images of the areas studied over 19 years and included in the OC CCI product helped to increase the analysed area, sometimes to a significant extent. Calculated from 1998 to 2016 as arithmetic means for the Barents, Bering, North, Norwegian and Greenland seas, the quantitative estimates of such an increase attained for each 8-day-averaged image reached, respectively, $\sim$ 107, 370, 31, 15 and 13 times. Thus, images were obtained with significantly larger cloud-free areas, ensuring a more accurate estimation of the borders of bloom areas, and their displacement, as well as of bloom areas per se.
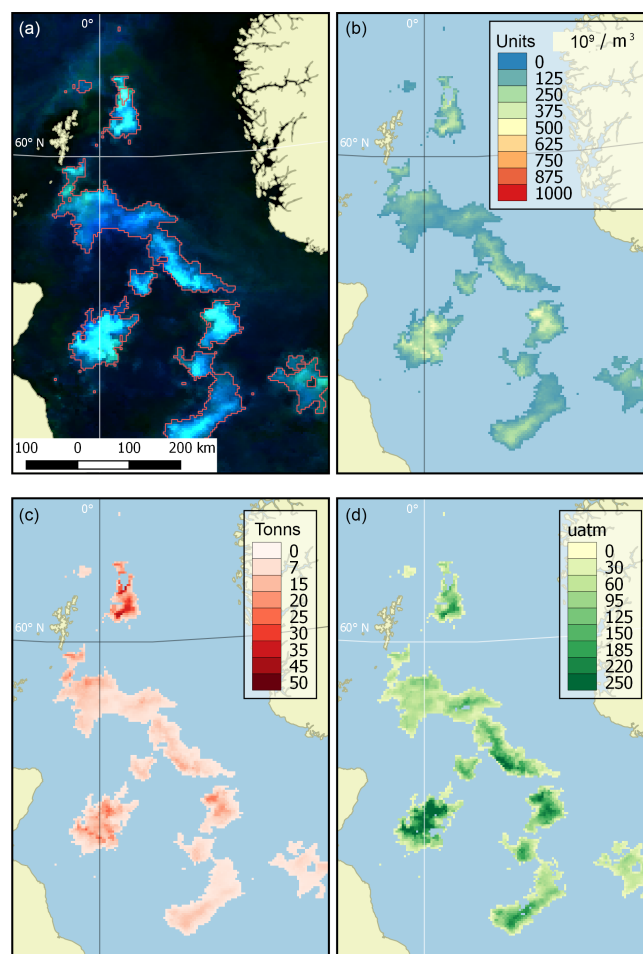

Figure 2. Example of dataset products (the North Sea, 9 June 2016). **(a)** Source OC CCI RGB imagery with the bloom mask contoured in red, **(b)** coccolith concentration (10^9 m^{-3}), **(c)** content of particulate inorganic carbon (tonnes) and **(d)** an increase in CO_2 partial pressure in water (µatm).

Examples of product visualisation (for the North Sea) are shown in Fig. 2.

3   Data sources

Data on R_{rs} in six channels (centred at 412, 443, 490, 510 and 670 nm) are from the OC CCI product (Ocean Colour Climate Change Initiative dataset, version 3.0, European Space Agency, available online at http://www.esa-oceancolour-cci. org/, last access: 10 January 2019).

For the bio-optical retrieval algorithm validation, we employed the PANGAEA database (https://www.pangaea.de/, last access: 10 January 2019) of the concentration of coccoliths within the target coccolithophore blooms in the North Atlantic, including the North and Norwegian seas (Charalampopoulou et al., 2008, 2011).

The bio-optical in situ database, spanning between 1997 and 2012 (16 years) was employed for ocean-colour satellite applications because it has global coverage (Valente et al., 2016). The data were acquired from several sources: MOBY (Marine Optical BuoY), BOUSSOLE (BOUée pour l'acquiSition d'une Série Optique à Long termE), AERONET-OC (AErosol RObotic NETwork-Ocean Color), SeaBASS (SeaWiFS Bio-Optical Archive and Storage System), NOMAD (NASA bio-Optical Marine Algorithm Dataset), MERMAID (MERIS Match-up In situ Database), AMT (Atlantic Meridional Transect), ICES (International Council for the Exploration of the Sea), HOT (Hawaii Ocean Time-series), and GeP&CO (Geochemistry, Phytoplankton, and Color of the Ocean). This database comprises a large number of variables, including the spectral remote-sensing reflectance, R_{rs} and chlorophyll a concentration.

Data on MLD were derived from the Montegut climatology (Montegut et al., 2004). Data on bathymetry inherent in the target seas were taken from the website http://www.ngdc.noaa.gov/mgg/bathymetry/arctic/arctic.html (last access: 10 January 2019; Jakobsson et al., 2012).

The GLobal Ocean Data Analysis Project (GLODAP) database (Key et al., 2015; Olsen et al., 2016), http://cdiac.ornl.gov/oceans/GLODAPv2/ (last access: 10 January 2019) was employed for pairing in situ NO_3 values at those points for which in situ pCO_2 values were available. In the cases when the desired NO_3 matching values were unavailable in the GLODAP database, the respective data were employed from the World Ocean Atlas 2013 (WOA13, NOAA, Garcia et al., 2013; https://www.nodc.noaa.gov/OC5/woa13/, last access: 10 January 2019).

The SOCAT v4 database (The Surface Ocean CO_2 Atlas; Bakker et al., 2016; http://www.socat.info/access.html, last access: 10 January 2019) comprises more than 6 million pCO_2 measurements performed at latitudes north of 40° N. The data employed by us from SOCAT V4 database met the following requirements: (1) measurements are conducted during 1998–2016 and within a 10 m top layer (if there were data from several depths, the measurements from the shallowest depth were used); (2) pCO_2 data should necessarily have both corresponding seawater salinity data and valid R_{rs} spectra; (3) a daily mean pCO_2 value was employed provided there were several in situ measurements; (4) pCO_2 measurements are conducted at a distance not less than 8 km offshore (to avoid the impact of adjacency effect on R_{rs} satellite data); (5) pCO_2 measurements were within the location and timing of *Emiliania huxleyi* blooming; and (6) data used from the SOCAT v4 database overlap the data from either the GLODAP database or the WOA13 climatology database (depending on which one was used for comparison).
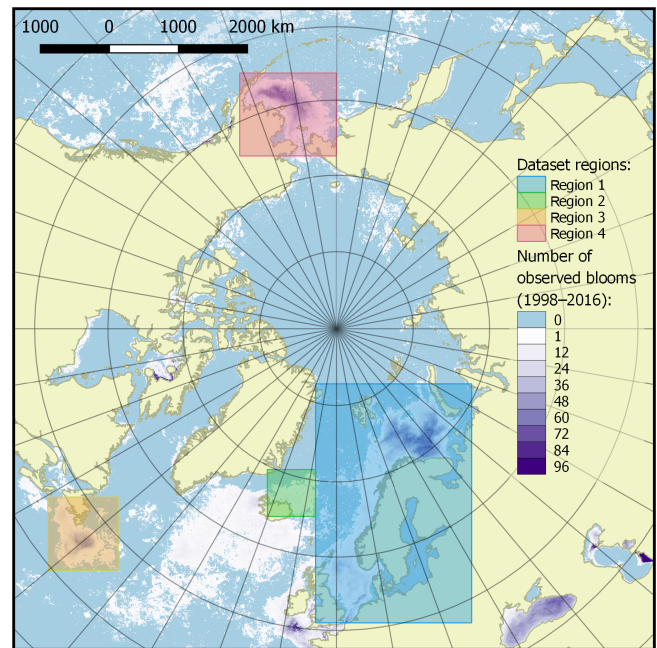

Figure 3. Dataset of target spatial regions. Regions are shown as coloured boxes, and the colour bar indicates the number of bloom observations in each pixel over the time period 1998–2016.

4 Data spatio-temporal domain

The published dataset covers a time period of 19 years, from 1998 to 2016, with a time resolution of 8 days (a total of 874 time periods), and a spatial domain with the total area of $11\,056\,800\,km^2$ at a resolution of $4 \times 4\,km$, divided into four regions described in Table 1 and shown in Fig. 3.

All data are represented in the Lambert azimuthal equal area projection with the parameters corresponding to the widespread NSIDC EASE-Grid North (EPSG: 3973) coordinate system.
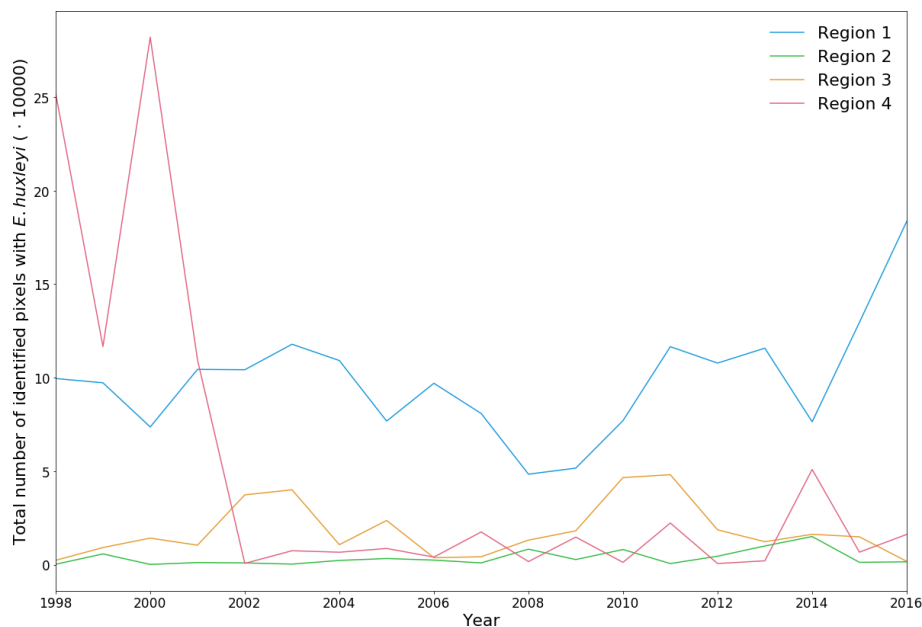
The selection of four regions in this work was made for several reasons. They include all seas where coccolithophore blooms usually occur in subpolar and polar regions of the Northern Hemisphere (North, Norwegian, Greenland, Barents, Bering and Labrador seas). The exclusion from our dataset of blooms occurring in the northern parts of the Atlantic Ocean (see, e.g. Holligan et al., 1993) was dictated by some technical restrictions: the hydro-optical model employed for obtaining coccolith concentration values was based prevalently on the data from high-latitude areas and thus should first be validated for geographically different marine environments such as open parts of the Atlantic Ocean.

5 Dataset overview

The 19-year period data covers four blooming regions differing in nature. This allows us to evaluate the bloom-related processes on different scales and time intervals in order to

Table 1. Description of spatial regions.

Region number	Extent coordinates (NSIDC EASE-Grid North, EPSG:3973)		Region area, km^2	Contained waters
1	X_{min}	−300 000.00	7 819 600	The Barents, Norwegian and North seas, the northern part of the Greenland Sea
	Y_{min}	−4 260 000.00		
	X_{max}	1 960 000.00		
	Y_{max}	−800 000.00		
2	X_{min}	−1 000 000.00	476 000	Southern part of the Greenland sea, western part of the Norwegian Sea
	Y_{min}	−2 720 000.00		
	X_{max}	−300 000.00		
	Y_{max}	−2 040 000.00		
3	X_{min}	−4 180 000.00	1 081 200	Southern part of the Labrador Sea, the North Atlantic Ocean part to the south of the Labrador Sea
	Y_{min}	−3 500 000.00		
	X_{max}	−3 160 000.00		
	Y_{max}	−2 440 000.00		
4	X_{min}	−1 400 000.00	1 680 000	The Bering Sea
	Y_{min}	2 500 000.00		
	X_{max}	0.00		
	Y_{max}	3 700 000.00		

Figure 4. Total number of identified pixels with *Emiliania huxleyi* for each blooming season in the period 1998–2016 within the four regions specified in Fig. 3.

reveal both interannual dynamics and seasonal variations of parameters relevant to the bloom phenomenon. *Emiliania huxleyi* blooms in the Arctic and sub-Arctic seas are characterised by significant instability: the difference in intensity of blooming in different years can be 10-fold. Figure 4 and Table 2 collectively illustrate for the above four marine regions the temporal dynamics in bloom intensity (i.e. bloom-

ing area). For example, in the Bering Sea (region 4), the most extensive blooms were observed exclusively from 1998 to 2001, but later on, their intensity decreased drastically. In region 1, mainly in the Barents, Norwegian and North seas, the blooming activity over the years we are reporting on was very irregular, with a peak in 2016.

Table 2. Total number of identified pixels with *E. huxleyi* for each blooming season in the period 1998–2016 within the four regions.

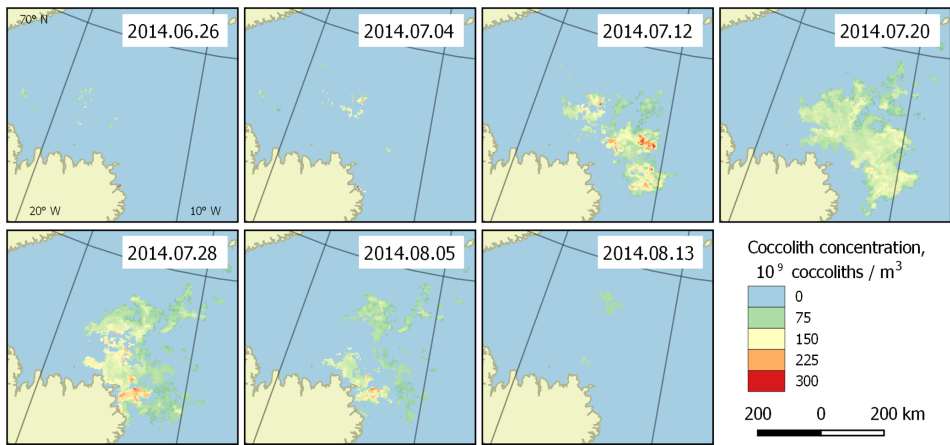| Year	Total number of pixels with *E. huxleyi*				Year	Total number of pixels with *E. huxleyi*			
	Region 1	Region 2	Region 3	Region 4		Region 1	Region 2	Region 3	Region 4
1998	99 538	214	2336	252 003	2008	48 399	8319	13 131	1656
1999	97 259	5754	9168	116 622	2009	51 620	2745	18 102	14 749
2000	73 642	138	14 205	282 046	2010	77 050	8110	46 591	1232
2001	104 425	1142	10 432	109 541	2011	116 555	603	48 101	22 259
2002	104 237	949	37 335	694	2012	107 791	4532	18 630	618
2003	117 877	312	40 018	7466	2013	115 764	10 011	12 302	2079
2004	109 156	2275	10 686	6657	2014	76 396	15 047	16 245	50 900
2005	76 768	3300	23 651	8679	2016	129 569	1265	14 890	6705
2006	97 004	2444	3729	4061	2017	183 546	1536	1779	16 184
2007	80 835	955	4237	17 505					

Figure 5. Bloom development in the Greenland Sea (region 2) in June–August 2014. The peak falls on 20 July.
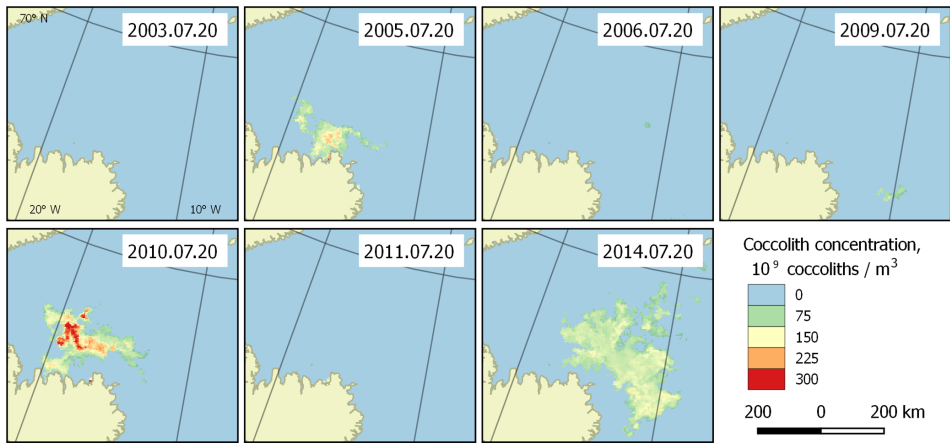

Figure 6. Bloom intensity in the Greenland Sea (region 2) on 20 July in different years. Its instability is obvious.

With the collected data, it is possible to highlight the patterns of development of the regularly occurring blooms. They can be characterised with the beginnings and ends of blooming periods, and the overall dynamics of coccolith concentration during the blooms. Such patterns can be established based on the published dataset. Figure 5 shows an example of bloom development in the Greenland Sea (region 2) in the period 26 June–13 August 2014. However, these periods are generally unstable, which is clearly seen in Fig. 6, which dis-
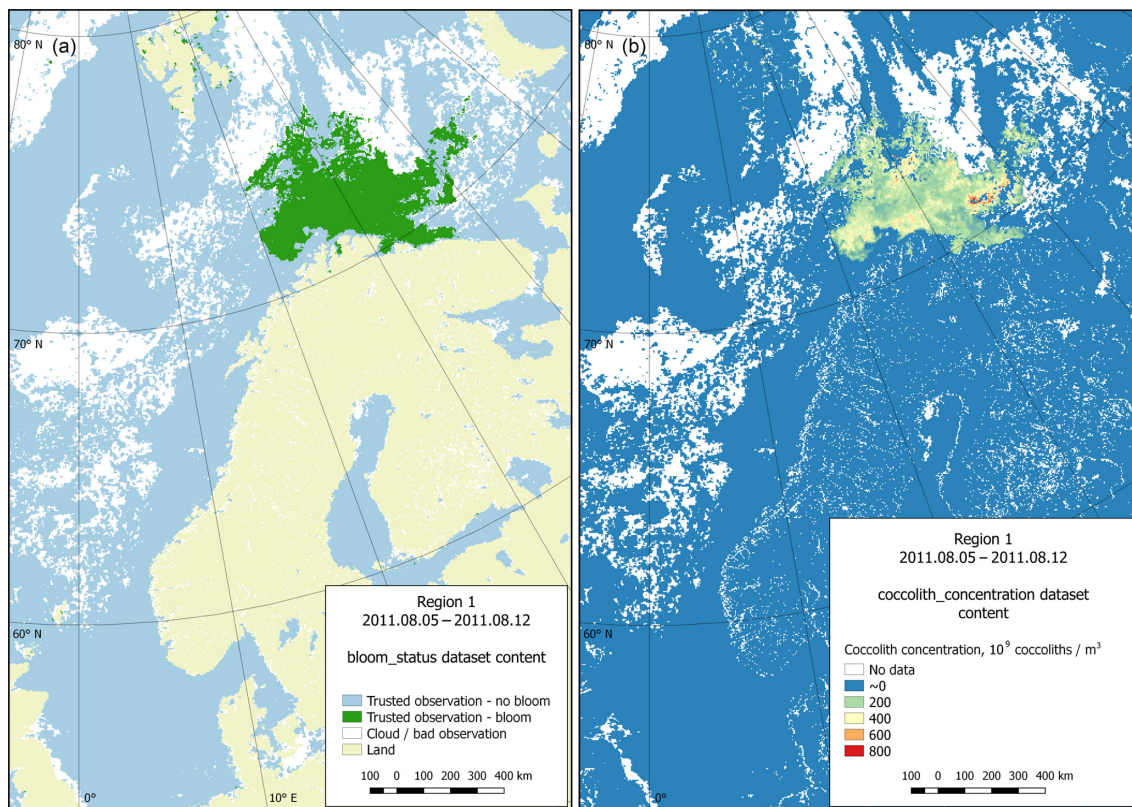
Figure 7. Dataset content example (region 1, 5 August 2011). **(a)** Bloom status sub-dataset visualisation; **(b)** coccolith concentration sub-dataset visualisation.

plays the blooming area configuration in 20 July for different years for the same area.

Technically, each dataset contains four sub-datasets: bloom status, coccolith concentration, particulate organic carbon content and CO_2 partial pressure in water driven by coccolithophores. The last three categories contain the directly calculated parameter values. The first sub-dataset contains information on the quality and content of the data. This information is organised as a set of flags attributed to data on reliable observations of the presence or absence of blooming, or inaccurate data (usually due to clouds), as well as data on the coast. Figure 7 provides both an example of a status matrix and the matrix containing coccolith concentration values.

6   Data availability

The dataset is available on Zenodo (Kondrik et al., 2018b; https://doi.org/10.5281/zenodo.1402033). Data granules are divided into directories by region and year, and each child directory contains files with 8-day data on the bloom status, coccolith concentration, PIC and $\Delta p\mathrm{CO}_2$. Data are stored in NetCDF4 format with GDAL-support, which allows immediate use of the data with any NetCDF-based or GIS software. Tips on how to read the data and on QGIS styles for fast visualisation are also provided.

7   Conclusions

We have composed a detailed 19-year dataset of *Emiliania huxleyi* blooms in the Arctic and sub-Arctic seas, including information about their influence on the carbon cycle in the ocean. These data are based mostly on satellite remote-sensing observations, but also on available shipborne measurements and results of processing with authorial algorithms. We hope that the publication of these data, on the one hand, will promote further studies aimed at elucidating *Emiliania huxleyi* bloom driving mechanisms and their forcing factors and, on the other hand, will facilitate an understanding of the patterns of this phenomenon distribution and its impact on the ocean and the atmosphere.

Author contributions.   DP is responsible for the theoretical background and methodology development. DK also contributed to theoretical background research and is responsible for the development and programming of data processing algorithms. EK conceived the dataset structure and contributed to programming of data processing algorithms, data analysis and visualisations. All authors equally contributed to the writing of the manuscript and data quality control.

References

Bakker, D. C. E., Pfeil, B., O'Brien, K. M., Currie, K. I., Jones, S. D., Landa, C. S., Lauvset, S. K., Metzl, N., Munro, D. R., Nakaoka, S.-I., Olsen, A., Pierrot, D., Saito, S., Smith, K., Sweeney, C., Takahashi, T., Wada, C., Wanninkhof, R., Alin, S. R., Becker, M., Bellerby, R. G. J., Borges, A. V., Boutin, J., Bozec, Y., Burger, E., Cai, W.-J., Castle, R. D., Cosca, C. E., DeGrandpre, M. D., Donnelly, M., Eischeid, G., Feely, R. A., Gkritzalis, T., Gonzalez-Davila, M., Goyet, C., Guillot, A., Hardman-Mountford, N. J., Hauck, J., Hoppema, M., Humphreys, M. P., Hunt, C. W., Ibanhez, J. S. P., Ichikawa, T., Ishii, M., Juranek, L. W., Kitidis, V., Kortzinger, A., Koffi, U. K., Kozyr, A., Kuwata, A., Lefevre, N., Lo Monaco, C., Manke, A., Marrec, P., Mathis, J. T., Millero, F. J., Monacci, N., Monteiro, P. M. S., Murata, A., Newberger, T., Nojiri, Y., Nonaka, I., Omar, A. M., Ono, T., Padin, X. A., Rehder, G., Rutgersson, A., Sabine, C. L., Salisbury, J., Santana-Casiano, J. M., Sasano, D., Schuster, U., Sieger, R., Skjelvan, I., Steinhoff, T., Sullivan, K., Sutherland, S. C., Sutton, A., Tadokoro, K., Telszewski, M., Thomas, H., Tilbrook, B., van Heuven, S., Vandemark, D., Wallace, D. W. R., and Woosley, R.: Surface Ocean CO$_2$ Atlas (SOCAT) V4, https://doi.org/10.1594/PANGAEA.866856, 2016.

Balch, W. M., Kilpatrick, K. A., and Trees, C. C.: The 1991 coccolithophore bloom in the central North Atlantic. 1. Optical properties and factors affecting their distribution, Limnol. Oceanogr., 41, 1669–1683, 1996a.

Balch, W. M., Kilpatrick, K. A., Holligan, P. M., Harbour, D., and Fernandez, E.: The 1991 coccolithophore bloom in the central North Atlantic. 2. Relating optics to coccolith concentration, Limnol. Oceanogr., 41, 1684–1696, 1996b.

Balch, W. M., Gordon, H., Bowler, B. C., Drapeau, D. T., and Booth, E. S: Calcium Carbonate Measurements in the Surface Global Ocean Based on Moderate-Resolution Imaging Spectrometer Data, J. Geophys. Res., 110, C07001, https://doi.org/10.1029/2004JC002560, 2005.

Balch, W. M., Bates, N. R., Lam, P. J., Twining, B. S., Rosengard, S. Z., Bowler, B. C., Drapeau, D. T., Garley, R., Lubelczyk, L. C., Mitchell, C., and Rauschenberg, S.: Factors regulat-

ing the Great Calcite Belt in the Southern Ocean and its biogeochemical significance, Global Biogeochem. Cy., 30, 1124–1144, https://doi.org/10.1002/2016GB005414, 2016.

Bukata, R. P., Jerome, J. H., Kondratyev, K. Ya., and Pozdnyakov, D. V.: Optical Properties and Remote Sensing of Inland and Coastal Waters, CRC Press, Boca Raton, FL, USA, 362 pp., 1995.

Charalampopoulou, A., Poulton, A. J., Tyrrell, T., and Lucas, M.: Surface Seawater Carbonate Chemistry, Nutrients and Phytoplankton Community Composition on a Transect between North Sea and Arctic Ocean, 2008, https://doi.org/10.1594/PANGAEA.763990, 2008.

Charalampopoulou, A., Poulton, A. J., Tyrrell, T., and Lucas, M: Irradiance and pH Affect Coccolithophore Community Composition on a Transect between the North Sea and the Arctic Ocean, Mar. Ecol. Prog. Ser., 431, 25–43, https://doi.org/10.3354/meps09140, 2011.

Copin-Montegut, C.: A new formula for the effect of temperature on the partial pressure of CO$_2$ in seawater, Mar. Chem., 25, 29–37, https://doi.org/10.1016/0304-4203(88)90012-6, 1988.

Garcia, H. E., Locarnini, R. A., Boyer, T. P., Antonov, J. I., Baranova, O. K., Zweng, M. M., Reagan, J. R., Johnson, D. R., Mishonov, A. V., and Levitus, S.: World ocean atlas 2013, vol. 4, Dissolved inorganic nutrients (phosphate, nitrate, silicate), Ocean Climate Laboratory, National Oceanographic Data Center, Silver Spring, Maryland, USA, https://doi.org/10.7289/V5J67DWD, 2013.

Gordon, H. R., Boynton, G. C., Balch, W. M., Groom, S. B., Harbour, D. S., and Smyth, T. J.: Retrieval of Coccolithophore Calcite Concentration from SeaWiFS Imagery, Geophys. Res. Lett., 28, 1587–1590, 2001.

Holligan, P. M., Fernández, E., Aiken, J., Balch, W. M., Boyd, P., Burkill, P. H., Finch, M., Groom, S. B., Malin, G., Muller, K., Purdie, D. A., Robinson, C., Trees, C. C., Turner, S. M., and van der Wal, P.: A biogeochemical study of the coccolithophore, *Emiliania huxleyi*, in the North Atlantic, Global Biogeochem. Cy., 7, 879–900, https://doi.org/10.1029/93GB01731, 1993.

Jakobsson, M., Mayer, L. A., Coakley, B., Dowdeswell, J. A., Forbes, S., Fridman, B., Hodnesdal, H., Noormets, R., Pedersen, R., Rebesco, M., Schenke, H. W., Zarayskaya, Y., Accettella, D., Armstrong, A., Anderson, R. M., Bienhoff, P., Camerlenghi, A., Church, I., Edwards, M., Gardner, J. V, Hall, J. K., Hell, B., Hestvik, O., Kristoffersen, Y., Marcussen, C., Mohammad, R., Mosher, D., Nghiem, S. V, Pedrosa, M. T., Travaglini, P. G., and Weatherall, P.: The International Bathymetric Chart of the Arctic Ocean (IBCAO) Version 3.0., Geophys. Res. Lett., 39, L12609, https://doi.org/10.1029/2012GL052219, 2012.

Jerome, J. H., Bukata, R. P., and Miller, J. R.: Remote sensing reflectance and its relationship to optical properties of natural water, Int. J. Remote Sens., 17, 43–52, 1996.

Johannessen, O. M.: Decreasing Arctic ice mirrors increasing CO$_2$ on decadal time scale, Atmospheric and Oceanic Science Letters, 1, 51–56, 2008.

Key, R. M., Olsen, A., van Heuven, S., Lauvset, S. K., Velo, A., Lin, X., Schirnick, C., Kozyr, A., Tanhua, T., Hoppema, M., Jutterström, S., Steinfeldt, R., Jeansson, E., Ishi, M., Perez, F. F., and Suzuki, T.: Global Ocean Data Analysis Project, version 2 (GLODAPv2), ORNL/CDIAC-162, ND-P093. Carbon Dioxide Information Analysis Center, Oak Ridge National

Laboratory, US Department of Energy, Oak Ridge, TN, USA, https://doi.org/10.3334/CDIAC/OTG.NDP093_GLODAPv2, 2015.

Kondrik, D. V., Pozdnyakov, D. V., and Pettersson, L. H.: Particulate inorganic carbon production within *E. huxleyi* blooms in subpolar and polar seas: a satellite time series study (1998–2013), Int. J. Remote Sens., 38, 6179–6205, https://doi.org/10.1080/01431161.2017.1350304, 2017a.

Kondrik, D., Pozdnyakov, D., and Pettersson, L.: Tendencies in Coccolithophorid Blooms in Some Marine Environments of the Northern Hemisphere According to the Data of Satellite Observations in 1998–2013, Izv. Atmos. Ocean. Phys.+, 53, 955–964, 2017b.

Kondrik, D. V., Pozdnyakov, D. V., and Johannessen, O. M.: Satellite evidence that *E. huxleyi* phytoplankton blooms weaken marine carbon sinks, Geophys. Res. Lett., 45, 846–854, https://doi.org/10.1002/2017GL076240, 2018a.

Kondrik, D., Kazakov, E., and Pozdnyakov, D.: Dataset of E. huxleyi blooms: spatio-temporal distribution and their impact on high-latitudinal marine environments (1998–2016) (Version 1) [Dataset], Zenodo, https://doi.org/10.5281/zenodo.1402033, 2018b.

Korosov, A. A., Pozdnyakov, D. V., Folkestad, A., Pettersson, L. H., Sorensen, K., and Shuchman, R.: Semi-empirical algorithm for the retrieval of ecology-relevant water constituents in various aquatic environments, Algorithms, 2, 470–497, https://doi.org/10.3390/a2010470, 2009.

Loveday, B. R. and Smyth, T.: A 40-year global data set of visible-channel remote-sensing reflectances and coccolithophore bloom occurrence derived from the Advanced Very High Resolution Radiometer catalogue, Earth Syst. Sci. Data, 10, 2043–2054, https://doi.org/10.5194/essd-10-2043-2018, 2018.

Montegut, C., Madec, G., Fisher, A., Lazar, A., and Iudicone, D.: Mixed layer depth over the global ocean: an examination of profile data and a profile-based climatology, J. Geophys. Res., 109, C12003, https://doi.org/10.1029/2004JC002378, 2004.

Olsen, A., Key, R. M., van Heuven, S., Lauvset, S. K., Velo, A., Lin, X., Schirnick, C., Kozyr, A., Tanhua, T., Hoppema, M., Jutterström, S., Steinfeldt, R., Jeansson, E., Ishii, M., Pérez, F. F., and Suzuki, T.: The Global Ocean Data Analysis Project version 2 (GLODAPv2) – an internally consistent data product for the world ocean, Earth Syst. Sci. Data, 8, 297–323, https://doi.org/10.5194/essd-8-297-2016, 2016.

Press, W., Teukolsky, S., Vettering, W., and Flannery, B.: Numerical recipes in C: The art of scientific computing, 2nd edn., Cambridge University Press, New York, USA, 1992.

Rost, B., Zondervan, I., and Wolf-Gladrow, D.: Sensitivity of phytoplankton to future changes in ocean carbonate chemistry: current knowledge, contradictions and research directions, Mar. Ecol. Prog. Ser., 373, 227–237, 2008.

Shutler, J., Grant, M., Miller, P., Rushton, E., and Anderson, K.: Coccolithophore bloom detection in the north Atlantic using SeaWiFS: Algorithm description, application and sensitivity analysis, Remote Sens. Environ., 114, 1008–1016, https://doi.org/10.1016/j.rse.2009.12.024, 2010.

Shutler, J. D., Land, P. E., Brown, C. W., Findlay, H. S., Donlon, C. J., Medland, M., Snooke, R., and Blackford, J. C.: Coccolithophore surface distributions in the North Atlantic and their modulation of the air-sea flux of CO_2 from 10 years of satellite Earth observation data, Biogeosciences, 10, 2699–2709, https://doi.org/10.5194/bg-10-2699-2013, 2013.

Valente, A., Sathyendranath, S., Brotas, V., Groom, S., Grant, M., Taberner, M., Antoine, D., Arnone, R., Balch, W. M., Barker, K., Barlow, R., Bélanger, S., Berthon, J.-F., Besiktepe, S., Brando, V., Canuti, E., Chavez, F., Claustre, H., Crout, R., Frouin, R., García-Soto, C., Gibb, S. W., Gould, R., Hooker, S., Kahru, M., Klein, H., Kratzer, S., Loisel, H., McKee, D., Mitchell, B. G., Moisan, T., Muller-Karger, F., O'Dowd, L., Ondrusek, M., Poulton, A. J., Repecaud, M., Smyth, T., Sosik, H. M., Twardowski, M., Voss, K., Werdell, J., Wernand, M., and Zibordi, G.: A compilation of global bio-optical in situ data for ocean-colour satellite applications, Earth Syst. Sci. Data, 8, 235–252, https://doi.org/10.5194/essd-8-235-2016, 2016.

Winter, A., Hendericks, J., Beaufort, L., Rickaby, R. E. M., and Brown, C. W.: Poleward expansion of the coccolithophore Emiliania huxleyi, J. Plankton Res., 36, 316–325, https://doi.org/10.1093/plankt/fbt110, 2014.